



HAL
open science

On the meaning of measurement uncertainty

Fabien Grégis

► **To cite this version:**

Fabien Grégis. On the meaning of measurement uncertainty. *Measurement - Journal of the International Measurement Confederation (IMEKO)*, 2019, pp. 41-46. 10.1016/j.measurement.2018.09.073 . halshs-01884155

HAL Id: halshs-01884155

<https://shs.hal.science/halshs-01884155v1>

Submitted on 10 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the meaning of measurement uncertainty

Fabien GRÉGIS

Univ. Paris Diderot / Univ. Tel Aviv

fabien.gregis@u-paris.fr

Post-doc fellow at the Cohn Institute, Tel Aviv University

Research associate at SPHERE, University Paris Diderot

Preprint version

Measurement **133**:41–46, 2019

DOI: [10.1016/j.measurement.2018.09.073](https://doi.org/10.1016/j.measurement.2018.09.073)

Abstract. This article discusses the definitions of “measurement uncertainty” given in the three successive editions of the International Vocabulary of Metrology (VIM) as well as a proposition that was recently suggested for the next edition of this document. It is argued that none of the definitions is satisfying. First, a thorough definition of measurement uncertainty should supply an explanation about the meaning of the concept, which is missing from the VIM2&3. Secondly, when the definitions do indeed try to delineate the meaning of measurement uncertainty, as is the case in the VIM1 and the possible next edition, the meanings provided are not accurate enough: the VIM1 version is flawed and the possible future definition appeals to an overly restrictive notion of “belief”. Alternative options are then proposed, based on the the observation that measurement uncertainty is a statement made by the experimenter about the measurand, that can only be inferred from what is accessible to one’s knowledge, and therefore rests on the hypothesis that no systematic error affects the measurement.

Keywords. Measurement uncertainty; Vocabulary of metrology; Epistemology of measurement

Contents

1	Introduction	3
2	VIM2 and VIM3 definitions avoid dealing with the meaning of measurement uncertainty	5
3	The VIM1 definition of measurement uncertainty is flawed	7
4	The suggested VIM4 definition of measurement uncertainty appeals to an overly restrictive notion of “belief”	10
5	Directions for improvement	13
6	Conclusion	16

1 Introduction

Measurement uncertainty is a key concept of metrology and scientific practice. Since the end of the 1970s and the recommendation INC-1 [13], lots of efforts have been invested in metrology in order to standardize the practice of uncertainty analysis and to reach an international consensus about the quantitative determination of uncertainties. This process led to the publication of the “GUM family of documents” (term borrowed from [35]), which comprises a series of guides, the *Guide to the expression of uncertainty in measurement* (GUM) [19] and its supplements [20, 21, 22, 23], in close relation with the *International Vocabulary of Metrology* (VIM) [30, 31, 24], another metrology guide setting up the definitions of critical terms of measurement.

These recent developments were instrumental in clarifying a number of frequently debated issues, among which the difference between measurement error and measurement uncertainty [10]. However, judging by the liveliness of the debates regarding measurement uncertainty today, whether inside or outside the metrological community, it appears that the adequate definition of the concept remains challenging. Despite documents like the GUM and the VIM, a lot of users (including science students [6, 5]) struggle to grasp the meaning of the concept and to appreciate its importance in experimental science. This article aims to challenge the definitions that were laid in the successive editions of the VIM. I argue that none of them is entirely satisfying, being incomplete and partly inaccurate, and I suggest an alternative orientation for the revision of these definitions.

The definition of measurement uncertainty in the *International Vocabulary of Metrology* (VIM) has changed significantly since the first edition in 1984. The successive definitions are listed in table 1 below, including a possible new definition suggested by the JCGM working group in charge of the conception of a future fourth edition of the document¹. None of these definitions is satisfying. First, the VIM2 and VIM3 definitions are incomplete. They merely point to what uncertainty purports to describe – a dispersion of values – while I claim that a thorough definition of measurement uncertainty should supply an explanation about the meaning of the concept. Secondly, when the definitions do indeed try to delineate the meaning of measurement uncertainty, as is the case in the VIM1 and the VIM4(?), the meanings provided are not accurate enough. The next sections will be first devoted to demonstrate these two shortcomings. I will then suggest a direction for improvement.

Section 1 gives a general overview of the problem by comparing the definitions provided by the different editions of the VIM. Section 2 argues that the VIM2 and VIM3 definitions are incomplete because they avoid discussing the meaning of measurement uncertainty. Section 3 argues that the VIM1 definition is flawed, and that it cannot easily be improved by acknowledging the probabilistic nature of measurement uncertainty.

¹Because of the provisional status of the latter definition, I will thereafter designate it as the “VIM4(?)” definition, as is done in [9].

