



HAL
open science

Assessing accuracy in measurement: The dilemma of safety versus precision in the adjustment of the fundamental physical constants

Grégis Fabien, Fabien Grégis

► To cite this version:

Grégis Fabien, Fabien Grégis. 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 Part A*, 2018, 10.1016/j.shpsa.2018.09.001 . halshs-01884061

HAL Id: halshs-01884061

<https://shs.hal.science/halshs-01884061>

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.

Assessing accuracy in measurement: The dilemma of safety versus precision in the adjustment of the fundamental physical constants

Fabien GRÉGIS

Cohn Institute, Univ. Tel Aviv & SPHERE laboratory, Univ. Paris Diderot
fabien.gregis@etu.univ-paris-diderot.fr

Studies in History and Philosophy of Science Part A
Preprint version, 2018-09-08

Contents

1 Introduction	2
2 The problem of unknowability	4
3 The adjustment of the physical constants: Birge's initiative	8
3.1 The principles of an adjustment	8
3.2 The speed of light in 1929	10
3.3 The Planck constant in 1929	13
3.4 Two roles of measurement uncertainty and a resulting tension	16
4 Accuracy versus precision in the adjustment of the constants	17
4.1 The perpetuation of Birge's initiative	17
4.2 Selection issues	18
4.3 Bender's criticism: uncertainty as a safe accuracy claim	19
4.4 Taylor's answer: uncertainty as insensitivity	22
5 Measurement uncertainty: objective and epistemic stances	25
5.1 The objective stance	26
5.2 The epistemic stance	27
6 Conclusion	29
Acknowledgements	30
References	30

Abstract. This article develops a historico-critical analysis of uncertainty and accuracy in measurement through a case-study of the adjustment of the fundamental physical constants, in order to investigate the sceptical “problem of unknowability” undermining realist accounts of measurement. Every scientific result must include a “measurement uncertainty”, but uncertainty cannot be evaluated against the unknown, and therefore cannot be taken as an assessment of “accuracy”, defined in the metrological vocabulary as the closeness to the truth. The way scientists use and interpret uncertainty in the adjustment activity illustrates how they try to overcome this predicament. I identify two operative roles of measurement uncertainty, in the comparison and in the combination of measurement results. This duality implies a tension between selection and combination when aggregating results, leading to a crucial question: should the evaluation of uncertainties favour safety over precision? I present contrasting answers and identify a specific account of physicists who implicitly try to conciliate realism with the problem of unknowability. I argue that this invites us to reconsider accuracy from a dynamical standpoint, as the gradual achievement of scientific progress through error correction. Finally, I present two interpretations of measurement uncertainty, objective and epistemic, which I criticize and suggest improvements to.

1 Introduction

The adjustment of the fundamental physical constants designates a crucial collective endeavour in precision physics, initiated in 1929 by the American physicist Raymond T. Birge¹ (1887–1980). An adjustment consists in collecting, at one given time, all the up-to-date best measurements in precision physics in order to produce a self-consistent set of temporary “recommended values” of the fundamental physical constants. At the same time, it allows a precise test of the agreement between experiment and the main physical theories involved. This activity is now perpetuated under the auspices of an international institution, the CODATA (Committee on Data for Science and Technology), which conducted a first adjustment in 1973², and since then regularly updates an official list of values. The latest list was published in 2014³, and, as an example, displays the recommended value for the Planck constant h as such⁴:

Value	$6.626\,070\,040 \times 10^{-34}$ J.s
Standard uncertainty	$0.000\,000\,081 \times 10^{-34}$ J.s
Relative standard uncertainty	1.2×10^{-8}
Concise form	$6.626\,070\,040(81) \times 10^{-34}$ J.s

Physicists do not simply provide a single value of h , that would be “the” unique accepted value of the constant, but they add a component known as the “measurement uncertainty”, emphasizing that there remains a margin of doubt that has been accounted for by tracking the possible additional sources of measurement error⁵.

¹Birge (1929)

²Cohen and Taylor (1973)

³The official report was published in 2016, see Mohr, Newell and Taylor (2016).

⁴NIST website, URL = <http://physics.nist.gov/cgi-bin/cuu/Value?h>. This recommended value will change with the upcoming reform of the International System of Units, which will be formally validated by the General Conference on Weights and Measures (CGPM) in November 2018 and implemented in 2019. The value of h will then be fixed in order to redefine the kilogram. All recommended values are freely accessible online: URL = <http://physics.nist.gov/cuu/constants> (last accessed on 2017-12-17).

⁵A very clear description of such a process, often called “uncertainty budgeting”, is proposed in Tal (2011) in the case of time measurements. As Tal also demonstrates, this is a model-based activity, Tal (2017b).

Measurement uncertainty combines epistemological and practical concerns. On the one hand, it is an echo of the inquest about the value of measurement with regard to truth, knowledge, confidence or belief. On the other hand, it is a quantitative parameter integrated into a sharply defined formalism of error analysis. Its determination is a crucial practical task for all experimentalists, enabling them to confront experimental results with each other, or with hypotheses, technical norms and tolerances. Despite two centuries of maturation in different fields of the natural and social sciences, lively debates continue to agitate the scientific community about the meaning of measurement uncertainty and its methods of determination. Consequently, uncertainty offers historians and philosophers a privileged ground to explore epistemological issues from a practice-oriented viewpoint. The adjustments of the physical constants, in particular, are at the intersection of two disciplines which have very much contributed to the development of uncertainty analysis: precision physics and metrology, the latter being the discipline dedicated to the methodological and conceptual foundations of measurement. The history of adjustments is thus filled with debates about the meanings of “error”, “uncertainty” and “accuracy”.

Questions about accuracy and uncertainty in measurement are being discussed with a growing interest by historians and philosophers and are, especially since the early 2000s, part of the core matter of the epistemology of measurement⁶. My analysis digs more specifically into issues about the nature and function of “accuracy” in science. As Eran Tal has shown, “measurement accuracy” is a polymorphic term, for which he identified at least five senses⁷. One question is how all its avatars relate to each other. In particular, how can a “metaphysical accuracy”, correlated to truth, be assessed? And how is it connected to an “epistemic accuracy”, which pertains to knowledge, confidence and belief? My analysis draws on an essential, quasi-kantian⁸ obstacle: if measurement is said to appraise “true” states of reality (“true values” in the vocabulary of measurement), then one is ineluctably led to a sceptical dead-end undermining realist accounts of measurement. Scientists have expressed this concern in a way that I shall designate as “the problem of unknowability”: true values are unknowable and therefore measurement accuracy cannot be evaluated.

Tal, while not being so much interested in issues about measurement realism (that tend to go beyond the scope of the *epistemology* of measurement), has nonetheless suggested that metaphysical accuracy could be “powerfully tested” by a robustness condition⁹ and further argued that measurement uncertainty is an instance of “predictive accuracy”, that is, the “uncertainty associated with [a] model-based prediction”¹⁰. I will adopt a different perspective. Even though I will not attempt to contest or to embrace measurement realism, I will use it as

⁶See for instance Chang (2004), Tal (2011, 2016), Boumans, Hon and Petersen (2014).

⁷Tal (2011), pp.1084–1085.

⁸As Marian points out about the correspondence theory of truth, “it is typically pointed out that one cannot step outside our own minds to compare our thoughts with mind-independent reality”, Marian (2015), section 9.2 (last consulted 2017-11-16). Marian traces it back to Kant, who stated that “since the object is outside me, the cognition in me, all I can ever pass judgment on is whether my cognition of the object agrees with my cognition of the object”, Kant (1992), pp.557–558.

⁹Tal (2011), pp.1090–1093.

