Studies in History and Philosophy of Science 65-66 (2017) 46–56

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

A structural interpretation of measurement and some related epistemological issues

Luca Mari [a,*], Paolo Carbone [b], Alessandro Giordani [c], Dario Petri [d]

[a] *Università Cattaneo - LIUC, Italy*
[b] *Università di Perugia, Italy*
[c] *Università Cattolica, Italy*
[d] *Università di Trento, Italy*

ARTICLE INFO

Article history:
Received 6 December 2015
Received in revised form
15 December 2016
Available online 18 August 2017

Keywords:
Measurement
Reliability of measurement
Stereotypes of measurement
Models in measurement
Objectivity and intersubjectivity in measurement

ABSTRACT

Measurement is widely applied because its results are assumed to be more reliable than opinions and guesses, but this reliability is sometimes justified in a stereotyped way. After a critical analysis of such stereotypes, a structural characterization of measurement is proposed, as partly empirical and partly theoretical process, by showing that it is in fact the structure of the process that guarantees the reliability of its results. On this basis the role and the structure of background knowledge in measurement and the justification of the conditions of object-relatedness ("objectivity") and subject-independence ("inter-subjectivity") of measurement are specifically discussed.

1. Introduction

Measurement is applied in a range of fields, from quantum physics to daily commercial transactions, because of its special epistemic role: its results are rightly assumed to be more reliable than, say, opinions and guesses (note that we will maintain here the distinction between measurement – a process – and measurement result – its outcome). The reasons for this claim are not however so clearly recognized, and in fact are overloaded with many stereotypes. The risk of allowing these stereotypes to persist, unidentified, in the public understanding of measurement, is that we may sooner or later find ourselves in a situation of "anything goes," with regards measurement: the consequence would be an inability to justify the reliability of measurement results and, ultimately, to distinguish measurement from mere opinion-making. Of course, exhibiting the rich body of knowledge around measuring instruments is not sufficient: engineering alone cannot prevent relativism. This is an endeavor that calls for an epistemological analysis, aimed at debunking the stereotypes that have been plaguing measurement and at providing a consistent justification of the special role of measurement itself.

We present such stereotypes as conceptually categorized in three main clusters, related to naive realist, operationalist, and representationalist interpretations of measurement. By highlighting and briefly discussing them, we show that they are unable to explain the epistemic role of measurement, which we claim has to be justified, instead, on a different ground. We then provide such a justification, in terms of a structural interpretation of measurement that takes into account, and remedies, the flaws of the stereotyped interpretations: measurement emerges as an object-related and subject-independent process, variously tied to background knowledge.

The paper is organized as follows. In Section 2 we identify and discuss three main clusters of stereotypes concerning the characterization of the measurement process and argue for the necessity of a better account of this process. In Section 3 we analyze two representative examples in which complex measuring processes are brought about, thus also showing the limits of the stereotyped interpretations. We suggest then, in Section 4, that the structural stages of measurement are given by the design and the realization of three different but related models, wherein pre-measurement and post-measurement knowledge about both the object under measurement and the property to be measured are at work. Starting from these cases and this analysis, Section 5 is devoted to arguing for object-relatedness and subject-independence as basic

* Corresponding author.
E-mail address: lmari@liuc.it (L. Mari).

http://dx.doi.org/10.1016/j.shpsa.2017.08.001
0039-3681/

criteria that distinguish measurement from other processes of evaluation. Given this perspective, we review the examples introduced in Section 3.

2. Stereotyping measurement

2.1. Cluster one: naive realist interpretations of measurement

A straightforward, non-sophisticated realist interpretation of measurement, whose fundamental ideas are still a common reference for some metrologists (see, e.g., Bentley, 2005), and possibly some philosophers of measurement (Michell, 2005), can be synthesized as follows: (1) what is measured are properties whose ratio is invariant: quantities; (2) ratios of quantities exist in the world, independently of any measurement process; (3) measurement is the process through which such pre-existing ratios are determined, i.e., discovered; (4) measuring systems receive as inputs the true values of the measured quantities and produce as outputs measured values of such quantities; (5) in ideal measuring systems the measured value is equal to the true value, and therefore an ideal measuring system is intended as the empirical implementation of the identity function; (6) measuring systems cannot be ideal due to their empirical nature; (7) hence measurements cannot be error-free.

These ideas fit well with an abstract approach to the entities that constitute the domain of the measurement processes, according to which what we measure is like the length of a segment in a mathematical space. However, while attractive for its structural simplicity, such a naive realist interpretation is problematic, with respect to its consequences both as a theory of measurement and as a fundamental understanding of the measurement process. On the one hand, as a basis for a theory of quantity, the risk of such a view is to neglect the role of experimental processes in the characterization of empirical quantities and to focus on pure mathematical characterizations (see Hölder, 1996 [1901]; Mundy, 1987). On the other hand, as a basis for a theory of the measurement process, the main problems are connected to the following four points[1]:

- the assumption that ratios of quantities exist independently of measurement processes, so that such ratios constitute what is to be measured independently of our intervention of the world;
- the assumption that these processes are ideally intended as empirical implementations of identity functions, so that their outputs are to coincide with their inputs;
- the consequent assumption that true values are out there in the world, so that the definition of the measurand is in principle not affected by uncertainty; and
- the consequent assumption that the notions of measurement accuracy and measurement precision can be defined without problems with reference to true values.

All these assumptions can be questioned, since they seem to be at odds both with the current reflection on measurement practice (in particular with respect to the role of the activity of ideal construction of the measurand) and with the current understanding of some fundamental metrological notions (in particular with respect to the concept of definitional uncertainty). Such a stereotypical position neglects the essential role of models both in the definition of the measurand and in the development of the measurement procedure and the interpretation of the results of its application (for further details, see, e.g., Frigerio, Giordani, & Mari, 2010; Teller, 2013).

2.2. Cluster two: operationalist interpretations of measurement

An opposing, yet also not uncommon (since Bridgman, 1927), interpretation of measurement is grounded on the assumption that the quantity which is measured is determined by the methods that we use to measure it, so that the quantity the measurement result is attributed to is precisely the one with which the measuring system interacted. Through the application of statistical methods, the repeatability of measurement is assessed and non-identical results are interpreted as caused by the imperfect behavior of the measuring system, and formalized in terms of measurement error.

Again, while attractive for its epistemological simplicity, such a purely operationalist interpretation of the process of measurement is problematic, since it is at odds with the fact that measurement results are pieces of information that can be reported and understood outside the specific experimental context in which they were obtained. Hence, such results are typically not attributed to the quantity with which the measuring instrument experimentally interacted, but to the measurand, defined by the *International Vocabulary of Metrology* (VIM) as the "quantity intended to be measured" (JCGM, 2012, 2.3). This reference to intentions emphasizes that a model of the measurand, as the individual quantity in the specific context of measurement, is unavoidably present. Generally, the measurand is different from the quantity with which the measuring instrument interacts, despite the best efforts of measurement engineering. That is why the core concept used today to take into account the non-ideality of measurement is no longer 'measurement error,' but 'measurement uncertainty'. According to the *Guide to the expression of uncertainty in measurement* (GUM), "The concept of uncertainty as a quantifiable attribute is relatively new in the history of measurement, although error and error analysis have long been a part of the practice of measurement science or metrology. It is now widely recognized that, when all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured" (JCGM, 2008, 0.2).