VIM1 (1984)	Uncertainty of measurement: an estimate characterizing the range of values within which the true value of a measurand lies.
------------------------	--

VIM2 (1993)	Uncertainty of measurement: parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could be reasonably attributed to the measurand.
------------------------	--

VIM3 (2008)	Measurement uncertainty: non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.
------------------------	---

VIM4(?)	Measurement uncertainty: parameter (or parameters) characterizing how well the (essentially unique) true value of the measurand is believed to be known
----------------	--

Table 1: Definitions of “measurement uncertainty” in the three editions of the *International Vocabulary of Measurement* (VIM) [30] [31] [24] and a possible new definition for the future [9].

Section 4 argues that the definition suggested for the next edition of the VIM, while making a sensible point, remains too narrow. Finally, in section 5, I make a suggestion for improving the definitions criticized previously.

2 VIM2 and VIM3 definitions avoid dealing with the meaning of measurement uncertainty

The VIM2 and VIM3 consider measurement as a process of *attribution* of values to the measurand [24, p.16]. Accordingly, their definitions of measurement uncertainty focus on descriptive features that are relevant in the process of attribution. The main difference between the two definitions resides in that the former uses conditional (“values that *could* be attributed”) while the latter uses infinitive (“values *being* attributed”). This is because measurement uncertainty is not considered by the VIM2 as part of the result itself. In the VIM2, the measurement result is the central value (the “value attributed to a measurand, obtained by measurement” [31, p.23]), and measurement uncertainty is then “associated with the result” in order to characterize it [31, p.23]. On the contrary, as explained in [9, p.S150], the VIM3 definition makes measurement uncertainty an integral part of the measurement result: since a “measurement result” is a “set of quantity values being attributed to a measurand (...)” [24, p.19], and since measurement uncertainty is the “dispersion” of this set of values, then measurement uncertainty is simply a descriptive parameter of the measurement result.

If we leave out this difference for the moment, it appears that both definitions are arguably appropriate regarding what they intent to focus on. They have the merit to clarify that measurement uncertainty is not a *property* of the measurand or of the measurement. Indeed, it is a *statement*, made by the experimenters *about* the result, with regard to their theoretical and experimental knowledge. In that sense, measurement uncertainty differs in nature from “measurement error” (or “measurement accuracy”) which are (unknown) objective features of the measurement result² (respectively the measurement process), and are independent from the knowledge of the experimenters. The VIM2 and 3 succeed in capturing this aspect of measurement uncertainty (contrarily to the VIM1 definition). The VIM3 statement could very well remain at the basis of a revised definition for measurement uncertainty.

However the VIM2 and VIM3 definitions leave the users without any clue about the meaning of the concept. None of them discloses how the attributed values are to be interpreted. Borrowing from [9], they do not even tell us *about what* one is uncertain about.

This predicament might very well be intentional on the part of metrologists, in order to avoid controversy on the meaning of the concept. The VIM2&3 definitions somehow play a role of placeholders: they merely state the existence of a concept, whose meaning is yet to be stipulated. Such a choice allows for the existence of different frameworks for the analysis and interpretation of measurement uncertainty. This could be fitting considering the context in which the VIM2&3 were composed: intense debates have emerged within the community of metrologists during these past decades with regard to the nature of the probabilities and the statistical methods to be used in the context of

²According to [9], the JCGM even tends to consider now that “measurement error should be defined as a quantity, which itself has a value” (p.S152).

measurement uncertainty evaluation [8].

Indeed, the traditional commitment to a frequentist interpretation of probabilities – where probabilities are long-term relative frequencies of occurrence of events – has been seriously challenged since the beginning of the 1970s. Instead, many articles of the contemporary literature in metrology advocate an alternative interpretation of probability, often called “epistemic”, in which probabilities describes rational degrees of belief³. This movement is especially visible in the *Guide for the expression of uncertainty in measurement* [19] (1993) and its first supplement [20]. Epistemic probabilities were first introduced as an answer to a specific shortcoming of the traditional frequentist approach, namely its inability to provide a probabilistic treatment of “systematic errors”. An increasing number of metrologists, statisticians and practitioners now advocate a fully Bayesian account of measurement as the best framework for the conceptualization of measurement uncertainty – which is part of the rationale for the introduction of the VIM4(?) definition [9].

Each of these frameworks, frequentist and Bayesian, suggests a different way to formalize, calculate, and interpret measurement uncertainty. This makes the latter’s definition even more challenging. The direction taken in the VIM2&3 may thus be understood as a neutral standpoint avoiding to take side in this controversy, being indifferent to the formalism used. Such a position might seem reasonable; but in that case, the VIM2&3 simply do not define the concept of measurement uncertainty.

Such a *status quo* is arguably unsatisfactory and, as a matter of fact, the VIM2 and VIM3 definitions contrast with both VIM1 and VIM4(?) definitions, which do tell something about the meaning of measurement uncertainty by introducing the so-called “true value” of the measurand⁴. I claim, however, that these definitions are not fully satisfying either. I will start by arguing that the VIM1 definition is incorrect.

³See [16] for a philosophical introduction to the concept and [4] for an explanation of its application to metrology. The *Guide to the expression of uncertainty in measurement* devotes an appendix to this question, see [19], p.57. [39] is often referred to as one of the seminal articles for the introduction of Bayesian methods in measurement uncertainty analysis.