¹⁰Tal (2017a), p.7.

a starting point – as being a classical category for the scientific actors – in order to initiate a historico-critical analysis in which no commitment is required on the realism issue. To this end, I will concentrate on the way measurement uncertainty is used and interpreted by physicists in the specific field of the adjustments¹¹. I will argue that the physicists involved in this field, while trying to deal with the problem of unknowability, have implicitly developed an epistemological position which leads to reconsider accuracy from a dynamical point of view, as the gradual achievement of scientific progress¹² through error correction. This position reconciles realism and the problem of unknowability while avoiding scepticism.

In section 2, I provide an instance of the problem of unknowability, based on the example of the history of the recommended values of the Planck constant. These values appear to converge, but with typical accidents that reveal a threat: unknown “systematic” errors. Section 3 analyses Birge’s pioneering work on the adjustments through two examples, the Planck constant and the speed of light. I identify two uses of measurement uncertainty in this process and underline a tension between selection and combination that triggered criticism among the community. This criticism is sketched in Section 4, where I also show how important actors of the field answered it. Their answer addresses the problem of unknowability by focusing on long-term progress, to be secured through error correction. This viewpoint suggests a dynamical account of accuracy. Lastly, in section 5, I briefly venture into a systematic account of measurement uncertainty. I discuss two typical directions in the interpretation of measurement uncertainty, objective and epistemic, which I criticize and try to improve.

2 The problem of unknowability

One of the main features of the success of modern science resides in its ability to reach outstanding levels of accuracy. Measurements of the physical constants have been improving in accuracy by one significant figure every 10 or 20 years in average since the beginning of the twentieth century. However, a closer look at the recommended values of the Planck constant h since 1941 reveals a more complicated story. Fig.1 displays these values along with their respective uncertainties and exhibits two features of particular interest.

First, there seems to be an overall convergence: over time, the values are coming closer to each other and the uncertainties decrease at a rather regular pace, despite some accidents. This first feature nicely fits a standard conception of scientific progress, as the values seem to tend asymptotically towards an unknown limit that one could be tempted to interpret as the actual, fixed but unknown, value of h . But secondly, on multiple occasions the error bars (the segments

¹¹I will purposely avoid entering into the statistical side of the debates, which would imply to dig into the methodology of uncertainty analysis, and which are not decisive to my argument. For a review of the epistemological ramifications of the evaluation of measurement uncertainties, see [de Courtenay and Grégis \(2017\)](#).

¹²The notion of “progress” in science is itself difficult to define, and is indeed at the centre of a lively philosophical debate today. Although the present case-study may certainly contribute to this general issue, my examination does not attempt any systematic account of scientific progress. Accordingly, I will not commit to any of the existing conceptions of progress in philosophy of science. The interested reader may refer to [Bird \(2007\)](#); [Rowbottom \(2010\)](#); [Niiniluoto \(2014\)](#) and [Dellsén \(2018\)](#) for examples of recent developments.

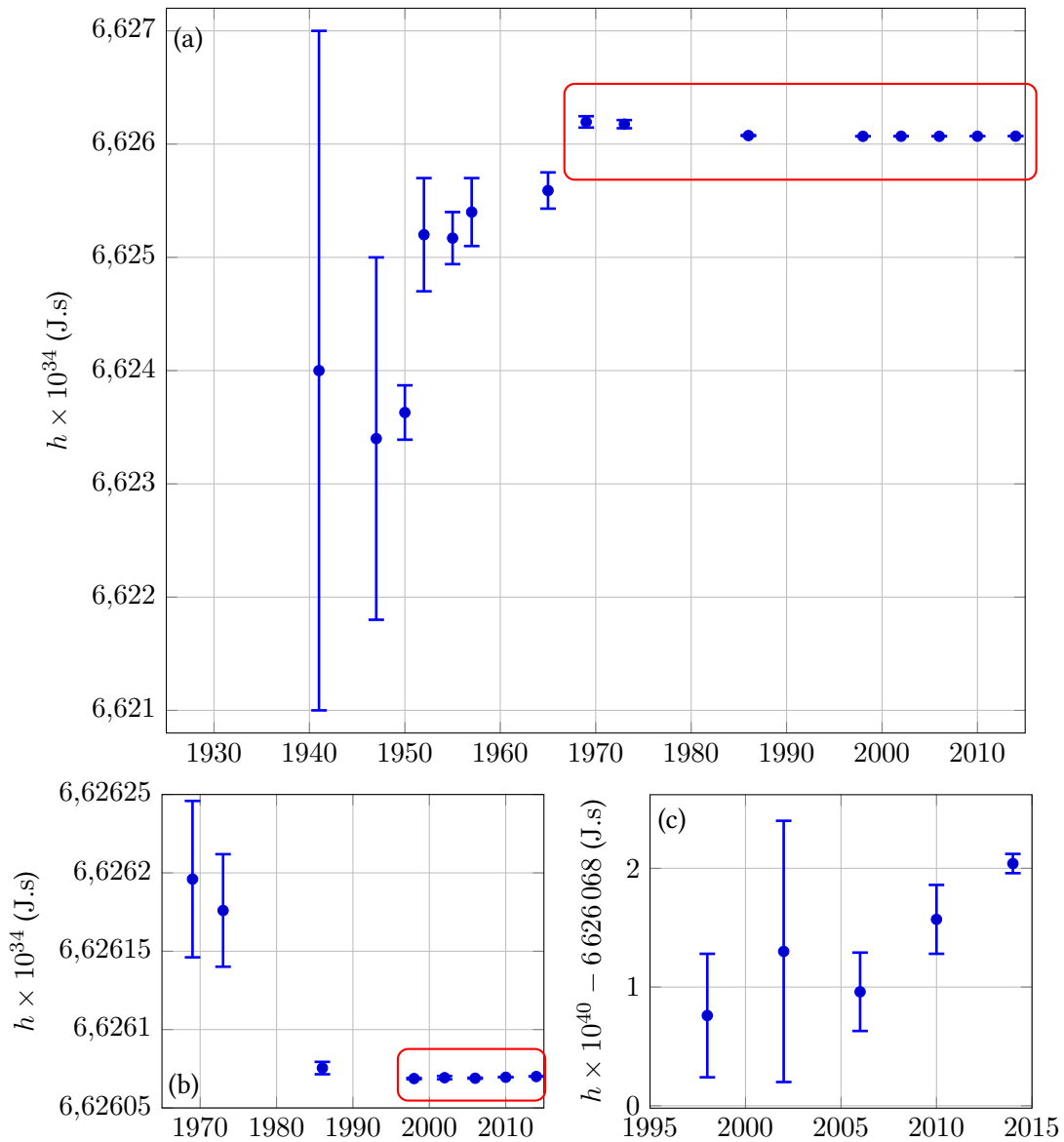


Figure 1: Evolution over time of the recommended values of the Planck constant h , as published by several different groups of scientists from 1941 to 1969, and by CODATA since 1973. Each result is represented by its central value (a dot) and its measurement uncertainty (the half-width of the “error bar”, namely the segment surrounding the central value). 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. Graphs (b) and (c) show that the overall convergence and the episodic jumps, observed on graph (a), persist in the later years. 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 an original figure constructed with data collected from the literature.

representing the measurement uncertainty on the graph) associated to two successive values do not overlap: there are a number of substantial leaps from one value to the other (e.g. from 1950 to 1952). A leap of this nature means the local presence of a significant discrepancy: the values disagree¹³.

Therefore, the story that this graph tells is not about a simple, linear progress. There is indeed nothing especially puzzling in observing that progress might not be uniform: so-called “systematic” errors may affect some of the results, and hopefully these can be eliminated or corrected in the future. The accidents displayed on the graph are nevertheless at the heart of interrogations about the meaning and the purpose of evaluating uncertainty in measurement. Is measurement uncertainty truth-related? In that case, how should the leaps be interpreted? Is it a knowledge claim? But then, how reliable is this claim? Is it a confidence claim? But then, how could any published result be trusted, when any recommended value expressed at one time can later be amended by more than what was indicated at first as the measurement uncertainty?