In summary, by interpreting it as a purely operationalist process, measurement is reduced to transduction, and the only option to give it an epistemic characterization is in terms of its inputs or its outputs, up to a point in which measurement is defined as "any precisely specified operation that yields a number" (Dingle, 1950, p. 11).

2.3. Cluster three: representationalist interpretations of measurement

With a conceptual path that started from Euclid and includes contributions by, among others, Newton and Maxwell, and more recently Campbell (1957), what in the classical world was a feature of measure – numerical ratios of additive entities – has become a definitional characteristic of measurement. More crucially, the emphasis on measurement as representation (Krantz, Luce, Suppes, & Tversky, 1971) has generated a shift from the original ontological claims which assumed the propensity of measurement to obtain truth, but for the presence of errors, to much weaker representability conditions. For example, according to Suppes an appropriate representation theorem "makes the theory of finite weak orderings a theory of measurement, because of its numerical representation" (Suppes, 2002). Even though weakly ordered entities do not satisfy

[1] These points can be intended as an expansion of "the two fundamental defects" already highlighted by Suppes (1954, p. 172) in the system of axioms proposed by Hölder: the infinite number of elements in the set of quantities and the transitivity of the relation of partial order among them.

in general the Euclidean conditions on a measure,[2] they are considered measurable because they can be represented by means of numerical values. The stereotype is manifest: if it is representable by means of numbers, then it is the outcome of a measurement.

The key concept here is representability, a condition formalized in terms of morphic mappings from empirical entities (either objects or their properties) to symbolic entities (numbers, with or without measurement units, but possibly also other entities in the generalization by Stevens (1946), that makes the representability by means of numbers sufficient but not necessary for measurability). Hence, "the question of measurement" is the one "about the possibility of using numbers to describe certain phenomena", and by representational theories it "has received answers in the form of testable conditions" (Doignon, 1993, p. 473).

This purely formal interpretation of measurement is also attractive for its epistemological simplicity, but the idea that measurement is definitionally equivalent to morphic mapping is so generic that it is just unable to distinguish between measurement and consistent representation (Mari, 2013; Mari, Carbone, & Petri, 2012).

Rather, what the so-called "representational theories of measurement" provide is an abstract framework for scale construction and meaningfulness of representation (Narens, 1985, 2002), and therefore at most for characterizing conditions of measurability. If, for example, a given quantity is selected as the unit, then the additive combination of two instances of that quantity has to be associated with the numerical quantity value 2, and so on. This is indeed a condition of morphic mapping which does not require any empirical action to be performed on given measurands, and is in fact preliminary to any such empirical action. In other terms, scale construction is a critical precondition for the execution of measurement but it is surely not measurement as such. Hence, a theory of scale construction cannot be a theory of measurement, but perhaps a component of such a theory.

This conceptual superposition has a justification. Representational theories were developed in the context of social sciences (Duncan, 1984), where in many cases the critical activity is the characterization of the property to be measured and then in particular the construction of an appropriate scale for it. As a consequence, the critical problem is actually measurability, while executing measurement can be sometimes as trivial as obtaining a questionnaire to be filled and then counting items checked in it. On the other hand, an epistemological presupposition – that plausibly we share with representational theorists – is that measurement of physical and non-physical properties is not fundamentally different. In the context of physical measurement, representational theories have been substantially neglected until now.[3] This fact can be interpreted as a sign of the practical uselessness of such theories in situations – like most cases of physical measurement – in which the source of complexity, and hence of interest, is actually (also) the execution of measurement, not (only) the characterization of its preconditions.

In summary, by interpreting it as a purely formal process, measurement is reduced to a functional transformation satisfying input-output conditions, and the only option to give it an epistemic characterization is in terms of the consistency of the transformation.

2.4. Beyond stereotypes

These three clusters of stereotypes share a common presupposition: measurement can be studied as a black box. While usually maintained as a tacit assumption, sometimes this is explicitly acknowledged, and even claimed as a characterizing feature of an appropriate conceptualization of measurement: "we are not interested in a measuring apparatus and in the interaction between the apparatus and the objects being measured. Rather, we attempt to describe how to put measurement on a firm, well-defined foundation" (Roberts, 1979, p. 3); "the theory of measurement is difficult enough without bringing in the theory of making measurements" (Kyburg, 1984, p. 7).

Our point should be clear now. A theory of the process of measuring cannot be developed without focusing on the specific features of the process itself. Indeed, even if abstracting from these features allows us to introduce elementary and accessible models of the process, such as the models emerging from the previous interpretations, it precludes the possibility of explaining the epistemic role customarily attributed to measurement and its results, which we have proposed to characterize in terms of *object-relatedness* and *subject-independence* ("objectivity" and "intersubjectivity" for short) (Frigerio et al., 2010; Mari et al., 2012). Such features must be justified in terms of the structure of the whole process which supports the related decision-making: starting from the specification of the goals towards which information on properties is required, to the analysis of the obtained results. Hence, measurement should be understood as a process performed by means of properly designed, set up, and operated measuring systems, whose outcome is traceable to measurement standards via calibration chains, i.e., metrological systems. The actual operation of measuring instruments – which we will call here "measurement execution" – is a necessary but surely not a sufficient condition: it must be complemented with knowledge – we will call it here "background knowledge".

Such a structural interpretation should then apply in principle to the measurement of both physical and non-physical properties.

3. Understanding measurement: two examples

A theory of measurement is a theory of a specific kind of process. In this perspective we propose a characterization of measurement based on a detailed analysis of the process of measuring, construed as an object-related (*objective*) and subject-independent (*intersubjective*) process which is run by agents who are endowed with specific resources and have particular theoretical or practical purposes. We assume that agents construct and utilize models in order to represent, and possibly operate on, portions of the world and perform measurements in order to acquire information on properties of these portions.[4] Moreover, agents use measurement results in order to improve their models and develop models in order to improve their measuring systems.

In this Section, we introduce and analyze two examples of measurement which aid the identification of crucial elements in a measurement process, being: firstly, the specification of the object under measurement, the definition of the considered general

[2] In the Elements (Book 5, "Proportion"), Euclid characterized quantities (or magnitudes, as they are also called) in terms of both ordering ("A magnitude is a part of a(nother) magnitude, the lesser of the greater, when it measures the greater." – definition 1) and ratio ("A ratio is a certain type of condition with respect to size of two magnitudes of the same kind." – definition 3) (Fitzpatrick, 2008, p. 130).

[3] The most noticeable exception is in the work of Finkelstein (1975, 2009).