⁴As [9, p.S150] explains, “Plain and simple, the GUM tells us that what we are uncertain about is the value, meaning true value, of the measurand.” The concept of “true value” of a quantity is without a doubt tricky and controversial. However, I will take Ehrlich’s statement as a sound starting point and will not question its validity here. I will purposely avoid any debate about the concept of “true value”, which I discuss further in [14].

3 The VIM1 definition of measurement uncertainty is flawed

The VIM1 defines measurement uncertainty as “an estimate characterizing the range of values within which the true value of a measurand lies”. This definition cannot work: it is deterministic⁵, and admittedly, knowledge is uncertain. A concrete example of why the definition fails is found in precision physics. The so-called “adjustments of the physical constants” (see for example [28]) offer to track the recommended values of the physical constants since 1929. The past and present recommended values of the Planck constant are shown on figure 1 (reproduced from [15] in which I examine this procedure in more details). This is where the definition fails: according to the VIM1 definition of measurement uncertainty, the true value of the Planck constant should be contained in all uncertainty intervals; and yet several of these intervals are disjoint⁶. A classical explanation to these leaps is the appeal to measurement error: it is likely that some of the measurements were affected by unknown systematic errors, which in turn make the VIM1 definition fail.

Perhaps the intervals need not to *always* overlap: one might simply argue that the VIM1 can be fixed by acknowledging the probabilistic nature of measurement uncertainty. Accordingly, the problem could be solved by turning the proposition into a probabilistic statement:

Amended VIM1 proposal. *Measurement uncertainty: an estimate characterizing the range of values within which the true value of a measurand **probably** lies.*

In this new definition, one must clarify what “probably” refers to. A traditional way to interpret this is in frequentist terms. In that case, the probability attached to the uncertainty refers to the expected *success rate* of the measurement process. The success rate is the relative frequency with which the uncertainty intervals (measurement results) cover the true value in the long-run⁷ (under infinite repetition of the experiment).

However, the amended VIM1 proposal stumbles upon a major problem: it can only account for error sources that have been identified⁸. The actual “success rate” of a measurement process is unknown: it can only be *predicted* under a set of hypotheses, including specific statistical models, *and the hypothesis that no unknown systematic error*

⁵One might argue that the definition only makes measurement uncertainty an “estimate” here. However, this specification is ambiguous and the rest of the definition still makes measurement uncertainty a characterization of where the true value effectively lies.

⁶One might simply argue that h itself could vary over time. Such a hypothesis cannot *a priori* be ruled out; however, no research work has yet suggested the possibility of a variation of this magnitude.

⁷This interpretation has known many debates in statistics and metrology. For an example of a recent disagreement about the meaning of “coverage intervals”, see [17, 25, 33, 26, 40, 41].

⁸As explained in the previous section, the frequentist appraisal of uncertainties also faces a critical challenge from a technical standpoint, namely the probabilistic treatment of systematic errors. Whether or not such a probabilistic treatment is actually possible is outside of the scope of this paper and does not really impact my analysis. What matters is that, even though the frequentist treatment of measurement uncertainty would include systematic errors, it could not ultimately process errors whose source has not been identified.