In the field of metrology, a classical way to interpret measurement uncertainty is to appeal to a realistic framework articulated to three key concepts: true value, error and accuracy. First, a particular quantity is characterized by its “true value”, a linguistic term which corresponds to the accurate representation of a (unique) state of reality¹⁴. Then comes measurement error. A perfect measurement of a physical quantity should yield its true value, but since no measurement is perfect, every result deviates from the true value. The amplitude of the deviation is defined as the measurement error. Finally, an experimenter expresses the outcome of error analysis by a statement of accuracy, understood here as the “closeness to the truth”¹⁵ – and corresponding to what Tal calls “metaphysical accuracy”¹⁶. Scientists are thus trying to give an indication about the effective *location* of the true value of the target quantity, and measurement uncertainty is meant to evaluate the final measurement error affecting the result. This account appears in the first edition of the *International Vocabulary of Metrology*, an essential metrology guide published by ISO in 1984, where measurement uncertainty is defined as:

Uncertainty of measurement: an estimate characterizing the range of values within which the true value of a measurand lies¹⁷

¹³Strictly speaking, this discrepancy can only be assessed on statistical grounds, since the error bar is not a deterministic interval but the dispersion parameter of a statistical distribution. My analysis remains merely based on qualitative claims, but taking into account the statistical nature of measurement uncertainty would not change the foundations of my argument.

¹⁴The concept of “true value” is controversial in metrology. Some scientists dislike this reference to “truth”, considered as too metaphysical to enter the metrology vocabulary, and its relevance has been intensely discussed. In this paper, I do not wish to elaborate on this issue. The interested reader can find a development in Grégis (2015).

¹⁵This is how Churchill Eisenhart defined the concept in his 1963 reference review of the field, which can be considered as one of the most comprehensive overviews of the traditional account. Contemporary metrology has retained this sense: the *International Vocabulary of Metrology* defines “measurement accuracy” as the “closeness of agreement between a measured quantity value and a true quantity value of a measurand [*the target quantity*]”, *Joint Committee for Guides in Metrology (JCGM) (2012)*, p.21.

¹⁶Tal (2011), p.1084.

¹⁷*International Organization for Standardization (ISO) (1984)*, p.16. The second (1993) and third (2008) editions of the document were written in a context of criticism, within metrology, of the traditional approach to measurement.

But the traditional account is quite problematic. Fig.1 obviously conflicts with the definition given in 1984: the plotted results cannot all contain the true value of h , since some of them are mutually incompatible. If one considers a correct result to be a result that contains the true value of the measured quantity, then one is led to the following conclusion: the recommended values of the physical constants cannot in any way be guaranteed as correct. Even acknowledging the probabilistic character of the values (whose calculus involves statistical models) is not enough: the discrepancies are too important¹⁸. Since the analysis of past recommended values indicates that, in retrospect, some of them might most probably be incorrect, then why could it not be also the case of the latest recommended values? This statement, illustrated by the graph of values of h , is an instance of a general issue with the traditional account of measurement. The main pitfall of this account is that it compels an experimenter to evaluate her result against the truth, which is unknown to her – knowing it being precisely the point of her measurements. As Tal explains,

“mind-independent truths about measurable quantities – for example, the exact true value of a quantity or the true equality of two magnitudes – are not cognitively accessible to metrologists and thus cannot serve as epistemic standards for measurement. [...] were metrologists required to evaluate the accuracy and stability of their measuring instruments against such truths, metrology would be mired in scepticism.”¹⁹

Scientists, and especially metrologists, have been addressing this issue with much emphasis in the latest decades, claiming that the true value is “unknowable”²⁰. Drawing on the latter expression, I propose to call this issue *the problem of unknowability*. The problem of unknowability entails that measurement accuracy cannot be evaluated – not as long as it refers to truth and true value. The explicit appeal to unknowability in metrology, in the early 1990s, resulted in repeated attempts to redefine measurement uncertainty in a way that departs from a requirement of truth and gets closer to an epistemic formulation²¹.

Before the early 1990s, during which the problem of unknowability was explicitly put forward as a serious problem in the scientific literature, the physicists involved in adjustments of the physical constants were already confronted with an implicit instance of this issue. My

Their definitions of measurement uncertainty are therefore quite different from the 1984 edition.

¹⁸One could indeed argue that measurement uncertainties are probabilistic in nature, and thus that uncertainty intervals are not required to overlap every time. Following this idea, a better definition of measurement uncertainty could for example be “the range of values within which the true value of a measurand frequently lies”, the frequency being attached to some probability. Such a redefinition supposes, though, that it is possible to evaluate measurement uncertainty on fully probabilistic grounds – a not so obvious task. Moreover, the redefinition would still not solve the issue at stake here, as the discrepancies are too frequent and too significant. Strictly speaking, demonstrating the latter claim would require an actual probabilistic analysis of the recommended values, and thus to delve into deep technicalities. This is not my intent here, and I will therefore not address the issue of the statistical significance of the discrepancies. I will rather use fig.1 as a qualitative illustration of the general problem of unknowability and refer to the physicist’s own points of view (exposed later) to show that they considered it as a problem themselves.

¹⁹Tal (2016), p.310.

²⁰See for example *Joint Committee for Guides in Metrology (JCGM) (2008)*, especially annex D.

²¹See Mari (2003), Grégis (2015).

point, in the next sections of this article, will be to show that these physicists have addressed the issue of unknowability from an original standpoint, trying to avoid scepticism without renouncing a traditional realist position. In order to reach this position, my analysis begins with the description of the first steps in the history of the adjustments.

3 The adjustment of the physical constants: Birge's initiative

The major constitutive principles of an adjustment of the physical constants were laid by Raymond T. Birge, an American spectroscopist working at the University of Berkeley (later named director of the physics department). In this section, I will concentrate on Birge's early work: this will enable me to identify two different roles for measurement uncertainty and discuss its possible interpretations, identifying the roots of an essential issue between combination and selection that I will discuss in the next section.

3.1 The principles of an adjustment

George Smith identifies the late nineteenth century, starting with the work of Joseph J. Thomson on the electron around 1897, as the emergence of a "full appreciation [...] of the evidence that can be obtained from precise measurement of fundamental constants"²². In the 1920s, scientists were engaged for several decades in regular high precision measurements of a continuously growing list of constants like the speed of light, the Planck constant, the elementary charge, or the gravitational constant. As Birge was exploring the collections of measurement results published at the time, he realized that apart from a few isolated cases, there had been no systematic attempt to put in common these data and to produce a critical analysis of the values of the fundamental physical constants. He was among the first to put the emphasis on the fact that the measurements of the physical constants had to be studied collectively. After some work on the Planck constant²³, Birge decided to perform a global review of the up-to-date knowledge of the constants, and published his conclusions in 1929 in an article entitled "Probable Values of the General Physical Constants"²⁴, in which he examined thirty-one constants. This article is regarded today as foundational for the practice of the adjustment of the physical constants²⁵.

An adjustment aims at answering the following question: given that there often are, at a given time, multiple determinations for each fundamental constant, sometimes performed by using different measurement methods, is it possible to identify a set of individual "best" (best-to-our-knowledge) values of these constants? The incentive for adjustments arises from a generalized observation: measurement results are not in full agreement with each other. Discrepancies in the empirical data are of two kinds. First, two different determinations of the

²²Smith (2010), p.566.