[4] We omit here any distinction between properties and quantities, and refer to the more generic case of properties; our position on this matter is presented in Giordani and Mari (2012b).

property,[5] and the definition of the measurand; second, the specification of the measuring system, including the choice of the measuring instruments and the design of the measurement procedure; third, the modeling activity underlying the measurement execution. The fact that the two examples discussed are significantly different yet, at the same time, admit the same structural description hints that measurement as such should be characterized in terms of its structure: this is our main thesis.

3.1. Example 1: weighing stars

In astrophysics, mass is a key property, since knowing the mass of a star allows us to infer some of its most significant characteristics and certain details of its evolution. Hence, measuring the stellar mass is crucial for developing and improving physical models. In this field, binary stars are of primary importance, since they provide a particularly accurate means of determining the masses of stars, through studying their gravitational interactions with other massive objects. In particular, given the universality of the gravitational interaction, the mass of binary stars can be measured according to the following principles and procedures.

Measurement task: measuring the mass of binary stars

General model: modeling celestial motions.

Let A and B be binary stars and choose a center of mass reference frame, so that:

$$m_A r_A = m_B r_B \tag{1}$$

where m_A and m_B are the masses of A and B, and r_A and r_B their distances from the center of mass. Let:

$$M = m_A + m_B \quad r = r_A + r_B \tag{2}$$

Then:

$$r_A/m_B = r_B/m_A = M/r \tag{3}$$

so that $r_A = m_B r/M$ and $r_B = m_A r/M$

In the case of circular orbits, the gravitational force $F_G = Gm_A m_B/r^2$ acting on the stars equals the centripetal force $F_C = m_A v_A^2/r_A$, where $v_A = 2\pi r_A/T$ is the speed of A and T is the common period. Thus:

$$Gm_B/r^2 = v_A^2/r_A = 4\pi^2 r_A/T^2 = 4\pi^2 m_B r/MT^2 \tag{4}$$

and so:

$$GM/4\pi^2 = r^3/T^2 \tag{5}$$

which is the improved version of Kepler's third law. The reference to A is no longer present here, and the total mass M can be determined by measuring the period T and the total distance r.

In the case of general elliptical orbits, the same relation can be derived (Carroll & Ostlie, 2007, ch. 2).

In addition, since (3) gives us the mass ratio, we obtain:

$$M = 4\pi^2 r^3/GT^2 \tag{6}$$

$$m_A = Mr_B/r = 4\pi^2 r^2 r_B/GT^2 \tag{7}$$

by (3) and (5)

$$m_B = Mr_A/r = 4\pi^2 r^2 r_A/GT^2 \tag{8}$$

by (3) and (5)

Hence, knowing the period T and the distances r_A and r_B from the center of mass enables us to measure the masses m_A and m_B of the stars in the binary system.

Specific model: modeling visual binaries.

A model of visual binary stars can be constructed as follows.

Step 1. assume that the orbit plane is perpendicular to the observer line of sight.

The ratio m_A/m_B coincides with $r_B/r_A = a_B/a_A$, where a_A and a_B are the major semi-axes of the ellipses. Furthermore, if the distance of the system from the observer is d, since $a_A, a_B \ll d$, then the angular distances α_A and α_B are such that $\tan\alpha_A = a_A/d$ and $\tan\alpha_B = a_B/d$. Thus we obtain:

$$m_A/m_B = \alpha_B/\alpha_A \tag{9}$$

Hence, knowing the angular distance allows us to determine the ratio of the masses, even though d is not known. Finally, knowing d together with the period allows us to determine the masses themselves, by using this instance of Kepler's third law:

$$m_A + m_B = (4\pi^2/\mathrm{G})\,(\alpha d)^3/T^2 \tag{10}$$

where $\alpha = \alpha_A + \alpha_B$.

Step 2. remove the assumption that the orbit plane is perpendicular to the observer line of sight.

If φ is the angle between the orbit plane and the sky plane, i.e., the plane perpendicular to the observer line of sight, then the ratio m_A/m_B coincides with $\alpha_B\cos\varphi/\alpha_A\cos\varphi$, and the instance of Kepler's third law is:

$$m_A + m_B = (4\pi^2/G)\,(\alpha\cos\varphi\; d)^3/T^2 \tag{11}$$

Hence, in order to know the total mass we have to determine φ. However, to deduce the angle of inclination it is sufficient to determine the apparent distance between the center of mass of the system and the foci of the apparent elliptical orbit. In fact, an ellipse tilted at an angle φ with respect to the sky plane is observed as an ellipse with different eccentricity and with the center of mass at a certain distance from the apparent foci. Therefore, the characteristic of the original orbit can be determined by comparing the observed ellipse with the projection of different ellipses in a mathematical model.

Measurement model: definition of the parameters to be measured and measurement methods.

To obtain the ratio of the masses we have to measure the period and the angle of the plane of the orbit. To obtain the masses we also have to measure the semi-axes of the orbits. All these data are to be extracted by analyzing the relative position of the stars (on the following steps, see the classic: Aitken, 1964, ch. 4).

Step 1. determining the relative position.

This can be done by using a telescope endowed with a filar micrometer. Typically, the brighter star is chosen as reference point and the position of the other star is registered as an angular distance, ρ, and a positional angle θ. Since the orbital period is in general of the order of years, the data relative to the positional

[5] Sometimes general properties, such as length, mass, etc., are called "kinds of properties".

angle, which is measured with respect to the equator of date, must be referred to the standard equator.

Step 2. determining the apparent ellipse.

This can be done by applying analytical methods for finding the best ellipse fitting a number of points determined by observation. The minimum number of points needed to carry out this operation is five. However, in order to obtain an accurate ellipse, data relative to an entire orbit are used and checked against Kepler's second law, which imposes that $\rho^2\, d\theta/dt$ be a constant.

Step 3. determining the actual ellipse.

This can be done by applying analytical methods to the apparent ellipse. Indeed, it turns out that seven standard parameters are sufficient in order to find the actual ellipse and that all of them can be obtained by operating on the equation of the apparent ellipse, leading to the determination of the period and the angle of the plane of the orbit.

Measurement execution: parameters and measurement setting

The previous modeling activity provides us with the necessary conditions for planning and performing a measurement process. Actually, what we have to do now is to choose an appropriate measuring system and to establish both the number of observations and the duration of the observation process. The process is then executed and the data that are obtained are used in order to acquire the desired results. In this part of the process it is necessary to check the calibration of the system, to compute the uncertainty associated to each measurement result, and to propagate the uncertainty up to the final result.

3.2. Example 2: weighing scientific research performances

The importance of assessing the research performance of individual researchers or institutions has considerably increased in the last few decades, the main reason being that significant economic resources are employed for funding such research projects. Hence, the problem arises of how a ranking of such performances should be constructed. The possibility of measuring these performances provides a potentially effective solution. In what follows, we try to present a general setting for understanding the current practice in measuring performances of scientific production (for a detailed introduction, see Vinkler, 2010, esp. ch. 7; for a specific metrological analysis, see Mari, Carbone, & Petri, 2015, esp. ch. 7.4).