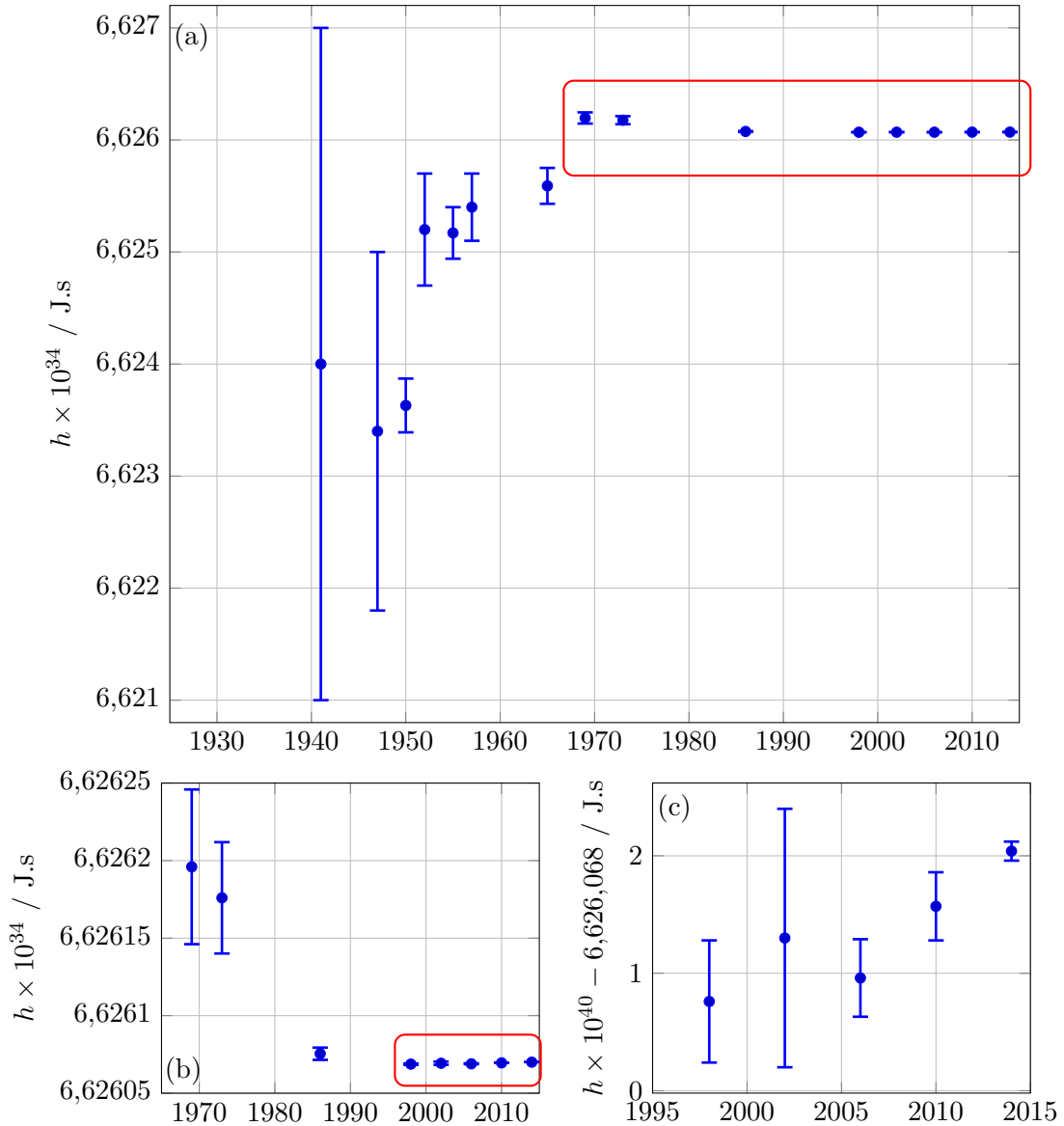


Figure 1: Evolution over time of the recommended values of the Planck constant h , along with their (1σ) uncertainties. It is instructive to zoom in on the later years of the graph – where measurement uncertainty is negligible compared to what it was in the middle of the century. Such zooms are indicated by the rectangles on the graphs (a) and (b), and end up in graphs (b) and (c) respectively. For practical reasons, two previous values, published by Birge in 1929 and 1932, were ignored in this graph because their plotting requires a change in scale that would obscure the rest of the observations. Both these values differ significantly from the 1941 value, despite a much higher uncertainty. This is a figure constructed with data collected from the literature, reproduced from [15].

*affects the measurement(s)*⁹. Nothing can guarantee the success rate of a measurement process, because nothing guarantees that no unknown systematic error affects a measurement. In a frequentist framework, the amended VIM1 definition still does not work.

My argument here is terminological: it does *not* entail that the frequentist interpretation is inoperative for measurement uncertainty evaluation. I will now carry on explaining my criticism of the VIM4(?) definition.

⁹The GUM itself acknowledges this: “only if there is a sound basis for believing that all of this has been done properly, with no significant systematic effects having been overlooked, can one assume that the measurement result is a reliable estimate of the value of the measurand and that its combined standard uncertainty is a reliable measure of its possible error.”, [19, p.51].

4 The suggested VIM4 definition of measurement uncertainty appeals to an overly restrictive notion of “belief”

By contrast with a characterization of uncertainty in term of “success rate”, metrologists have explored another path. As early as 1970 [11, 38], some physicists suggested to venture into the terminology of “betting odds”: one should be ready to *bet* on one’s own results, the reward being indexed on the probability attached to the uncertainty. The spread of Bayesian statistics within measurement science is in part inherited from this idea. For several decades, metrologists have emphasized the subjective status of a measurement result: while it was traditionally conceived as an objective claim about the measurand itself, Bayesians argued that it rather refers to *the state of knowledge of the agent* (or community of agents) who investigates the measurand¹⁰. In that context, an uncertainty statement is not a claim about *where the true value lies*, but about *what the agent believes* (alternatively: *knows*) about the true value. Walter Bich, for example, characterizes measurement uncertainty as “the logical reciprocal of state of knowledge” [4, p.2155]. In a way, the amended VIM1 proposal suggested in section 3 could fit into this background – with the appropriate interpretation of “probably” as a degree of belief¹¹. Metrologists, though, have undertook to find a better, more explicit formulation, encapsulated in the VIM4(?) definition: “parameter [...] characterizing how well the (essentially unique)¹² true value of the measurand is believed to be known”. In comparison to the VIM1, this definition has the merit of acknowledging that the “true value” of the measurand is unknown.