²³Birge (1919)

²⁴Birge (1929)

²⁵A. Carl Helmholtz considers Birge as the "pioneer" in this field, Helmholtz (1980), p.63. Philip Mirowski supports this claim when he explains that Birge's "major achievement was single-handedly to create the role of the meta-analyst in atomic physics", Mirowski (1994), p.576.

same physical constant rarely yield the exact same result. Secondly, the values of the constants drawn from experiment are almost never in a perfect fit with the web of theoretical relations in which they are involved. One may attempt to explain both kinds of discrepancies either by supposing that various measurement errors affect the experimental data or by identifying theoretical setbacks. The adjustment is founded on the first option: it supposes that the theory is correct, and thus that as long as the discrepancies are small, it is possible to reduce the errors by combining the results together. In some cases, the discrepancies may not be small, and combining the results would then imply stretching the original data quite outside of their uncertainty bounds, which is interpreted as an anomaly. The adjustment thus helps to reveal significant discrepancies that could have a theoretical explanation (although not always), and therefore constitutes “a new test of the validity of all our theoretical preconceptions and their experimental verification over the widest possible domain”²⁶. When the adjustment does not reveal any significant discrepancy, it illustrates the epistemic virtues of what Stephen Stigler called the “combination of observations”²⁷, namely the fact that combining slightly discordant results often yields more accurate values. This principle was not obvious in early precision physics, as witnessed by the physicists Joyce Bearden and John Thomsen:

In the early work in atomic physics there was a tendency to select one best experiment as fixing a particular constant or ratio. For example, Millikan’s oil drop experiment was long regarded as giving the most reliable value of e . This approach gradually changed to a more critical review of all pertinent data.²⁸

Technically speaking, the protocol of adjustments revolves around a “least-squares fit” which can roughly be assimilated to a sophisticated kind of simultaneous averaging of all constants – in which the assessment of uncertainties takes a crucial place. However, in his 1929 article, Birge had not yet designed the complete machinery. He rather chose to proceed with what he later dubbed the “method of successive approximations”²⁹, in which the constants were evaluated not simultaneously but successively. Birge began with the speed of light in vacuum c , which could be inferred from measurements without involving any other physical constant. He collected the best up-to-date determinations of c and drew from them its “most probable value”. Once the case of c was settled, its most probable value could be used as a fixed parameter whenever c was involved in a physical equation used to determine another physical constant. Birge could then repeat the same procedure on and on by iteration, spanning all the constants.

In what follows, I will take the example of two specific constants, the speed of light in vacuum c and the Planck constant h , in order to show how Birge processed them in his 1929 article. This will enable me to illustrate two major epistemic roles of measurement uncertainty, in the comparison and in the combination of measurement results.

²⁶Cohen and DuMond (1965), p.540.

²⁷Stigler (1986), p.11.

²⁸Bearden and Thomsen (1957), p.270.

²⁹Birge (1932b), p.228.

3.2 The speed of light in 1929

The speed of light in vacuum c was the first to be examined in Birge’s 1929 article³⁰. Drawing on earlier publications, Birge identified three remarkable determinations of c . At that time, Albert Michelson’s work was a reference in the field: he made direct measurements of the speed of light by using the rotating-mirror method³¹. Using the results he gathered between 1921 and 1926, Michelson obtained a value of $c = 299\,796 \pm 4$ km/s.³² Birge praised the quality of Michelson’s work to the point that he judged, in the case of direct measurements, that “this investigation so far surpasses in accuracy any of the preceding that it alone need be considered”³³. However, other indirect measurements were also worth examining.

Jean Mercier, a student of Henri Abraham at the ENS in Paris, had investigated the behaviour of stationary electrical waves in a pair of parallel copper wires³⁴. Electromagnetic theory indicated that the velocity of these waves should be equal to c , provided the wires were perfect conductors. By measuring simultaneously and independently the wavelength and the frequency of the stationary waves, and by correcting the bias due to the imperfect conductivity of the wires, Mercier obtained $c = 299\,700$ km/s, with a probable error estimated as 30 km/s.

Two decades earlier, Edward B. Rosa and Noah E. Dorsey, at the US National Bureau of Standards, had also proposed an indirect measurement of c based on another experimental principle³⁵. In the beginning of the twentieth century (since 1873 in Britain), the usual system of units used by physicists was the CGS system (for “Centimetres-Grams-Seconds”). However, two different CGS systems were used in electricity, an “electrostatic” and “electromagnetic” one. As a consequence of the electromagnetic theory of light, the ratio of the electric charge of a particle in the electrostatic system over the electric charge of that particle in the electromagnetic system was equal to the speed of light in vacuum. Therefore, by realizing the measurement of

³⁰Birge (1929), section A, pp.8–10.

³¹Michelson (1927)

³²In Birge’s article, uncertainty assessments were expressed as “probable errors”, a popular concept in many publications of the first half of the twentieth century. “Probable error” is one of the various concepts and terms that have been used since the beginning of statistical analysis of measurement errors to describe what is now regrouped under the term “measurement uncertainty”. It was defined “as the numerical magnitude which the true error is as likely to exceed as not” (Birge, 1929, p.4). While contemporary “measurement uncertainty” and Birge’s use of “probable error” are not strictly equivalent, they convey the same general idea. Importantly, “probable error” was often restricted to the evaluation of statistical errors, but Birge himself used it with an extended scope, encapsulating uncertainties about constant errors (Birge, 1929, p.4) and thus closer to today’s “measurement uncertainty”. In Birge’s framework, the difference between the concepts mainly lies in some details about their statistical appraisal. In particular, “probable error” was based on a statistical model supposing that statistical errors are distributed according to a normal distribution, whereas this hypothesis is not required in the analysis of contemporary “measurement uncertainty”. For statistical errors, when the normal distribution hypothesis stands for both concepts, one may simply apply a conversion factor of approximately 1.48 to compute an uncertainty attached to a probability of 0.68 from a probable error (attached to a probability of 0.5). For the present paper, these statistical nuances are not relevant to my argument. When discussing Birge’s account, I will use the term “probable error”, in order to remain faithful to the original text, while applying the conversion factor when comparing it to later uncertainties. Overall, I will consider “probable error” and “measurement uncertainty” as depicting the same general idea, even though they correspond to slightly different quantitative realizations.

³³Birge (1929), p.9.

³⁴Mercier (1924)

³⁵Rosa and Dorsey (1907)

an electric charge in both systems, Rosa and Dorsey were able to derive $c = 299\,790$ km/s. After exchanging with the authors, Birge evaluated the corresponding probable error as 10 km/s.

Birge restricted his adjustment of c to the three previous results, which are represented on fig.2. The figure illustrates in particular that although Rosa and Dorsey's result was "in beautiful agreement with Michelson's recent value of c "³⁶, there was, on the contrary, a significant disagreement between Mercier's result and the two other ones. Indeed, Mercier's and Michelson's values are different; but more compellingly, they cannot be reconciled within the margin of their probable errors (qualitatively: on the graph, Mercier's result does not overlap with any of the other ones). Evaluating the significance of a disagreement between two values is one of the primary roles of measurement uncertainty.

Therefore, Birge faced here a problem. Let us recall, at this stage, that Birge's intention was to draw a "best value" of c out of the collected data. The case of c proved difficult because it consisted in aggregating significantly contradictory information. Did it imply that one group of results had to be discarded? Conversely, was it better to try to accommodate all the results? From a general viewpoint, solving this kind of problem introduces a tension between two methodological choices: arbitrage versus compromise. As I will show in the next section, this tension grew more critical in the decades following Birge's first adjustment.

