

PAPER • OPEN ACCESS

A meta-structural understanding of measurement

To cite this article: Luca Mari *et al* 2016 *J. Phys.: Conf. Ser.* **772** 012009

View the [article online](#) for updates and enhancements.

Related content

- [Accordion-like metamaterials with tunable ultra-wide low-frequency band gaps](#)
A O Krushynska, A Amendola, F Bosia *et al.*

Recent citations

- [Leslie Pendrill](#)
- [The Psychometric Modeling of Scientific Reasoning: a Review and Recommendations for Future Avenues](#)
Peter A. Edelsbrunner and Fabian Dablander
- [Data generation methods across the empirical sciences: differences in the study phenomena's accessibility and the processes of data encoding](#)
Jana Uher



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

A meta-structural understanding of measurement

Luca Mari¹, Andrew Maul^{21#}, David Torres Irribarra³, Mark Wilson⁴

¹ School of Engineering, Università Cattaneo – LIUC, Castellanza (VA), Italy

² Graduate School of Education, University of California, Santa Barbara, CA, USA

³ MIDE, Pontificia Universidad Católica de Chile

⁴ Graduate School of Education, University of California, Berkeley, CA, USA

E-mail: amaul@education.ucsb.edu

Abstract. It is not always clear to what extent the logic and vocabulary of measurement as used in different scientific disciplines are mutually coherent, nor how measurement can be demarcated from, say, opinion. In recent decades there have been a number of attempts to provide necessary and/or sufficient sets of conditions for when measurement is achieved, usually in terms either of inputs (e.g., whether an evaluated property is a quantity), or outputs (e.g., whether a procedure assigns numbers according to a rule). We argue instead that the public trust attributed to measurement is best justified in terms of the structural features of the process rather than of its inputs or outputs.

1. Introduction

In both scientific and lay discourse, the concept of measurement is commonly associated with precision, accuracy, and trustworthiness, yet it remains remarkably difficult to provide a fully satisfactory definition of the concept, as the vast array of proposed definitions witnesses (e.g., [1]). At least two basic reasons concur to explain this difficulty. First, as the scope of the concept of measurement has broadened, it is not always obvious what – if indeed anything – is common among all the processes claimed to be measurements. Second, the concept of measurement has come to be characterized according to purely formal criteria, thus abstracting from the concrete realization of the process, due in part to the fact the evaluation of non-physical properties cannot conform to the traditional structure of measuring systems operating on the basis of the transduction implemented by a physical sensor. As a consequence, theoretical interpretations of measurement have become so abstract that they may be unable to provide a convincing and useful demarcation of measurement from formally similar processes that are generally thought to lack epistemic authority, such as most instances of the expression of subjective judgments and opinions.

One may question whether working on the definition of ‘measurement’ is a worthwhile endeavor. Our position on this matter is practical: there is a social interest in sharing scientific and technical vocabulary across disciplines, particularly in the case of an infrastructural activity like measurement [2], and there is a social acknowledgment of the epistemic authority of measurement, which has

¹ # To whom any correspondence should be addressed.

Authors’ names are ordered alphabetically. One of the authors is a member of the Joint Committee on Guides in Metrology (JCGM) Working Group 2 (VIM). The opinion expressed in this paper does not necessarily represent the view of this Working Group.



critical consequences in particular in terms of public trust attributed to the outcomes of putative measurement processes and the resources devoted to such processes.

In synthesis, the position we propose here is that measurement has to be characterized as a process by its operative structure, not by the specification of the functional relationship connecting its inputs to its outputs. As a corollary, our claim is that any purely black-box (meta-)model cannot adequately account for fundamental features of measurement.

2. Defining measurement

Under the general hypothesis that measurement is a process that operates on inputs (at least the measurand, in the case of direct measurement methods) to produce outputs (at least the measurement result), the idea is that measurement is an instance of the meta-model:

$$\text{input} \rightarrow [\text{transformation}] \rightarrow \text{output}$$

specified as:

$$\text{input} \rightarrow [\text{evaluation}] \rightarrow \text{output}$$

given that measurement produces values of properties (instead of, e.g., physical transitions or objects). The acknowledgment that not all evaluations are measurement calls for a further specialization, and thus the characterization of what measurement is, obtained by differently focusing on one or more of these three entities [3, 4], and in particular claiming that the necessary and/or sufficient conditions for measurement can be located in one or more of these entities.