There is a downside to this improvement: the meaning of “belief” and “knowledge” in the definition are not straightforward and have to be clarified. Indeed, in his terminological review, Charles Ehrlich adds a comment right after the suggested definition for measurement uncertainty:

a Note could then elaborate on what belief means: ‘Note: the degree of belief, expressed as a probability, is based on the available information (obtained from the measurement and possibly also from other sources), and assumes that no mistakes have been made in performing and reporting on the measurement.’ [9, p.S150]

This cannot be satisfying. Admittedly, the note introduces an important idea: that the conclusions driven by an uncertainty statement are conditioned to some hypotheses. However, it is too vague and addresses the general issue of belief and knowledge too simplistically. My claim is that reducing measurement uncertainty to what one believes

¹⁰The epistemological ramifications of this shift are described in [8].

¹¹“Looking ahead a little to the next editions of the VIM and GUM, a definition of measurement uncertainty might be considered more along the lines of the definition in the first edition of the VIM shown earlier, with the important addition of the concept of degree of belief, expressed as a probability” [9, p.S150].

¹²The “essentially unique” mention belongs to a whole other issue about “definitional uncertainty” that needs not to be developed here.

is too trivial: uncertainty should better be understood as stating what *the circumstances of the experiment(s) performed tell the experimenter* to believe. Then, the experimenter can hold this belief or not, depending on how *confident*¹³ he is regarding the measurements and the models involved in the analyses¹⁴. Such confidence is independent of the measurement uncertainty itself¹⁵.

An illustration of this tension between uncertainty and belief is given by a notorious example¹⁶: in 2011, the international collaboration OPERA announced the possible observation of faster-than-light neutrinos [1]. The result obtained was quite significant, with a standard uncertainty approximately six times smaller than the difference between the measured muon neutrino velocity and the speed of light in vacuum. If this result is to be interpreted in the light of the VIM4(?) definition, then one has to conclude that the OPERA collaborators *believed they knew* that the true speed of neutrinos was greater than the speed of light with a probability (degree of belief) well beyond 99%. But they did not actually believe that. The first publication [1, p.22], as well as an early CNRS statement [7], were very cautious with regard to the potential high impact of the result, and they immediately pushed for further investigations of unknown errors in the experimental chain. Indeed, this skepticism was justified one year later, when the CERN announced that “the original OPERA measurement can be attributed to a faulty element of the experiment’s fibre optic timing system” ([7], updates of 2012). In the end, as the CERN Research Director expressed himself, “although this result isn’t as exciting as some would have liked, it is *what we all expected deep down*” ([7], updates of 2012 – my emphasis).

This example illustrates that a low uncertainty does not automatically entail a high belief – not until, at least, it is made sure that no unknown systematic error has been overlooked. Such a situation typically happens, in physics, when different experiments give discordant results. Let’s mention, for instance, the case of the “17-keV neutrino” documented by Allan Franklin [12]; the more recent “proton radius puzzle”¹⁷, briefly sketched by Nadine de Courtenay and I [8]; or the measurement of the neutrino lifetime

¹³Confidence may typically arise from robustness, when different measurement methods converge towards a common conclusion. In that case, the more methods are used, the less likely it is that the results are affected by significant unknown systematic errors.

¹⁴Same goes, then, for other agents who might have to rely on the result and whose confidence in the result depends on different parameters, among which their own trust in the source. This is an instance of what the philosopher John Hardwig called “epistemic dependence” [18], namely the fact that knowing something almost always implies relying on other people’s knowledge, and thus on (rational) trust. Metrology, as the “hidden infrastructure” [34] of science and technology, has the mission to secure the robustness of these relationships of trust.

¹⁵This might be even more striking from the perspective of manufacturing industry, see [27].

¹⁶I mentioned this example earlier in [15].

¹⁷Different methods of measurement of the proton radius yielded two discordant groups of results, and the CODATA Task Group on Fundamental constants had to decide on how to proceed in view of the 2014 adjustment of the physical constants. The committee eventually chose to select one of the two groups of results, for complex motives that would be too long to analyse here. It turns out that it was the one to which was attached the larger uncertainty.

[29]. All these examples have in common the fact that when discussing the relevance of the results, measurement uncertainty and belief are somewhat disconnected. The same idea prevails outside of physics, for instance when evaluations from experts conflict¹⁸. As a conclusion, it would be more accurate to claim that measurement uncertainty is connected to belief *through* confidence. Indeed, it is the combination of measurement uncertainty and confidence in the result that triggers a given degree of belief. More generally, the issue of belief in measurement involves dynamics of confidence and trust that are not easily reducible to probabilities and uncertainty analysis. This leads to the much broader question of the methodology of science and its social organization (an idea further developed in [8]), of which measurement uncertainty is only an ingredient.

The argument remains valid within the Bayesian scope. The appeal to “belief” in the definition of measurement uncertainty is understandable in the context of a Bayesian approach where probabilities are by nature degrees of belief. But it is important to realize that Bayesian statistics define a very specific *formal* background for the quantification of rational beliefs from the perspective of probability theory. They are first and foremost a method to manipulate these beliefs quantitatively and to formalize how new information affects them. Therefore, probabilities do quantify degrees of belief in that case, but not necessarily those of the cognitive agent who is confronted to the issues about confidence that I mentioned earlier. A way to summarize this could be to say that, in this context, Bayesian probabilities deal with what one *ought to* believe, given one’s prior knowledge and given the empirical and theoretical information one collects, *as long as* one is confident in the fact that this information is correct. Again, one’s belief relies on the confidence in the information collected.