On the one hand, one could claim that one group of results is either in error or unreliable, and decide to put it aside: this is the arbitrage choice, leading to a selection in the data. But the adjuster could very well have no clear-cut indication about the compared reliability of the results. Thus, on the other hand, he could try to find a way to combine all the results as fairly as possible: this is the compromise choice. Applied to Birge's adjustment of the speed of light, the compromise choice could have been performed, for instance, by expanding the measurement uncertainties associated to Mercier and Michelson's results in order to make them compatible. The adjusted value of c would then have covered every input result, and the resulting uncertainty would have been larger than the uncertainty attached to any of the individual values collected by Birge. This compromise attitude is illustrated in fig.2-b.

Birge himself did not explicitly address this issue, but it is very clear that he adopted the arbitrage choice. He did not try to accommodate the input values, but resolved the discrepancy by excluding Mercier's value, which he felt was less reliable. Indeed, he immediately commented Mercier's result, stating that "because of uncertainties in the theory, the error [in Mercier's value] may be (and according to Michelson's result actually is) much greater"³⁷ than that reported by Mercier himself. So, to support his choice of selection, Birge appealed to "uncertainties in the theory" and did not rely on the amplitude of the probable errors (even though they were also in favour of Michelson's result). The reliability of the uncertainties themselves was at stake, and Birge's remark indicated that the evaluation of the probable error in Mercier's result was likely to be incomplete. Indeed, with his comment on Mercier's result, Birge was in fact emitting an *external* judgement about the quality and reliability of the measurements at

³⁶Birge (1929), p.10.

³⁷Birge (1929), p.9. Although Mercier himself did notice the disagreement between his result and the previous determinations of c , he was unable to explain it (Mercier, 1924, p.178). Mercier only suggested ideas of possible sources of error, without being able to investigate them further. Therefore, his evaluation of the probable error did not include what Birge referred to as "uncertainties in the theories".

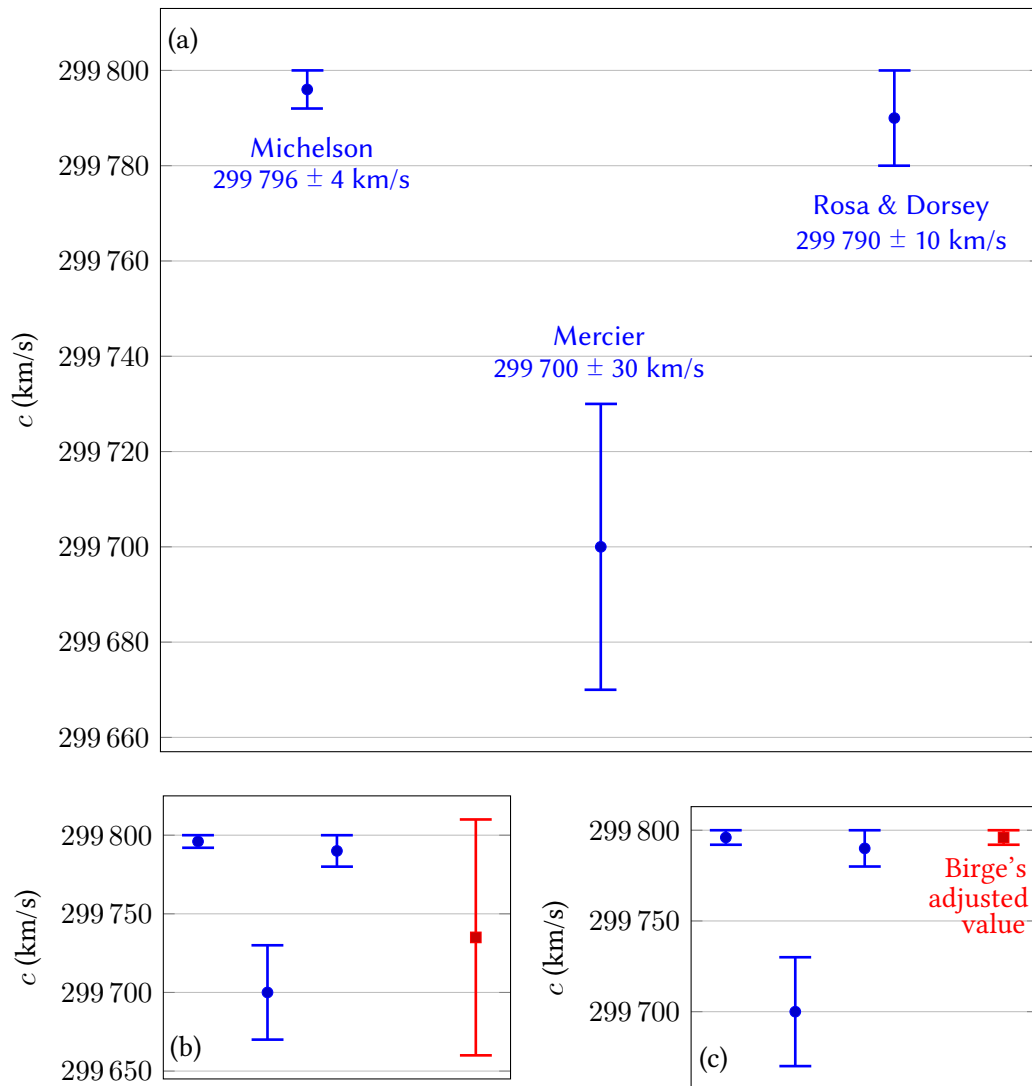


Figure 2: Plot of the three determinations of the speed of light in vacuum c collected by Birge in 1929 (figure a). The results are represented by a segment whose centre is the measured value and whose half width is the probable error attached to the result. Roughly, two results are in good agreement if the segments representing them are overlapping. Mercier's result is significantly different from the two other ones. On figure (b) is shown a rough example of a possible conservative adjustment (result on the right-side, with the central value represented by a square). In that case, the adjusted value is chosen in order to accommodate all input values of the adjustment. In particular, the probable error is expanded so that it covers the probable error of all input results. Birge did not choose this option. Instead, he excluded Mercier's value from the pool. Figure (c) shows Birge's adjusted value, which is, eventually, Michelson's value (result on the right-side, with the central value represented by a square). This is an original figure constructed out of the tables provided in Birge's 1929 article.

hand, which he was using in order to select the appropriate data for the adjustment. Accordingly, he had doubts about the accuracy of Mercier’s result and did not consider its probable error as a relevant indication of how confident he could be in the result.

Having excluded Mercier’s result, Birge calculated an adjusted value of c out of the two other results. Birge explained that the contribution of Rosa and Dorsey’s measurement to the final adjusted value was negligible. Therefore, he decided that the adjustment outcome could be approximated as Michelson’s value – which is not to be interpreted as a dismissal of Rosa and Dorsey’s result.

In the process described here, measurement uncertainty was only used to evaluate the significance of the discrepancy in the input data, but *not* to resolve this discrepancy. I will come back to this point later on, after having discussed a second example from Birge’s 1929 article: the case of the Planck constant h . I already emphasized the role of measurement uncertainty in the *comparison* of experimental data; Birge’s account of the Planck constant will enable me to highlight another important feature of measurement uncertainty, in the *combination* of measurement results.

3.3 The Planck constant in 1929

Later in his article, Birge identified six different methods by which h had been determined with high accuracy in 1929³⁸. Each of these six methods exploit a physical phenomenon in which h is involved, setting up a quantitative derivation of h by linking it to some measurable physical properties (wavelengths, electrical measurements, etc.). Moreover, these determinations involve a number of various fundamental constants already treated in the previous sections of Birge’s article – among which, in particular, the elementary charge e . In what follows, I will only provide a brief overview of the two first determinations of h described by Birge, in order to give an idea of how this was carried on. The exact content of each method is unnecessary to understand the core of the issues that I want to tackle here.