Measurement task: measuring the scientific research performance of a researcher

General model: modeling scientific research performances.

The first basic assumption is that research performance is determined by the quality of the products of the research, so that in measuring performance what has to be evaluated is indeed the quality of such products. A second basic assumption is that the quality of a scientific product is proportional to the impact of the product on the scientific community, in view of both the profundity of the results obtained and the innovation in the relevant fields. Hence, two laws can be stated, under the assumption that the relevant properties can be compared at least on ordinal scales:

$$p <_P p' \Leftrightarrow r(p) <_R r(p') \tag{1}$$

$$r <_R r' \Leftrightarrow imp(r) < imp(r') \tag{2}$$

where p, p' vary over performances, r, r' vary over scientific products, $r(\text{-})$ is a function returning the global product of a performance, $imp(\text{-})$ is a function returning the global impact of a product, $<_P$ is an order on the set of performances, $<_R$ is an order on the set of scientific products, and $<$ is the order to be measured, i.e., the order concerning the impact of a product on the scientific community.

Some remarks:

(i) $<_P$ is the relation to be established, whose existence is assumed and not further discussed. In order to establish $p <_P p'$, given (1), it is sufficient to establish that $r(p) <_R r(p')$;

(ii) $<_R$ is an intermediate relation to be established, whose existence is also assumed and not further discussed. In order to establish $r <_R r'$, given (2), it is sufficient to establish that $imp(r) < imp(r')$;

(iii) the function $r(\text{-})$ can be empirically determined, provided a set of criteria for fixing what is to be considered as a scientific product is put forward and intersubjectively acknowledged;

(iv) the function $imp(\text{-})$ can be empirically determined, provided a set of criteria for measuring the impact of a scientific product is put forward and intersubjectively acknowledged;

(v) a final function for composing the impact of scientific products has to be provided. In fact, $r(p)$ is the global product of a performance, which could be constituted by different specific products. Hence, a final principle concerning a sort of superposition of effects is assumed, so as to make $\{r_1, \ldots, r_n\} <_R \{r'_1, \ldots, r'_n\}$ meaningful. This last point is particularly sensible, since it is difficult to think of an order preserving superposition function on ordinal properties which is both meaningful and useful.[6]

Specific model: modeling the impact of a scientific product.

Here the key idea is to assess the impact of a product by making the following assumptions. First, only journal articles are considered as scientific production. Second, journal articles derive their impact from the impact of the journal in which they are published, so that to all articles published in the same journal, the same impact, which is the impact of the journal, is attributed. Third, a weight is defined for each journal, based on how many times its articles are cited by other papers in the same and in other journals. Finally, the weight is adjusted so as to take into account self-citations, age, and possibly other traits.

Measurement model: definition of the parameters to be measured and measurement methods.

The method used here is the impact factor (IF). According to this method, the impact of a journal is the average number of citations received by the papers of the current year, that were published during the two preceding years (a journal has IF = 2 in 2015 if on average the articles published in 2013 and 2014 received 2 citations each in 2015). Once the method is selected, the value of a performance can be determined as follows.

Step 1. determine the impact of each product.

If $\{r_1, \ldots, r_n\}$ is the list representing the scientific production, then $imp(r_i) = \text{IF}(j(r_n))$, where $j(r_n)$ is the current impact factor of the

[6] A typical ordinal function between lists can be defined according to one of three well-known strategies: (1) all the items of one of the lists are better than all of the items of the other list; (2) all the items of one of the lists are better than the corresponding items of the other list, when such items exist; (3) some of the items of one of the lists are better than the corresponding items of the other list, when such items exist, and none of the other items is worse than a corresponding item of the other list, when such items exist. Still, it is evident that no such strategy is useful in cases where the best item of a list is better than the best item of the other, while the last item is worse than the last item of the other.

journal where the article was published. Hence, $\{imp(r_1), \ldots, imp(r_n)\}$ is a list of numbers, representing the global production.

Step 2. determine the impact of the global production.

If $\{imp(r_1), \ldots, imp(r_n)\}$ is the list representing the global production, then (imp, n), where

$$imp = \sum\{imp(r_1), \ldots, imp(r_n)\}/n$$

is the pair of numbers representing the chosen performance. The ordering on different pairs is lexicographic: the second numbers enter the ordering when the first numbers are equal.

Measurement execution: parameters and measurement setting

As in the previous case, the modeling activity provides us with the necessary conditions for planning and performing a measurement process. Once the list of product is given, the process can be executed at once. In this case, no calibration problems arise, and uncertainty in identifying the list can be neglected.

4. Understanding measurement: background knowledge

As the examples introduced above illustrate, measurement is a complex process whose structure is ideally describable in such a way that a measurement task is solved through the progressive introduction of a general model, a specific model, and a measurement model, and finally the measurement execution. This structure is now concisely analyzed from the point of view of the knowledge that it involves, also to highlight the intersubjective traits of this knowledge. As we will see, measurement turns out to be a highly theoretical activity, based on a vast amount of background knowledge, primarily related to general laws and closure assumptions.

4.1. Background knowledge in a measurement process

Stage 1. measurement task

Measurement is a process made on purpose, and therefore its very first step is to provide a description of the property intended to be measured, possibly with additional information on the available resources and the constraints to be taken into account.

This requires that both the object under measurement and the kind of the property to be measured are identified. The object under measurement is typically modeled as an object in a certain state within a certain state space. Therefore, in order to measure a property of an object in a certain state (i) a general property has to be selected in the state space and (ii) the object has to be assumed to be in one state of the selected state space, corresponding to the individual property intended to be measured, i.e., the measurand.[7]

In the two examples presented above the measurement task is, then, measuring the mass of binary stars and the scientific research performance of a researcher respectively.

Stage 2. general model

Information on a particular instance of a general property can be provided either through measuring instruments designed to interact with that general property or through measuring instruments designed to interact with properties that are suitably related to that general property. The choice of whether, and how, such other properties are to be used in the measurement leads to a distinction between direct and indirect methods of measurement.

In direct methods, the fact is exploited that measuring instruments are designed for (i) interacting, *according to laws*, with the object under measurement relative to a general property and (ii) identifying the individual property instantiated by that object relative to that general property. Hence, by interacting with the object under measurement, the measuring system ideally selects both a unique dimension of the object state space and a unique element of that dimension. This is the case in which measurement instruments are adopted that are specifically designed to provide information on instances of the relevant general property, so that all possible information on other general properties related via laws to the one intended to be measured is either neglected or leads to them being interpreted as influence properties (JCGM, 2012, 2.52).

In indirect methods, the fact is exploited that general properties can be connected, *according to laws*, with each other, so that one dimension of the state of an object within a certain state space can be inferred from data concerning other dimensions of the same state. This is the case in which measurement instruments are adopted that are specifically designed to provide information on instances of general properties other than the one intended to be measured, and the relevant laws are exploited to infer information on the general property intended to be measured. It should be apparent that the application of indirect methods is based on the application of direct methods, since the properties which allows us to infer the value of what we intend to measure are to be measured in a direct way.