Just as my criticism of the VIM1 definition was not a criticism of frequentism in itself, a similar comment should be made here: at no point does the argument displayed here attack the foundations of Bayesianism or its application to metrology. In fact, the argument does not suggest to take any side in the controversy that opposes frequentist to Bayesians in metrology and other areas of science. It merely points out that the notion of “belief” is not a simple one and that its use within the Bayesian formalism cannot be directly imported into a general definition of measurement uncertainty.

¹⁸Yet another example – to be explored – could be conflicting opinion polls.

5 Directions for improvement

The analyses of the VIM 1 to 4 provide us with two important lessons. First, the VIM2&3 remind us that measurement uncertainty is a *statement*, made *by* the experimenter *about* the measurand. This does not appear in the VIM1 and is not made explicit in the VIM4(?). More precisely, measurement uncertainty is a descriptive parameter concerning an *inference* about the true value of the measurand, founded on a series of idealizations, approximations and models. Based on the measurement outcomes, but also on statistical, experimental and theoretical models, as well as on previous tabulated data and theoretical knowledge, scientists infer that the value of the measurand should be located somewhere within the uncertainty interval.

Secondly, what appears from the limits of the VIM1&4 is that they are both concerned, though differently, with a common problem that could be called “problem of unknown systematic errors”¹⁹. Indeed, the discussions engaged previously lead to an apparently obvious but fundamental point in the appraisal of measurement uncertainties: *measurement uncertainty can only be inferred from what is accessible to our knowledge*. When experimenters do not have any information at all about measurement errors, they cannot account for it *at all* in their uncertainty calculus. Accordingly, an uncertainty budget only deals with *identified*²⁰ sources of uncertainty, and nothing is said about the *unknown unknowns* [32]. It would not have made any sense, in the initial uncertainty budget of the OPERA experiment, to add a line concerning “possible issues with optical fibers” if those issues were not identified in the first place. One cannot account for the unknown – and should not try to do so. Metrologists are well aware of this: they have for long criticized “safe” or “conservative” approaches of measurement uncertainty, where uncertainty is “often made deliberately large” [19, Annex E, p.54]. They advocate “a realistic rather than a ‘safe’ value of uncertainty” instead. As a consequence, measurement uncertainty cannot simply be the logical reciprocal of state of knowledge. Instead, it is a claim about what one knows that one does not know.

I now wish to sketch different options to improve²¹ the definition of measurement uncertainty while taking the two previous lessons into account. These options are displayed below; they are inspired from the historical and philosophical analysis I provided in [15]. The definitions are based on one major rationale: that it is not imperative to explain in the definition what the probability refers to. Would this be considered as necessary, the definitions could be amended but would become more technical.

First, the definition could start with a common preamble to all propositions, inspired

¹⁹Elsewhere I suggested that, borrowing from the terminology used in the GUM, it could be called the “problem of unknowability” [15].

²⁰I am not implying here that scientists always know the causes of the uncertainties – a statistical dispersion, for example, is a source of uncertainty but its (very multiple) causes are difficult, if not impossible, to establish.

²¹In that process, I do not claim in any way that these options are perfect. I am aware in particular that simplicity is not their priority. I do hope, though, that they can shed light on important aspects relative to the meaning of measurement uncertainty.

from the VIM3 definition:

Measurement uncertainty: *Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the experimental data and on other available information. [To be completed: see below]*

Then, a first option to proceed is to adopt the terminology of the traditional framework of measurement uncertainty – designated as the “error approach” to measurement in the VIM3, and to which “the objective of measurement (...) is to determine an estimate of the true value that is as close as possible to that single true value” [24, p.viii]. A possible definition is:

Proposition 1 (“error approach” framework)

Measurement uncertainty: *Non-negative parameter [etc. (see above)]. Measurement uncertainty characterizes the range of values within which the [essentially unique] true value of the measurand probably lies, according to the experimental results. It is inferred from a set of hypotheses, including physical and statistical models and the assumption that no unknown systematic error affects the measurement.*

If one prefers to express measurement uncertainty in epistemic terms, the following propositions may be suggested:

Proposition 2a (in terms of belief)

Measurement uncertainty: *Non-negative parameter [etc. (see above)]. Measurement uncertainty characterizes how well the (essentially unique) true value of the measurand is believed to be known, provided that one is confident that no unknown systematic error affects the measurement.*

Proposition 2b (in terms of knowledge)

Measurement uncertainty: *Non-negative parameter [etc. (see above)]. Measurement uncertainty characterizes what one knows that one does not know, based on the sources of uncertainty that one was able to identify.*

Note: an uncertainty statement makes no claim about the possible unknown unknowns - i.e. unidentified sources of uncertainty.