The first method relies on a theoretical connexion between h and the “Rydberg constant for an infinite mass” R_∞ , through the following equation (expressed in CGS units):³⁹

$$R_\infty = \frac{2\pi^2 e^5}{h^3 c^2 e/m} \quad (1)$$

where e is the elementary charge (the absolute value of the charge on an electron), and e/m is the ratio of the elementary charge over the mass of the electron (also called the “specific charge of the electron”⁴⁰). If the four constants c , e , e/m and R_∞ have been determined by specific independent means, then h can be derived from this equation. This was indeed the case in Birge’s account (R_∞ involving some more constants, themselves discussed in earlier sections of Birge’s article). By gathering all these values and their corresponding probable error, Birge

³⁸Birge (1929), section M, pp.48–28.

³⁹The equation would have a slightly different expression if written in the International System of Units.

⁴⁰Birge (1929), p.43.

was able to calculate a value for the Planck constant⁴¹: $h = (6.547 \pm 0.011) \cdot 10^{-27}$ erg.sec.

The second method of determination was due to the future Nobel Prize Ernest O. Lawrence, at Berkeley, who improved a classical experiment by Franck and Hertz in which an electron beam is driven across an atomic vapour – in that case a mercury vapour – until a ionization is observed⁴². Both the ionization potential V of the mercury and the wavelength λ of the light produced by this ionization can be measured and related to h by the following equation:

$$\frac{hc}{\lambda} = eV \quad (2)$$

This supports a determination of h as long as e and c have already been obtained by other means. Birge collected the value $h = (6.560 \pm 0.015) \times 10^{-27}$ erg.sec.

The four other cases are similar in their general form to the two methods described above, namely (i) computation from the Rydberg constant and (ii) ionization energies. Birge additionally listed determinations based (iii) on interference experiments of X-rays through a calcite crystal; (iv) on the photoelectric effect; and on measurements concerning black-body radiation phenomena related (v) to Wien's Displacement Law and (vi) to the Stefan-Boltzmann Law. Birge thus gathered six different values of h , illustrated on fig.3 (p.15) which displays a good overall agreement. Therefore, contrarily to the case of the speed of light, no arbitrage was needed to derive an adjusted value. The next stage was not a selection (as in the case of c), but an aggregation, which was done using a weighted average. According to a classical result of error analysis, the best average of the data is obtained by weighting each input value by the squared inverse of its probable error. By indexing the values of h as h_i and their respective probable errors by u_i , one comes to the conclusion:

$$h_{\text{adjusted}} \propto \sum_{i=1}^6 \frac{h_i}{u_i^2} \quad (3)$$

Equation (3) is a consequence of the probabilistic model of data used in the least-squares method, as Birge would undertake to demonstrate in a later article in 1932⁴³, and this equation remains today an essential tool for the adjustments. Moreover, once the adjusted value has been obtained, a similar method is used to calculate the probable error on the adjusted value. This led Birge to the following value and probable error for the adjusted Planck constant (also represented fig.3): $h = (6.547 \pm 0.008) \times 10^{-27}$ erg.sec.

The core of the combination law embodied in equation (3) rests on two ideas. First, all results are not equally reliable. Secondly, although the results are unequally reliable, it is still worth to combine them, provided the more reliable results are attributed a greater weight. Here, the (squared) probable error is taken *as a measure of unreliability of each input result*,

⁴¹Birge expressed the values in the CGS system of units. The unit for h was therefore the erg.sec, where an erg is a unit of energy that corresponds to 10^{-7} J.

⁴²Lawrence (1926)

⁴³Birge (1932a)

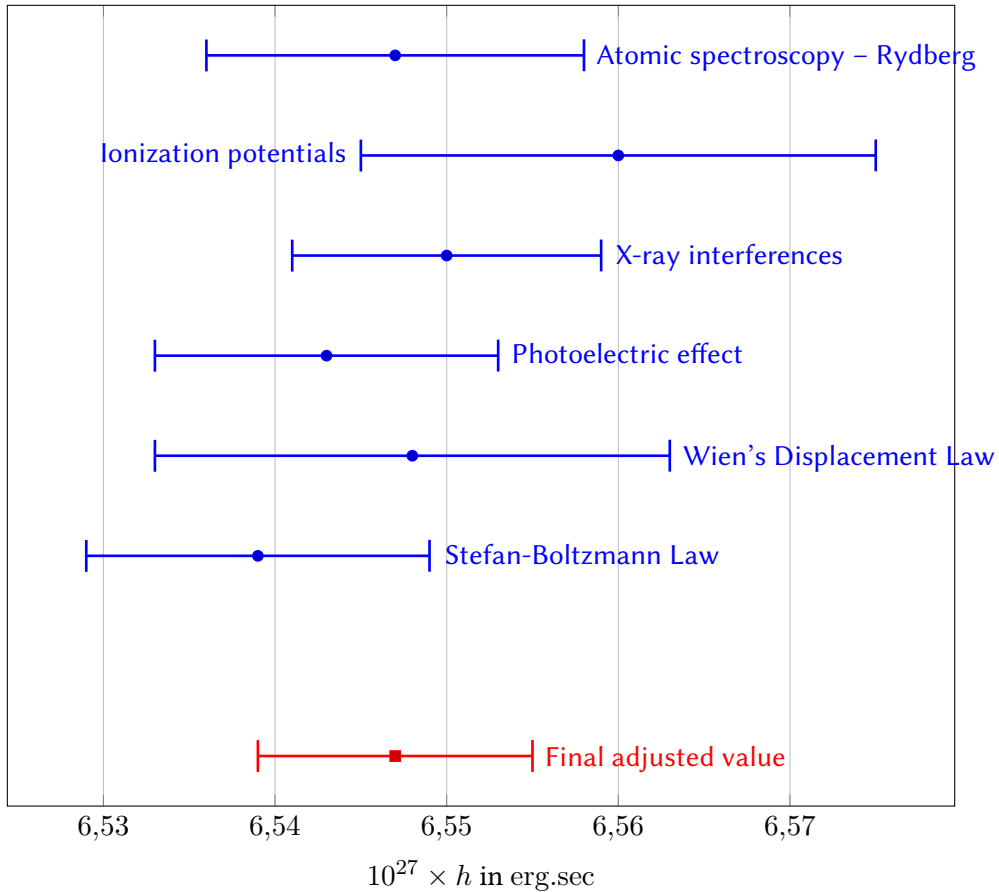


Figure 3: Six different measured values of h collected by Birge for his 1929 adjustment (central values represented by circles) and the adjusted value calculated out of Birge's adjustment (central value represented by a square). All six results are in good agreement, and the adjusted value is therefore a weighted average of the six individual values. The respective weights are decided upon the corresponding probable errors. The probable error on the adjusted value depends both on the dispersion of the six individual values and on the probable error on the elementary charge e , a constant that impacts the determination of h in all six methods. This is an original figure constructed out of the tables provided in Birge's 1929 article.

“reliability” being taken literally: the larger the probable error, the less the experimenter will effectively rely on the result in the combination. The weighted average is legitimate *only* on the basis of the hypothesis that no unknown so-called “systematic” error affects one or several of these input results. In the case of h , such a hypothesis was made because the results were in good agreement. Yet, this agreement did not guarantee that all sources of error had been identified.

3.4 Two roles of measurement uncertainty and a resulting tension

The previous examples demonstrate two levels of use of measurement uncertainty. At the first level, measurement uncertainty is used in comparisons, to quantify the significance of an agreement or a disagreement between measurement results. Such comparisons only test for consistency: a set of results is deemed consistent or not with regard to some hypothesis. For example, the disagreement between Michelson and Mercier’s values is incoherent with the hypothesis that they are both accurate measurement of the speed of light in vacuum. Depending on the outcome of this consistency test, different scenarios can be considered.