Note that both direct and indirect measurement methods are based on the identification of a set of laws linking general properties. The set of laws on which both a direct and an indirect measurement rest constitutes the *general model* underlying a measurement process.

In both examples presented above indirect methods of measurement are adopted. To be sure, the mass of a binary star is inferred from data concerning angular distances and periods and the scientific research performance of a researcher is inferred from data concerning the number of publications and the journals in which they appeared. In both cases, the available data are obtained through direct methods: e.g., angular distances are measured by using telescopes endowed with micrometers, which are designed, on the basis of the laws of geometrical optics, precisely for measuring these kinds of distances, while the number of publications is simply measured by counting. In addition, it is worth noting that this is the stage at which idealizations concerning relevant laws are introduced. Thus, in modeling a binary system according to the laws of classical mechanics we are assuming a level of idealization according to which the system is a classical object, while in modeling a scientific performance in terms of a set of publications we are assuming that such a performance is wholly determined by a set of published results.

Stage 3. specific model

Once a general model is given, it has to be specified with respect to the kind of object that bears the property intended to be measured. The outcome is the specific model underlying the measurement process, including a model of the measurand, a specific instance of the general property intended to be measured (JCGM, 2012, 2.3), as identified in the general model.

This is the stage at which a possible non-null definitional uncertainty (JCGM, 2012, 2.27) is evaluated, and where idealizations and approximations concerning the object are introduced, allowing us to neglect some aspects of the actual world. In particular: (i) idealizations are typically involved in the assumption that the object under measurement has certain characteristics or that the

[7] For simplicity this description neglects the issues related to non-null definitional uncertainty, including those characteristic of quantum measurement.

object is located such that it is closed off from external influences (e.g., in modeling a binary system, we assume that the center of mass of the system is fixed; in modeling the impact of a scientific product, we assume that all articles published in the same journal have the same impact); (ii) approximations are typically involved in the assumption that some quantities can be substituted for other quantities or that some quantities are constant (e.g., in modeling a binary system, we equate the angular distance and the tangent of the angular distance; in modeling the impact of a scientific product, we assume that a scientific product manifests its relevance only in terms of citations).

Stage 4. measurement model

Once the specific model is given, the selection of the measuring system can be performed, including both measuring instruments, one for each property whose value will allow us to evaluate the measurand, and analytical tools, allowing us to compute the measurand given the values of such properties.

This is the stage at which the model equation is involved, for each measured property, and instrumental uncertainty enters the picture.

This characterization of the stages preceding the execution of the actual measurement shows that, in order to perform a measurement of the property of an object, we have to be able:

C1. *to identify the object*: the measuring system is expected to interact with the identified object, or to objects related to the identified one, and the measurement result is attributed to a property of that object;
C2. *to operatively specify the general property of interest*;
C3. *to have access to one or more properties of the object and possibly of the environment*.

Hence, before executing a measurement, the following problems have to be solved:

P1. *identification*: are we able to identify the object whose property we intend to measure?
P2. *measurability*: are we able to produce a procedure for measuring the general property?
P3. *accessibility*: are we able to produce a procedure for measuring the relevant individual property?

The solution of these problems rests on a body of *background knowledge* concerning how a general property is correctly classified under a property evaluation type (see Giordani and Mari (2012a) where: first, the concept of property evaluation type is presented as a generalization of scales of measurement in Stevens' (1946) sense); second, how an access to the individual property is suitably specified.

As to the first point, it should be possible: (i) to exhibit the scale construction which enables us to assign the general property to its type; (ii) to specify in which way the measuring systems we want to apply can be calibrated, so as to ensure the traceability of the measurement results to the primary measurement standards and therefore the *subject-independence* (i.e., the intersubjectivity) *of the results*.

As to the second point, it should be possible: (i) to exhibit the link which enables us to access the individual property; (ii) to specify in which way the laws used to construct the link can be justified, so as to ensure the *object-relatedness* (i.e., the objectivity) *of the process*.

These points are significant for the task of acquiring information on a certain object: the knowledge of the possibility of measuring a general property P by measuring Q and connecting P with Q by means of appropriate laws is not useful if we are unable to measure Q with an appropriate measuring system. Similarly, the knowledge that an appropriate measuring system for measuring Q is available is not useful if we want to measure P and we have no appropriate laws connecting P with Q.

The intersubjectivity of the results and the objectivity of the process are fundamental conditions that allow us to distinguish measurement from opinion making. Recognizing the importance of these conditions helps us to understand why the first cluster of stereotypes provides a distorted notion of measurement. In fact, a measuring instrument may interact in an appropriate way with an object under measurement and produce information on a property which has a determinate ratio with another property taken as unit, and at the same time this could not warrant the claim that the obtained results are intersubjective, if the measuring instrument is not properly calibrated. Vice versa, even a well calibrated measuring system might not be sufficient to enable an objective process, if, for example, measuring instruments are not operated according to the expected procedure.

4.2. The structure of background knowledge of measurement

The background knowledge underlying measurement can be classified as knowledge applied before and after executing a measurement. Let us call it *pre-measurement knowledge* and *post-measurement knowledge* respectively.

4.2.1. Pre-measurement knowledge

Finding a solution to the previous problems amounts to finding:

S1. a model of the object under measurement;
S2. a model of the general property;
S3. a model of the measurand.

The last point includes two sub-problems:

S3.1. *a definitional problem*, concerning the existence and the uniqueness of the property, which can be considered as either preexistent to the measurement process or as generated by the process;
S3.2. *a theoretical problem*, concerning the relation of the property intended to be measured with other (influence) properties.

The following assumptions are to be taken into account.

First, the object under measurement has to be modeled as the support of the property while the measurement is executed, and it should be modeled as stable during the process (we are not taking dynamic measurement into account here), stability being the condition that ensures that the information obtained by means of measurement is correctly attributed to the object.

Second, the general property has to be modeled according to a property evaluation type connected with a measurement procedure that is in principle realizable. The attribution of a type to the property should follow from the availability of a procedure for constructing a scale of that type, which would thus allow us to calibrate the measuring system in all circumstances.

Third, the model of the individual property should enable us to solve both the definitional and the theoretical problems. The model should justify the assumption of existence and uniqueness of the measurand at the moment of the measurement execution and should describe how the properties with which the measuring instrument empirically interacts are connected with the measurand.

Furthermore, and most important, the theoretical chain has to be modeled linking the measurand to the set of properties with which the measuring instruments interact. In this respect, a critical difference is that between:

(a) the theoretical laws involved in the *construction of a measuring system* for the measurand: these laws are essential in measuring instruments for which there exists a distinction between the measured property and the indication property of the instrument. In particular, for measuring instruments which operate according to some transduction effect, the process of transduction is modeled in the light of these laws;
(b) the theoretical laws involved in the *computation of a numerical value* for the measurand: these laws are essential in ensuring that the measurand is effectively accessible. As the case of weighing stars demonstrates, it is very possible to be in a condition wherein: (i) the relevant object is identifiable (a star is identifiable), and; (ii) the relevant general property is measurable (mass is measurable), but; (iii) the relevant individual property is not accessible (the mass of a single star is not accessible). In this case, the mass of a star is accessible just because laws of gravitational interaction allow us to link the mass and the motion of a star in a binary system.