A first category of views, such as those given by Bridgman [5] and Dingle [6], could be characterized as focusing on the outputs of the process. Dingle, for example, defines measurement as “any precisely specified operation that yields a number,” thus referring to an aspect of the output of the process (i.e., that it is numerical) as a necessary and sufficient condition for measurement. In the absence of an account of what forms of precision are necessary, the requirement that the procedure be “precisely specified” seems to demand sufficient clarity regarding what the operations of the procedure are, but places no demands concerning the inputs of the procedure nor the manner in which they are transformed into outputs.

A second category of views focuses instead, or sometimes additionally, on the inputs of the process. For example, Michell (e.g., [7]), in line with the Euclidean tradition, defines ‘measurement’ as the assessment of a quantity, the measurand, in comparison to a second quantity, the unit. According to this view, whether or not a given property is a quantity is an ontological issue, and is a pre-condition for measurement. In this view, a property that is not structured as a quantity is a priori not measurable: thus a feature of the input, namely that it is a quantity, is a necessary condition for measurement. Arguably, Michell’s view also requires that the output of the process be numerical – specifically, that it be expressed as a ratio of a quantity value to a standard unit.

A third category of views characterizes measurement in terms of the formal relationship between inputs and outputs, but is silent regarding the operative structure of the transformation by which the relationship is implemented as a causal mechanism that induces variations in the outputs from variations in the inputs. Representational theories of measurement (e.g., [8]), for example, define ‘measurement’ in terms of a morphism between an empirical relational system and a numerical relational system; thus, in this view, a particular feature of the formal dependence of the output on the input is a (necessary and) sufficient condition for measurement.

As we have argued elsewhere (e.g., [1], [9], [10]), while each of the views described above succeeds in capturing valuable intuitions about measurement, none of them provides a fully satisfactory set of conditions that could universally be used (a) to distinguish measurement from non-measurement processes, nor (b) to distinguish better (or more trustworthy, useful, etc.) from worse instances of measurement. Views in the first category tend to fail to disallow instances of rule-based numerical assignment that have no epistemic value, such as procedures based on subjective judgment such as (formalized) guesses or statements of opinion, or precisely-specified but arbitrary rules. Views in the second category tend to be too restrictive, insofar as they would disallow many widely accepted cases of measurement in the physical sciences in addition to, potentially, all cases in the social

sciences, regardless of epistemic or pragmatic value. Finally, views in the third category tend to be incapable of distinguishing processes that generate dependable knowledge from processes that do not but are otherwise functionally similar evaluations.

3. Rethinking the definitional problem

As it has been hinted at in the previous section, debates concerning the definition of measurement may conflate two different questions: (a) how should measurement be defined, as distinct from other types of evaluations? and (b) how can we judge the dependability of the information provided by an evaluation? In principle the two questions could be considered redundant if measurement were defined as any evaluation that yielded dependable knowledge, but this would broaden the scope of the concept of measurement to potentially include essentially all forms of inquiry, and is *prima facie* inconsistent with every formal and lay conception of measurement of which we are aware – for example, it is commonly accepted that not all measurements guarantee the same high dependability, and not all opinions are flawed by the same low dependability. Hence, dependability of empirical information is definitely a worthwhile target, but the problem is how such dependability can be obtained and assessed as a stable and publicly justifiable feature of a given process.

Particularly when the required quality of the results is not so high (relatively low resolution, sensitivity, precision, etc.), an experienced, skilled individual may express her/his opinions in such a way that they would be accepted as dependable (and discovered to be morphically related to some set of empirical relations). But for an evaluation to be able to produce results that can be publicly trusted, and thus considered a measurement, the process must be at least in principle replicable, and even a successful result is not considered the outcome of a scientific or technological process if the way in which it has been obtained is not at least in principle publicly explainable.

As we have argued elsewhere [1, 3, 4, 10, 11], such a public trust is primarily grounded in the demonstrated capability of providing sufficient object-relatedness (“objectivity”) and subject-independence (“intersubjectivity”) of the information produced by those processes that are traditionally structured as measurements (and are traditionally called “measurements”). The extent to which a given process yields results with these epistemic virtues, in turn, depends on its operative structure, which includes the design, setup, and implementation of the replicable transformation of inputs into outputs, and therefore unavoidably on the structure of the process, which must be then construed according to an open-box (meta-)model. *Ceteris paribus*, the better researchers understand the causal chain leading from inputs to outputs, and thereby understand the conditions under which one could expect this causal chain to remain invariant, and the better such understanding is implemented in the actual process, the more trust can be placed in the objectivity of the results. And, *ceteris paribus* again, the better researchers understand the standards and the processes by which they can be compared with each other in a traceability chain, and thereby understand the conditions under which one could expect these standards to remain stable, the more trust can be placed in the results as for their intersubjectivity.