If the data are in good agreement, then they can be combined without complication, as was done in the case of the Planck constant. This is the second level of use of measurement uncertainty, which is used to weight the data in the combination. Since a lower uncertainty corresponds to a greater weight (and conversely), uncertainty may be taken here to evaluate the reliability of the result within the combination. I call this evaluation an “internal evaluation” of reliability, because measurement uncertainty is a part of the expression of the result itself.

If, on the contrary, the data significantly disagree, then the contradiction of the original data makes a genuine combination quite problematic. The disagreement has revealed an anomaly: something went wrong in the determinations of the results – for instance, the uncertainties might have been underestimated. In that case, the internal evaluations of reliability provided by the measurement uncertainties are not sufficient, and the disentanglement of the anomaly requires – at least in the arbitrage attitude adopted by Birge – to formulate some external judgement about the results: the adjusters have to apply an external evaluation of reliability. For example, such external evaluations may be judgements about the credibility of the experimenters themselves, about the quality of the measurement method used, about the background theory and model involved, etc. That is what Birge did when he excluded Mercier’s result based on “uncertainties in the theory”.

This whole analysis shows that the practice of adjustment revolves around a fundamental tension between *selection* and *combination*. Although an adjustment is aimed at aggregating knowledge about the physical constants, it sometimes appeals to a selection phase which by definition requires excluding information⁴⁴. This seems at first glance at odd with the basic idea of an adjustment, and as I will explain in the next section, Birge’s way to proceed has not always been supported by all scientists. The choice of an arbitrage between the input values summons some external, subjective judgement, and, as we are about to see, this generated a

⁴⁴In the general case, the determination of a constant may involve both selection, as displayed in the case of e , and combination through weighted averages, as displayed in the case of h .

debate that spread in the scientific community quickly after Birge's first efforts.

4 Accuracy versus precision in the adjustment of the constants

4.1 The perpetuation of Birge's initiative

In 1932, three years after his pioneering article, Birge published another adjustment focused on a restricted pool of four physical constants⁴⁵. He undertook this new adjustment in the context Arthur Eddington's theoretical prediction⁴⁶ of the value of the fine-structure constant α . An English hydrodynamicist, Wilfrid Bond, had published two articles in which he defended a good agreement between Eddington's determinations and recent measurements⁴⁷. Although Birge rejected Bond's conclusions⁴⁸, he was seduced by some elements of Bond's methodology, and decided to revise his own account of adjustments accordingly. He thereby improved his 1929 "method of successive approximations" which had at least one major drawback: it introduced an artificial hierarchy between the constants, depending on the order in which they were analysed. Bond's technique enabled adjusters to conceive of a simultaneous treatment of all the constants considered. Birge set up such a simultaneous treatment in the 1932 article and thus laid down the technical premises of the general method to be used in the future⁴⁹, while taking position against Eddington's predictions about the fine-structure constant.

The program outlined in Birge's 1929 and 1932 articles⁵⁰ was a success, quickly spreading in the community and being reinvested by several other physicists⁵¹. All these more or less individual attempts were to be consolidated in the end of the 1960s into an international collaboration piloted by the CODATA, which triggered a first adjustment in 1973, and a second one in 1986. During the third CODATA adjustment in 1998, it was decided to perform adjustments at a regular pace, every four years. This decision was made possible in particular by the progresses of computers: whereas Birge pointed out in 1932 that "the necessary calculations might easily take centuries to carry out"⁵², the execution of the adjustment algorithm itself is now a matter of seconds. Birge's individual initiative, which contained the germs of a collective study of the knowledge of the physical constants, and which had explicit standardization aims, turned into an institutionalized tradition of adjustments involving the community of physicists

⁴⁵[Birge \(1932b\)](#)

⁴⁶See [Kragh \(2003\)](#) for a history of the fine-structure constant including the debates about Eddington's determinations.

⁴⁷[Bond \(1930\)](#), [Bond \(1931\)](#).

⁴⁸[Birge \(1932b\)](#), p.231.

⁴⁹This led in particular to graphical analysis tools later known as "Birge-Bond diagrams".

⁵⁰Birge further perfected this program in multiple articles focused on one or a couple of specific constants, and then produced two general reviews in 1941 and 1945 ([Birge, 1941, 1945](#)). A later conference given in 1957 sums up his viewpoint on this activity ([Birge, 1957](#)).

⁵¹Let's mention, among others, the attempts of Dunnington in 1939 ([Dunnington, 1939](#)); of Cohen and DuMond in 1948, 1952, 1955 (with Layton and Rollett), 1959 and 1965 ([DuMond and Cohen, 1948, 1953; Cohen et al., 1955; Cohen and DuMond, 1958, 1965](#)); of Bearden, with Watts in 1951 and with Thomsen in 1957 ([Bearden and Watts, 1951; Bearden and Thomsen, 1957](#)); and also those of Taylor, Parker and Langenberg in 1969 ([Taylor, Parker and Langenberg, 1969](#)).

⁵²[Birge \(1932b\)](#), p.231.

at a large international scale, with the CODATA as a major epistemic authority.

The overall success of the activity of adjustments did not prevent criticism. In particular, the selection stage of the adjustment was the object of a recurring discussion until the 1970s – and remains today a topic of debate.

4.2 Selection issues

As early as 1948, Jesse W. DuMond and E. Richard Cohen pointed out that “the procedure of obtaining adjusted values by weighted averages of many different experiments [...] has been severely criticized in many quarters”⁵³. One of the principal sources of dissent in the community came from the issue of the selection of the input data.

As I illustrated with the case of the speed of light, Birge resolved disagreements within the pool of input data by an arbitrage about the data to be selected, and a similar attitude was adopted in later adjustments. In 1965, DuMond and Cohen dedicated a significant part of their adjustment report to an explanation and a justification of what they called the “rejection of data”⁵⁴. A disagreement between input data was taken as a trace of an unknown error affecting one of the results and preventing any combination of data, because it would then be reported on the other measurement results with which it would be combined.

However, excluding results is controversial. When Birge discarded Mercier’s value from the adjustment, he downgraded the result by suggesting that it could be less accurate or less reliable than what Mercier had himself claimed. Should Birge have excluded Michelson’s result instead? And if there were reasons to believe that Mercier’s result was not as good as claimed, why weren’t they included into Mercier’s own evaluation of the probable error? In principle, a disagreement between two results could only bring to the conclusion that at least one of them was incorrect, without telling anything about *which* one was incorrect. By selecting Michelson’s result, Birge had provided his own external evaluation of reliability, and *judged* that he had a satisfying answer to the data inconsistency. This appeal to judgement was emphasized by Birge⁵⁵ as well as by Cohen and DuMond, who pointed out that it “must involve, to some degree, the subjective element”⁵⁶. But then, what did they mean by “subjective” and how did they avoid the slope from subjectivity to arbitrariness? Either the adjusters had access to additional information (which could be their own insight), or they were just taking wild guesses, hoping for their choice to reveal correct in the future.

Eventually, are the selection choices fair? Are they fair from an epistemic point of view? Additionally, are they fair from an institutional point of view? These questions are at the heart of criticisms against the adjustment procedure. Interestingly, there is a sociological element to this: the selection process implies a reward and control system. For a research group, having its result incorporated within the effective pool of input data of an adjustment can be perceived as a validation of its work by the community – and conversely. As a consequence, the physicists

⁵³DuMond and Cohen (1948), pp.83-84.

⁵⁴Cohen and DuMond (1965), pp.577–584.

⁵⁵Birge (1929), p.6.

⁵⁶Cohen and DuMond (1965), p.539.

performing the adjustments are put in a privileged position from where they have to judge their peers' work even if they do not necessarily have new theoretical or experimental information to do so. One can wonder, then, how and by whom can such a task be accomplished properly and fairly. This became even more important once the CODATA was put in charge of the adjustments.