Finally, measurement uncertainty sources must be modeled. In general, uncertainty depends on various sources, primarily related to the measurand, the measurement procedure, the environment, and the instruments. A suitable interpretation of how uncertainty originates can be obtained by an analysis of the equation that represents the dependence of the measurand on the properties that are involved in the measurement process (a still open issue is whether the definitional uncertainty should be intended as the minimum threshold of measurement uncertainty or it should be included in the uncertainty budget (JCGM, 2012, 2.33)). Indeed, in typical measurements, the property to be measured is related to the indication provided by an instrument according to the *model equation*,

$$Y = f(X,\boldsymbol{Z}) \tag{1}$$

which models the measurand Y as connected via f to the instrument indication X and a set of other properties $\boldsymbol{Z}$.[8] In particular, in direct methods of measurement the indication X is interpreted as the effect produced by Y in a context characterized by the influence properties $\boldsymbol{Z}$. Thus, a model equation can be viewed as a sort of inversion of the *transduction equation*,

$$X = f^*(Y,\boldsymbol{Z}) \tag{2}$$

which models the causal connection whose output is the indication X, thus represented as the effect of Y given $\boldsymbol{Z}$. In particular, the measurand Y is interpreted here as one of the causes producing X in a context characterized by $\boldsymbol{Z}$.

Hence, in order to model the interaction between the instrument and the property directly measured by the instrument we have to:

1. identify Y, X, and $\boldsymbol{Z}$;
2. identify the interaction between Y, X, and $\boldsymbol{Z}$;
3. model the interactions between Y, X, and $\boldsymbol{Z}$ in terms of the transduction equation;
4. introduce the non-idealities due to incomplete knowledge concerning points 1, 2, and 3;
5. obtain the model equation by inverting the transduction equation.

Uncertainty is connected with this modeling activity in as much as: (i) the introduction of models of entities and properties involves idealization, and so a certain distortion of how the entities and properties actually are; (ii) the introduction of models of coupled systems involves the specification of transduction laws that link properties of different kinds, and so other (influence but not only) properties and interaction laws enter the picture; (iii) the realization of entities determined with respect to certain properties, the standards, involves production, and so a certain distortion of how the entities to be realized ideally are (for more detailed analysis of the components of uncertainty and its dependence on the modeling activity, see Sommer & Siebert, 2006; Giordani & Mari, 2012a).

4.2.2. Post-measurement knowledge

The main problem, after having obtained data from a measurement execution, is that of producing a model of the data and to use it to obtain a measurement result. It is at this stage that the rows of data produced by interpreting the outputs of a measuring instruments are converted into a result that can be then reasonably attributed to the measurand. Two processes are involved here: (i) the inversion of the law-like chain that connects the measurand with the set of properties that are directly measured by the measuring instruments; (ii) the very construction of the model specifying the form of the measurement result and how it is obtained from available data. The latter is typically an estimation of a measured value and the corresponding measurement uncertainty, but other options are possible, such as an interval of values or a probability density/mass function.

In order to get a model of the data and extract the result, well-known statistical procedures are applied to the available data, including the outputs of the measuring instruments. The equations relating the measured properties to the measurand have to be inverted, and all the approximations and idealizations which are involved in the use of such equations have to be taken into account. It is at this stage that the connection between measurement theory and theory of measurement error/uncertainty is located, thus highlighting that the statistical treatment of the data is only a component of the measurement process.

5. Understanding measurement: objectivity and intersubjectivity

The analysis proposed so far shows that measurement is a structured process, and that no black box models can adequately interpret and distinguish it from weaker forms of property value assignments, such as judgment by experience or opinion making. Indeed, the identification of the various elements of the background knowledge is significant in the assessment of two crucial claims: (1) that a certain process is indeed a measurement process; (2) that a measurement process is correctly carried out. Let us then conclude our analysis by highlighting the way in which the foregoing measurement elements are to be considered in the light of the basic features of object-relatedness and subject-independence ("objectivity" and "intersubjectivity" for short) that we assume to be necessary in a measurement.

[8] Note that $\boldsymbol{Z}$ can be further subdivided into a set $\boldsymbol{Z}_I$ of properties whose values and uncertainties are assigned in the current measurement, and a set $\boldsymbol{Z}_E$ of quantities whose values and uncertainties are assigned by external sources.

5.1. Objectivity of the process

Here a measurement process is said to be objective, with respect to an object considered as support of a given property *P*, if it satisfies the following two conditions: (1) it involves an interaction with the object in virtue of *P*, or in virtue of a property connected with *P* according to laws; (2) its results only depend on *P*, or at least they depend only on *P* once other contextual properties are fixed. Thus, in the case of the measurement of the mass of binary stars, the process of measuring is supposed to be objective. In fact: (1) it involves an interaction with a binary star in virtue of a property that is connected according to laws with the mass of the star; (2) the result obtained by running the process is supposed to depend only on the mass of the star once the relevant parameters are fixed. In particular, as we have seen, the process involves the interaction between what is observed through a telescope, a micrometer, and a clock. The quantities thus measured, i.e., the angular distance between the stars and the period of apparent revolution, are connected, via laws, to the distance of the stars and the period of revolution which, in turn, are connected via laws to the mass of the stars. Similarly, in the case of the measurement of the scientific research performance of a researcher, the process of measuring is supposed to be objective. In fact: (1) it involves an interaction with an author in virtue of a property that is connected according to laws with the performance of the author; (2) the result obtained by running the process is supposed to depend only on the performance of the author once the relevant parameters are fixed. In particular, the process involves counting the relevant products, computing the impact factor of the relevant journals, and computing a certain index. The quantity thus measured, i.e., the impact of the global production, is then connected, via laws, to the research performance.

5.2. Intersubjectivity of the results

Here a measurement result is said to be intersubjective if it is invariant with respect to the substitution of the involved subjects. Thus, in the case of the measurement of the mass of binary stars, both the determination of the apparent distance between the stars and the computations that return the desired result are supposed to be independent of the observers and the computers employed thanks to the appropriate calibration of the involved measuring instruments. Similarly, in the case of the measurement of the scientific research performance of a researcher, both the journal impact factor data and the computations that return the desired result are supposed to be public and therefore independent of the observers.

5.3. Reviewing the examples in view of justifying the involved knowledge

Let us review the previous examples to emphasize how the structure of the measurement process that we have proposed supports the solution of the problems of objectivity and intersubjectivity, and at the same time is useful for appreciating some significant differences that emerge in the measurement of a physical property and a non-physical one.

5.3.1. Justification of the knowledge involved in the construction of the general model

The aim of a general model is to provide a link between the objects whose properties we intend to measure and the properties we are able to measure given our set of measuring instruments. At this stage, all the solutions characterizing the activity of pre-measurement modeling are to be provided.