Despite their differences, these three propositions do actually express a common general idea. If one is confident in the result, one should expect the true value to lie within the uncertainty interval. Therefore proposition 2a entails proposition 1. The reciprocal is also true: if the true value probably lies in the uncertainty interval, then one should believe it – provided, again, confidence in the result. Understanding measurement uncertainty as the result of a risky inference at least partly conciliates the various standpoints about measurement (frequentist vs Bayesian, “error” vs “uncertainty” approach). All propositions have the advantage of affording room enough for discussing the merits of the different methods (and especially of the statistical models) involved in uncertainty analysis.

The comparison of various results obtained about the same measurand (or more complex comparisons like the adjustments of the physical constants) provides a robustness test for measurement uncertainty evaluation [37, 2]. In particular, two discordant results about the same measurand, for example two recommended values of the Planck constant that do not agree (see fig.1) indicate that the confidence in these results might have to be reevaluated. But the existence of such a conflict *is* compatible with the definition of measurement uncertainty given in propositions 1, 2a and 2b, whereas it invalidates the VIM1 definition. The observation of discordant results is a scientific issue that requires a resolution, whether by identifying previously unknown systematic errors, revising background models, or discovering new theoretical knowledge. Accordingly, measurement uncertainty is an indicator of the sensitivity (as introduced by Pierre-Hugues Beauchemin and Kent Staley [3, 36]) with which scientists are able to probe these kinds of conflicts and scientific theories in general.

6 Conclusion

The definitions of measurement uncertainty provided in the first three editions of the VIM and in the upcoming fourth edition are incomplete, and they do not accurately explain the meaning of the concept when they try to do so. A proper definition of measurement uncertainty should state the meaning of measurement uncertainty. An option would be to combine the VIM3 definition with an interpretation of measurement uncertainty in terms of the true value of the measurand. The VIM4(?) attempt points towards an interesting direction but its use of “belief” and “knowledge” is too narrow and does not acknowledge the problem of unknown systematic errors. The definition of measurement uncertainty should thus acknowledge that measurement uncertainty can either be seen as a descriptive parameter of an inference about where the true value of the measurand lies; or a parameter characterizing what one should believe as long as one is confident in the results; or, alternatively, a descriptive parameter of what one knows that one does not know. All three propositions converge regarding the meaning of measurement uncertainty, and explain for example its role in theory testing. Devising the definitions in a document like the VIM is a demanding and arduous task. Surely my propositions ignore a lot of constraints and difficulties that dictate the work of the writers of the document. I nonetheless hope they may contribute to the necessary improvement of the definition of measurement uncertainty.

Acknowledgements

The author wishes to thank Nadine de Courtenay and Olivier Darrigol for their reviews on earlier drafts of the paper – all remaining errors are mine. Moreover, I am truly grateful to the members of the metrology community with whom I was able to exchange on related issues at different stages of my research. Special thanks go to Walter Bich, Michèle Désenfant, Charles Ehrlich, Luca Mari, Nicolas Fischer and Marc Priel. This work was supported by the Cohn Institute for the History and Philosophy of Science and Ideas at Tel Aviv University and by University Paris Diderot.

References

- [1] ADAM, T. *ET AL.* (2011), Measurement of the neutrino velocity with the OPERA detector in the CNGS beam, *ArXiv*, pp.1–24.
- [2] BASSO, A. (2017), The appeal to robustness in measurement practice, *Studies in History and Philosophy of Science Part A*.
- [3] BEAUCHEMIN, P.-H. (2017), Autopsy of measurements with the ATLAS detector at the LHC, *Synthese*, 194(2), pp.275–312.
- [4] BICH, W. (2012), From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty, *IEEE Transactions on Instrumentation and Measurement*, 61(8), pp.2153–2159.
- [5] CAUSSARIEU, A. et TIBERGHEN, A. (2017a), Adapter l’enseignement pour donner du sens aux incertitudes de mesures, *Bulletin d’Union des Physiciens*, 111(998), pp.1111–1126.
- [6] CAUSSARIEU, A. et TIBERGHEN, A. (2017b), When and Why Are the Values of Physical Quantities Expressed with Uncertainties? A Case Study of a Physics Undergraduate Laboratory Course, *International Journal of Science and Mathematics Education*, 15(6), pp.997–1015.
- [7] CERN (2011), OPERA experiment reports anomaly in flight time of neutrinos from CERN to Gran Sasso, *CERN Press Release*. 23th September, 2011. <http://press.cern/press-releases/2011/09/opera-experiment-reports-anomaly-flight-time-neutrinos-cern-gran-sasso>
- [8] de COURTENAY, N. et GRÉGIS, F. (2017), The evaluation of measurement uncertainties and its epistemological ramifications, *Studies in History and Philosophy of Science Part A*, 65-66, pp.21–32.
- [9] EHRLICH, C. (2014), Terminological aspects of the *Guide to the Expression of Uncertainty in Measurement* (GUM), *Metrologia*, 51(4), pp.S145–S154.
- [10] EHRLICH, C., DYBKAER, R. et WÖGER, W. (2007), Evolution of philosophy and description of measurement (preliminary rationale for VIM3), *Accreditation and Quality Assurance*, 12, pp.201–218.
- [11] FRANKEN, P. (1971), Comments on the Assignment of Experimental Uncertainties, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.507–508, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- [12] FRANKLIN, A. (1995), The appearance and disappearance of the 17-keV Neutrino, *Reviews of Modern Physics*, 67(2), pp.457–490.