In context, the level of specificity in this understanding (and thereby, the degree of warranted trust) will inevitably depend on the context and purpose for which the measurement results are to be used, and, of course, on historical evidence that it has been found to be dependable in the past.

4. Two examples

Measurement instruments are designed for a wide variety of purposes, and even different instruments designed to measure the same property may take on different forms depending on the purpose² for which the results are to be used and the context in which the instrument will operate. For example, the

² For the case of physical properties and their measurement, this dependence is often related to the expected quality (precision, resolution, etc.) of the results, so that different instruments take on different form depending on the quality of the results demanded by a specific purpose.

property programming proficiency could be assessed for a number of distinct purposes, among them (a) the selection of candidates for a position in a professional setting, and (b) the diagnosis of specific patterns of misunderstandings for educational purposes.

In the first instance, the emphasis may be on deciding whether the candidate is able to successfully undertake the duties entailed by the position; as such the instrument could consist of a simulation of these duties, and an evaluation of the performance of the candidate on each of the relevant tasks. In this instance it may be that the relevant information is whether (or the degree to which) candidates are able to successfully execute the duties of the position, independently of the reasons they are able to do so – and it may be that there are different combinations of knowledge, skills, abilities, and other individual characteristics that would enable a candidate to be successful. The final evaluation could take any number of forms depending on the context: it could be a dichotomous classification (qualified / not qualified), with measurement uncertainty expressed in terms of the probability of an accurate classification; alternatively, it could take the form of rank-ordering all candidates from most qualified to least qualified, especially if some predetermined number of candidates were to be selected and the goal was to identify the most qualified individuals; as yet another possibility, each candidate might be given a score reflecting their overall level of proficiency, where interpretation of the score may be aided by reference to population norms and/or specific sets of criteria associated with different scores [12].

In the second instance, given the desire for utility of the test for pedagogical purposes, a much greater emphasis might be placed on understanding the specific elements of knowledge and skill that go into successful or unsuccessful completion of specific tasks; this in turn could lead to different sorts of tasks being designed and included in the assessment. In contrast to the first instance, which left the specific reasons for success or lack thereof relatively unexamined, this sort of test would likely foreground the need for a theory-based explanation of the processes leading to successful task completion at the level of the individual – that is, domain-specific cognitive theory. Again, the results of the assessment could take several forms depending on the context and the preferences of the stakeholders involved, ranging from categorical classification of individuals on a number of specific forms of knowledge and skill, to a continuous estimate of the individual's overall level of proficiency, aided by a theory-grounded interpretative device such as a 'construct map' [12].

In terms of the meta-model mentioned in section 2, these two scenarios plausibly have the same input—that is, programming proficiency of the individual.³ The manner in which programming proficiency is transformed into the outputs of the test are different, as shaped by the selection of specific tasks, task features, and contextual considerations, but plausibly involve many of the same cognitive, motivational, and affective processes at the level of the individual. The outputs are distinct, and need not be quantitative or even numerical in either case—though they could be, if this was judged by the relevant stakeholders to be the most useful manner in which to present and communicate information about programming proficiency.

As discussed in the previous sections, at least two questions could be asked of each of these instances: (a) does the evaluation qualify as measurement? and (b) does the evaluation yield results that are dependable, for their intended purposes and in their intended contexts? The second question is clearly of greater interest to the relevant stakeholders than the first, and in turn depends, as previously noted, on the degree to which the design, setup, and implementation of the replicable transformation of inputs into outputs is well-understood and implemented.

As a second example, the assessment of wind speed in the context of sailing could either be by using a cup anemometer, a sonic anemometer or observation using the Beaufort Scale. The cup anemometer is “the most basic type of anemometer [which] consists of a series of cups mounted at the end of arms that rotate in the wind,⁴” while the sonic anemometer, a more recently-developed tool in

³ Alternatively, it could be argued that the inputs are distinct, given that the pragmatic concerns motivating the creation and use of the instrument also serve to shape the definition of the target attribute itself [9].