When measuring the mass of binary stars, the general model allows us to link the mass of a celestial body to distances and angles. In this case, the basic element is a set of laws which provide the relevant links, and which are currently acknowledged as sufficiently accurate for our purposes. The problems of identifying the object under measurement, measuring of the general properties involved, and having access to the measurand, are then solved in accordance with such laws. In fact: (1) the objects under measurement are identifiable through optical interactions with instruments; (2) the general property, mass, is known to be measurable according to a ratio scale; (3) the measurand can be accessed due to the theoretical links given by the laws. In addition, the laws that determine the motion and interaction of celestial objects, at different levels of accuracy and ideality, are accepted. In particular, Kepler's laws about the shape of the orbit and the velocity through the orbit are only approximations of what can be obtained by applying Newton's laws, and Newton's classical laws are in turn only approximations of Einstein's relativistic laws. Still, the error deriving from the assumption that a binary system is gravitationally closed and that the orbits of the stars in such a system are ellipses is justly considered negligible if compared with the uncertainty derived from measuring astronomical distances using telescopes and micrometers.

Such laws have two crucial features: (F1) they are accepted as a consequence of our successes in applying them in different circumstances for modeling different systems, where the success of such applications and models is independent of the success we can obtain in the case under scrutiny; (F2) they are selected from a set containing different independent laws that are applicable to the same case, whose application would lead, as is theoretically predictable and operatively confirmable, to similar results, within an acceptable degree of uncertainty. When assessing the research performance of an individual researcher, the laws used for ordering performances given the order on the corresponding products, and for ordering the product given the order on their overall impact, seem to be acceptable. In addition: (1) the objects under measurement are evidently identifiable, being texts; (2) the general property, quality of performance, is hypothesized to be measurable according to an ordinal scale; (3) the measurand can be accessed due to the theoretical links given by the laws.

However, there are some important differences between the two examples here. As just mentioned, physics tells us that mass is structurally connected to many other properties, in a network that guarantees the cross-validation of different measurements of different properties. Were the radical doubt of objectivity put forward – are we really measuring what we intend to measure? – several alternative measurements could be performed to obtain an answer. On the other hand, nothing similar to physics exists to provide a well-grounded connection of research performance to other properties, with the consequence that in this case the general property is more typically defined directly by the measurement procedure itself. Furthermore, the two previously highlighted features are not present in this case: against F1, the employed laws are not accepted in virtue of the success of their application in cases which are independent of the one under scrutiny; against F2, they are not part of a set containing different independent laws that are applicable to the same cases. This difference has been further elaborated by Finkelstein in terms of strongly vs weakly defined measurement (Finkelstein, 2003).

5.3.2. Justification of the knowledge involved in the construction of the specific model

The object under measurement is modeled according to the theories that better fit the trade-off between accuracy and idealization. Hence, when measuring the mass of binary stars, the model is based on the idealization of stars, considered as point masses, and on an agreed concept of star mass. When assessing the research performance of an individual researcher, the object under measurement is simply modeled as a set of texts, while the procedure for determining the impact is based on the hypothesis that impact factor of journals conveys the appropriate information for achieving the proposed aim, i.e., for making the impact accessible.

Again, there are some important differences between the two examples. The model of the object under measurement, in the first case, can be progressively refined, thus obtaining a set of mutually compatible, non-competing models ordered according to their degree of accuracy in view of the idealizations involved in the construction of the models themselves. For example, a model of binary stars as point masses subjected only to the corresponding gravitational interactions can be refined by removing some approximations. By contrast, different models of scientific production can lead to significant differences in the results. For example, admitting monographs as evidence of scientific performance can change in a drastic way the impact of authors' production (Henrekson & Waldenström, 2011). They are then incompatible, competing models: taking them together does not generally improve the knowledge of the measurand.

5.3.3. Justification of the knowledge involved in the construction of the measurement model

The two examples are critically different on this matter, even if, in both cases, the distinction between the measuring instrument and the object under measurement is unproblematic, and the means of calibrating the measuring instrument is known.

When measuring the mass of a star, the method of measurement is selected among a class of consistent non-competing methods. So, the applicability of the selected measurement procedure can be tested against the application of other independent methods, like spectroscopic analysis (Carroll & Ostlie, 2007, ch.7).

When assessing the research performance of an individual researcher, the method of measurement is selected among a class of inconsistent competing methods. In particular, among the methods in use for measuring scientific performances, four general approaches can be followed. A measurement model can be based on weighted journal publications, as in the IF method, or on citations of most cited works, or on the number of publications, or on a suitable combination of the previous criteria (admittedly, these can be also intended as different ways to define the measurand). An analysis of the outcomes provided by these method shows that the resulting rankings are strongly inconsistent (for a striking example, see Henrekson & Waldenström, 2011), thus indicating that the attribution of the result is essentially dependent on the method which is used.

5.3.4. Justification of the knowledge involved in the measurement execution

Measurement requires the empirical interaction of the object under measurement and a calibrated measuring instrument; an operation that is performed in a context that generally influences the outcomes of both the interaction and the calibration, thus reducing the expected objectivity of the process. As much is well known, hence sophisticated instrumentation techniques have been developed with the aim of reducing such effects.

5.4. Reviewing the examples in view of their being instances of measurement

As we have seen, the two examples we have proposed have the same structure, but differ as to the justification of their assumptions. In particular, the procedure involved in measuring research performance exhibits the following deficiencies:

1. as to the general model, the general property of research performance is not structurally connected in a network that allows us to check measurement results via cross-validation, and the employed relations are accepted neither by virtue of a previous success nor by virtue of their fitting in a relevant theoretical framework;
2. as to the specific model, different models of the same phenomenon can be constructed that are neither comparable on a common theoretical or mathematical basis nor consistent with respect to the results that can be obtained by employing them;
3. as to the measurement model, different methods of measurement can be selected among a class of inconsistent competing methods, whose success cannot be assessed by means of independent measurement procedures.

These issues prompt consideration as to whether this example is properly a case of measurement. On the other hand, the concept 'measurement' is not uniquely defined, and tentatively encompassing definitions – such as the one of the VIM, measurement being a "process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity" (JCGM, 2012, 2.1) – pay the price of a generic characterization. In response to three stereotypical clusters of interpretations, we have adopted the standpoint that measurement is a source of dependable information not because we know *that* we can rely on the information it produces (the concept 'low quality measurement' is perfectly admissible), but because we know *how much* we can rely on it. In this perspective, the foregoing analysis shows that the key feature of a measurement is the structure of the process. To be sure, the adoption of certain general models, specific models, and measurement models – i.e., the contents of such a structure – has to be *justified* relative to both the objectivity of the measurement process they make realizable, and the intersubjectivity of the results they make obtainable. Thus, while the procedure adopted for assessing research performance exhibits the *right structure* for being a candidate measurement procedure, it might be considered as still not having the *right level of justification* for being designated such. This pluralistic vision of measurement not only acknowledges the context-dependence of the evaluation of reliability of measurement results, but also takes the evolutionary development of measurement processes into account.