- [13] GIACOMO, P. (1981), News from the BIPM, *Metrologia*, 17(2), pp.69–74.
- [14] GRÉGIS, F. (2015), Can we dispense with the notion of “true value” in metrology?, in SCHLAUDT, O. et HUBER, L. (dir.), *Standardization in Measurement: Philosophical, Historical and Sociological Issues*, pp.81–93, Pickering & Chatto.
- [15] GRÉGIS, F. (2019), Assessing accuracy in measurement : The dilemma of safety versus precision in the adjustment of the fundamental physical constants, *Studies in History and Philosophy of Science*, 74, pp.42–55.
- [16] HACKING, I. (2001), *An Introduction to Probability and Inductive Logic*, Cambridge University Press.
- [17] HALL, B. D. (2008), Evaluating methods of calculating measurement uncertainty, *Metrologia*, 45(2), pp.L5–L8.
- [18] HARDWIG, J. (1985), Epistemic Dependence, *The Journal of Philosophy*, 82(7), pp.335–349.
- [19] JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008a), *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*, Sèvres: JCGM.
- [20] JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008b), *Evaluation of measurement data - Supplement 1 to the “Guide to the expression of uncertainty in measurement” - Propagation of distributions using a Monte Carlo method*, Sèvres: JCGM.
- [21] JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2009), *Evaluation of measurement data - An introduction to the “Guide to the expression of uncertainty in measurement” and related documents*, Sèvres: JCGM.
- [22] JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2011), *Evaluation of measurement data - Supplement 2 to the “Guide to the expression of uncertainty in measurement” - Extension to any number of output quantities*, Sèvres: JCGM.
- [23] JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2012a), *Evaluation of measurement data - The role of measurement uncertainty in conformity assessment*, Sèvres: JCGM.
- [24] JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2012b), *International vocabulary of metrology - Basic and general concepts and associated terms (VIM)*, Sèvres: JCGM. 3rd edition. 2008 version with minor corrections.
- [25] LIRA, I. (2008), On the long-run success rate of coverage intervals, *Metrologia*, 45(4), p.L21.
- [26] LIRA, I. (2009), On the meaning of coverage probabilities, *Metrologia*, 46(6), pp.616–618.

- [27] LOFTUS, P. et GIUDICE, S. (2014), Relevance of methods and standards for the assessment of measurement system performance in a High-Value Manufacturing Industry, *Metrologia*, 51, p.S219.
- [28] MOHR, P. J., NEWELL, D. B. et TAYLOR, B. N. (2016), CODATA recommended values of the fundamental physical constants: 2014, *Reviews of Modern Physics*, 88, p.035009.
- [29] MUMM, P. (2018), Resolving the neutron lifetime puzzle, *Science*, 360(6389), pp.605–606.
- [30] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) (1984), *International vocabulary of basic and general terms in metrology*.
- [31] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) (1993), *International vocabulary of basic and general terms in metrology*.
- [32] POMMÉ, S. (2016), When the model doesn't cover reality: examples from radionuclide metrology, *Metrologia*, 53(2), p.S55.
- [33] POSSOLO, A., TOMAN, B. et ESTLER, T. (2009), Contribution to a conversation about the Supplement 1 to the GUM, *Metrologia*, 46(1), pp.L1–L7.
- [34] QUINN, T. J. (2002), Metrology, its role in today's world, in LIRA, I. (dir.), *Evaluating the Measurement Uncertainty: Fundamental and Practical Guidance*, pp.1–23, Bristol and Philadelphia, Institute of Physics Publishing.
- [35] RIDLER, N. M. et SALTER, M. J. (2014), Evaluating and expressing uncertainty in high-frequency electromagnetic measurements: a selective review, *Metrologia*, 51(4), pp.S191–S198.
- [36] STALEY, K. W. (2017), Securing the Empirical Value of Measurement Results, *The British Journal for the Philosophy of Science*, p.axx036.
- [37] TAL, E. (2011), How Accurate is the Standard Second?, *Philosophy of Science*, 78(5), pp.1082–1096.
- [38] THOMSEN, J. S. (1971), Some Aspects of Least-Squares Adjustments of Constants, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.503–505, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- [39] WEISE, K. et WÖGER, W. (1993), A Bayesian theory of measurement uncertainty, *Measurement Science and Technology*, 4(1), pp.1–11.
- [40] WILLINK, R. (2010a), On the validity of methods of uncertainty evaluation, *Metrologia*, 47(1), pp.80–89.

- [41] WILLINK, R. (2010b), Probability, belief and success rate: comments on ‘On the meaning of coverage probabilities’, *Metrologia*, 47(3), pp.343–346.