⁴ Cup anemometer. (n.d.) The American Heritage® Dictionary of Student Science, Second Edition. (2014).

the measurement of wind speed, has no moving parts and “measures the projection of the wind velocity vector on the acoustic path” [13]. Finally, the Beaufort Scale relies on a combination of observable environmental indicators plus the judgement of a human observer to produce an estimate of wind speed. We can use any of these three to answer a question about wind speed in order to set the sails accordingly.

The selection of one of these methods involves changes in the first three of the four components of the model introduced here: that is, changes in the process structure, the instrument, or the procedure. To begin with, these methods rely on different underlying theories—a sonic anemometer arguably relies on acoustics theory while the cup anemometer and the Beaufort Scale do not—tracing different causal narratives justifying each procedure. Of course, the instrument in each case is changed, varying from moving cups in one case, to a static device in the sonic anemometer, up to the use of environmental cues integrated by a human observer. It is also worth noting that the use of anemometers versus observations based on the Beaufort scale also changes the structure of the outcome, in one case modelling it as an ordinal attribute and in the other modelling it as quantitative. Similarly, each of these methods of assessing wind speed will require different procedures in order to obtain measurement results. Despite their differences, all these approaches can potentially be replicable, reliable and explainable measurements that allow us to achieve the goals defined by the context. Interestingly, varying the context could make some of these procedures more or less adequate, as a different context could make some approaches incompatible with the goals of that measurement need; for instance, if a context of application is that one that requires the use of wind speed as a quantity to be incorporated into an equation, then that would be a poor match for the results produced through the use of the Beaufort Scale. This variety of possible measurement approaches, which involve alternatives in the process structure, instrument, and procedure can be found both in the physical and the social sciences.

5. Concluding thoughts

Measurement can be characterized as a structured process, designed to be performed by given devices according to given specifications to produce information in a given context [11]. This results in a meta-structural understanding of measurement processes composed of the (1) process structure, (2) instrument, (3) procedure, and (4) context. This paper has not aimed to provide a conclusive demonstration of the thesis that this meta-structural understanding of measurement is superior to what we have referred to as the “black box” (meta-)models of measurement implied by some other perspectives, though we hope that we have laid the groundwork for an alternative way of thinking about measurement and measurement quality. Further scholarship will expand on this argument, in the hopes that by focusing on the extent to which structural features of the process yield results with epistemic value in given contexts, rather than on prescriptive stances regarding what should or should not count as measurement, coherent dialogues can take place across disparate areas of inquiry, while still respecting the fact that measurement results play a privileged role as a source of publicly accepted knowledge.

References

- [1] L. Mari, A quest for the definition of measurement, *Measurement*, 46, 2889-2895, 2013.
- [2] JGCM, International Vocabulary of Metrology (VIM) – Basic and General Concepts and Associated Terms (2008 edition with minor corrections), 2012.
- [3] A. Frigerio, A. Giordani, L. Mari, Outline of a general model of measurement, *Synthese*, 175, 123–149, 2010.
- [4] L. Mari, P. Carbone, D. Petri, Measurement fundamentals: A pragmatic view, *IEEE Trans. Instr. Meas.*, 61, 8, 2107–2115, 2012.
- [5] P. Bridgman, *The logic of modern physics*. Macmillan: New York, 1920.
- [6] H. Dingle, A theory of measurement, *The British Journal for the Philosophy of Science*, 1, 5-26, 1950.

- [7] J. Michell, The logic of measurement: A realist overview, *Measurement*, 38, 285–294, 2005.
- [8] D. Krantz, R. Luce, P. Suppes, A. Tversky, *Foundations of measurement*, vol. 1: Additive and polynomial representations. New York: Academic Press, 1971.
- [9] A. Maul, D. Torres Irribarra, M. Wilson, The philosophical foundations of psychological measurement, *Measurement*, 79, 311–320, 2016.
- [10] L. Mari, A. Maul, D. Torres Irribarra, M. Wilson, Quantification is neither necessary nor sufficient for measurement, *Journal of Physics: Conference Series*, 459, 2013.
- [11] L. Mari, M. Wilson, A structural framework across strongly and weakly defined measurements, *proc. IEEE I2MTC 2015*, 1522–1526.
- [12] M. Wilson, L. Mari, A. Maul, D. Torres Irribarra, A comparison of measurement concepts across physical science and social science domains: instrument design, calibration, and measurement, *Journal of Physics: Conference Series*, 588, 2015.
- [13] J.C. Wyngaard, Cup, Propeller, Vane, and Sonic Anemometers in Turbulence Research, *Annual Review of Fluid Mechanics*, 13, 399–423, 1981