6. Conclusions

The practice of developing measurements for increasingly complex objects is fostering the view in which empirical and informational components are intertwined and driven by target: designing and then performing a measurement process involves goal-setting, theoretical assumptions, modeling, experiments, calculation, information interpretation and decision, all of them mutually related in a feedback structure. Measurement is a goal-driven process, and as such the harmonization of specified target and the resources used is a crucial component for characterizing that which makes a "good" measurement.

The analysis that we have proposed here provides an epistemological basis for a conceptual framework in which the main tasks expected in a measurement and their inter-relations of such tasks are identified (we have more extensively presented the structural and procedural aspects of such a framework in Petri, Mari, & Carbone, 2015). While the naive black box atomic models of measurement might remain pragmatically acceptable whenever the required quality of results is much lower than the capability of the adopted methods and instruments, our analysis has highlighted the theory-ladenness of measurement, at the same time showing that the public trust socially attributed to measurement has structural reasons, for which producing quantitative data or being a morphic representation are at most necessary, but definitely not sufficient, conditions.

This is in explicit opposition with the thesis that Feyerabend effectively presented as: "the events, procedures and results that constitute the sciences have no common structure" (Feyerabend, 1993 [1975], Introduction to the Chinese Edition. p. 1). We claim that it is instead exactly the structure of the process that distinguishes measurement from representational processes such as opinion making.

Such a framework aims at achieving similar goals as technical standards (UNIDO, 2006): to guarantee fitness for purpose of the measurement processes that need to be designed and implemented; to favor the development of compatible activities among different fields when performing measurement; to guard against misinterpretation of concepts; to allow improvements in how efficiently measurement is performed; to contribute to better communication and understanding of measurement outcomes, and; to help remove barriers to abstracting measurement across different fields and domains of application.

Acknowledgements

The referees and the editors of the Special Issue supported us excellently in the improvement of this paper. We thank them warmly.

References

Aitken, R. J. (1964). *The binary stars*. New York: Dover.
Bentley, J. P. (2005). *Principles of measurement systems*. Harlow: Pearson.
Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan.
Campbell, N. R. (1957). *Foundations of science*. New York: Dover (Reprinted edition of: Physics – The elements, 1920)
Carroll, B. W., & Ostlie, D. A. (2007). *An introduction to modern astrophysics*. San Francisco: Addison-Wesley.
Dingle, H. (1950). A theory of measurement. *British Journal of the Philosophy of Science, 1*, 5-26.
Doignon, J. P. (1993). Geometry is measurement but measurement theory is more than geometry. *Journal of Mathematical Psychology, 37*, 472-476.
Duncan, O. D. (1984). *Notes on social measurement – Historical and critical*. Russell Sage Foundation.
Feyerabend, P. (1993). *Against method – Outline of an anarchistic theory of knowledge* (3rd ed.). London: New Left Books (Originally published, 1975)
Finkelstein, L. (1975). Fundamental concepts of measurement: Definitions and scales. *Measurement and Control, 8*, 105-111.
Finkelstein, L. (2003). Widely, strongly and weakly defined measurement. *Measurement, 34*, 39-48.
Finkelstein, L. (2009). Widely-defined measurement – An analysis of challenges. *Measurement, 42*, 1270-1277.
Fitzpatrick, R. (2008). Euclid's Elements of geometry. the Greek text of J.L. Heiberg (1883–1885) edited, and provided with a modern English translation http://farside.ph.utexas.edu/Books/Euclid/Euclid.html.
Frigerio, A., Giordani, A., & Mari, L. (2010). Outline of a general model of measurement. *Synthese, 175*(2), 123-149.
Giordani, A., & Mari, L. (2012a). Measurement, models, and uncertainty. *IEEE Transactions on Instrumentation and Measurement, 61*(8), 2144-2152.
Giordani, A., & Mari, L. (2012b). Property evaluation types. *Measurement, 45*, 437-452.
Henrekson, M., & Waldenström, D. (2011). How should research performance be measured? A study of Swedish economists. *The Manchester School, 79*, 1139-1156.
Hölder, O. (1996). The axioms of quantity and the theory of measurement (Translation of *Die Axiome der Quantität und die Lehre vom Mass*, 1901, by J. Michell and C. Ernst). *Journal of Mathematical Psychology, 40*, 235-252.
JCGM 100:2008. (2008). Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM). Joint Committee for Guides in Metrology (originally published, 1993) http://www.bipm.org/en/publications/guides/gum.html
JCGM 200:2012. (2012). International vocabulary of metrology – Basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology (2008 editions with minor corrections) http://www.bipm.org/en/publications/guides/vim.html
Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A.. (1971, 1989, 1990). Foundations of measurement, vols. 1–3. New York: Academic Press.
Kyburg, H. E., Jr. (1984). *Theory and measurement*. Cambridge: Cambridge University Press.
Mari, L. (2013). A quest for the definition of measurement. *Measurement, 46*, 2889-2895.
Mari, L., Carbone, P., & Petri, D. (2012). Measurement fundamentals: A pragmatic view. *IEEE Transactions on Instrumentation and Measurement, 61*(8), 2107-2115.
Mari, L., Carbone, P., & Petri, D. (2015). Fundamentals of hard and soft measurement: A pragmatic and structural perspective. In A. Ferrero, D. Petri, P. Carbone, & M. Catelani (Eds.), *Modern measurements: Fundamentals and applications* (pp. 203-262). Hoboken: Wiley.
Michell, J. (2005). The logic of measurement: A realist overview. *Measurement, 38*, 285-294.
Mundy, B. (1987). The metaphysics of quantity. *Philosophical Studies, 51*, 29-54.
Narens, L. (1985). *Abstract measurement theory*. Cambridge: MIT Press.
Narens, L. (2002). *Theories of meaningfulness*. New York: Routledge.
Petri, D., Mari, L., & Carbone, P. (2015). A structured methodology for measurement development. *IEEE Transactions on Instrumentation and Measurement, 64*(9), 2367-2379.
Roberts, F. S. (1979). *Measurement theory – With applications to decision-making, utility and social sciences*. London: Addison-Wesley.
Sommer, K. D., & Siebert, B. R. (2006). Systematic approach to the modeling of measurements for uncertainty evaluation. *Metrologia, 43*, 200-210.
Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680.
Suppes, P. (1954). A set of independent axioms for extensive quantities. *Portugaliae Mathematica, 10*(4), 163-172.
Suppes, P. (2002). *Representation and invariance of scientific structures*. Stanford: CSLI Publications.
Teller, P. (2013). The concept of measurement-precision. *Synthese, 190*, 189-202.
UNIDO. (2006). *Role of standards – A guide for small and medium enterprises*. United Nations – Industrial Development Organization. https://www.unido.org/fileadmin/media/documents/pdf/tcb_role_standards.pdf
Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford: Chandos Publishing.