Selecting the input data has very concrete consequences: since inconsistent data are excluded from the pool, the uncertainty on the output adjusted values is reduced. And since the uncertainties are reduced, one could expect the adjustments to provide overconfident evaluations of the uncertainties of the adjusted values – and ultimately, end up with accidents in the history of the recommended values, as the ones observed about the Planck constant in fig 1 (section 2). Physicists quickly grew aware of this feature. Bearden and Thomsen mentioned it in a 1957 report⁵⁷. Barry N. Taylor, William H. Parker and Donald N. Langenberg also pointed out this tendency and connected it to the input selection of the data:

(I)t must be realized that the final values of the constants resulting from an LSA [*least-squares adjustment*] are generally obtained from a highly expurgated group of experimental input data. Most of the inconsistent or “bad” data are discarded [...] This censorship results in output values which are so consistent that they would appear inviolate. Such confidence is not justified!⁵⁸

A few years later, Taylor even explained that this feature “has been the rule rather than the exception”⁵⁹. In 1984, reflecting on the history of the values of physical constants, Max Henrion and Baruch Fischhoff named “surprise index” the rate of these accidents. They claimed that “examination of historical measurements and recommended values for the fundamental physical constants shows that the reported uncertainties have a consistent bias towards underestimating the actual errors”, and concluded that “the most common problem is overconfidence”⁶⁰.

Why would physicists formulate such uncertainties if they knew that “such confidence is not justified”? This question was explicitly discussed during an international conference, organized in 1970 in the United States at the National Bureau of Standards (today the National Institute for Standards and Technology). I propose to analyse some parts of the debates, which illustrate the fundamental tension between arbitrage and compromise, and between selection and combination, that I already pointed out – and which result in a tension between safety and precision.

4.3 Bender’s criticism: uncertainty as a safe accuracy claim

In 1970, a few years after the creation of the CODATA, scientists from various fields of precision physics gathered to discuss the latest progress in the measurement of the physical constants,

⁵⁷Bearden and Thomsen (1957), pp.273–274.

⁵⁸Taylor, Parker and Langenberg (1969), p.379.

⁵⁹Taylor (1971), p.497.

⁶⁰All quotes are from Henrion and Fischhoff (1986), p.791.

and ended up raising the following question: “Should least-squares adjustments of the fundamental constants be abolished?”⁶¹ The debate was introduced by the US astrophysicist Peter Bender, and appeared as a direct consequence of the issues related to data selection. Bender first stated the problem encountered in the treatment of discordant data:

When the uncertainties assigned by the experimenters seem reasonable on physical grounds but the results clearly disagree, one point of view is that there is no meaningful way in which we can assign a most probable value and an uncertainty.⁶²

Although he did not think that this problem should lead to abandon least-squares adjustments, Bender was not satisfied with the way it was taken care of in the usual adjustment accounts:

Up until now people doing evaluations have felt compelled to make crisp yes-or-no decisions on how to handle discrepancies, and no allowance has been made in the final uncertainties for the possibility that the wrong choice was made.⁶³

In Bender’s view, the selection stage had to be amended in a way which can be briefly summarized as follows. Bender’s main idea was that three options could be thought of in reaction to two results being discordant: (i) the first result was incorrect; (ii) the second result was incorrect; (iii) the uncertainties were too small. Options (i) and (ii) called for a resolution by selection, and option (iii) called for estimations of expanded uncertainties that would cover the initial discrepancy. Bender underlined that when there was no *a priori* reason to favour one option over the other ones, one had to make a compromise between all the solutions – this is where his account deviates from the traditional appeal to a selection stage. He suggested attributing a probability to each option: “I would have [the evaluator] estimate probabilities for each of the hypotheses being correct”⁶⁴. The final, adjusted result could then be calculated as a weighted average of the outcomes that were obtained through each one of the solutions to options (i), (ii) and (iii), the weights being taken as the probabilities of each option. Bender emphasized that it enabled judgement to come by degrees – by finding compromises between discrepant data – and not as a binary selection.

Bender’s solution may look like a mere technical trick. However, its realization does not matter as much as its conceptual rationale: the discussion concerns the very function of measurement uncertainty. Bender was emphasizing that one of the main aims of an adjustment was to provide the most correct values as possible. As he stated,

[Most users] would rather use values which [...] someone who has studied the problem carefully feels are as close as one can come to the *true ones* at the time of their adoption. [...] The major question is how to make the value and uncertainty that are chosen as unbiased as possible.⁶⁵

⁶¹Langenberg and Taylor (1971), pp.493–525.

⁶²Bender (1971), p.493.

⁶³Bender (1971), p.494.

⁶⁴Bender (1971), p.494.

⁶⁵Bender (1971), pp.493–494 (emphasis is mine).

He repeated this desire in his conclusion:

(A) reasonably *fair* estimate of the actual uncertainty in the result [is] the most important goal. To summarize, the basic desire is to find a way to avoid having the quoted uncertainty in the results be *systematically too small* because of throwing out data.⁶⁶

It must be noted that Bender's aim was not to *improve* the accuracy of the measurements (a task devoted to the experimenters themselves): his concern was about the *assessment* of accuracy in measurement. Bender was trying to prevent cases in which apparently solid results would prove incorrect in the future – *i.e.* cases in which the true value of a target quantity would turn out, under further research, to lie outside of the uncertainty interval. Rather than applying an input selection to the data, his proposal consisted in taking into account all (reasonably) possible outcomes, thus producing larger uncertainties more likely to encompass the actual measurement error. Such an improvement should then be observable in the reduction of the long-run “surprise index” in the values of the physical constants. In other words, Bender wanted measurement uncertainty to be a safe evaluation of the measurement accuracy (closeness to the truth).

This approach may be further examined in the light of the “secure evidence framework” recently developed by Kent Staley⁶⁷. Staley defines the “security of the evidence that the data provide in support of the measurement result”⁶⁸ as the resistance of “evidential conclusions” to the “error scenarios” that are possible to the knowledge of a given epistemic agent. The securing of evidence requires the pursuit of different “strategies”, among which a “weakening strategy”⁶⁹ which consists, applied to the present case, as the extension of uncertainties to make results “immune to otherwise threatening error scenarios”. In the case of the adjustments, the discrepancy in the input data demonstrates the possibility of certain error scenarios (even though the causes of the errors are not known). Choosing to filter the input data amounts to discarding these specific error scenarios. Bender's account, on the contrary, is a “weakening strategy” that advocates to include these error scenarios in the evaluation of uncertainty, thus increasing the security of the adjusted values at the cost of their decreased precision. Bender is arguably concerned with security only, but I am using here the word “safe” instead of “secure” for two reasons. First, it was used by the physicists themselves, as will show a later quote by Cohen and DuMond, a use that recurs in a recent metrology document⁷⁰. Secondly, being relative to the knowledge of an agent, “security” is an epistemic concept. In the context the problem of unknowability, my use of “safe” refers here to the result being unlikely to be proved incorrect in the future, with regard to known *or unknown* error scenarios.

Bender's standpoint fits into the traditional account of measurement that I sketched in section 2: the value after the \pm sign (today's measurement uncertainty) is conceptualized as a

⁶⁶Bender (1971), p.494 (emphasis is mine).

⁶⁷Staley (2014, 2017)

⁶⁸Staley (2014), p.3.

⁶⁹Staley (2014), p.16.

⁷⁰Joint Committee for Guides in Metrology (JCGM) (2008), p.54.

safe accuracy claim – accuracy being understood as closeness to the truth. As I will now show, not all scientists concurred on such a position.

4.4 Taylor’s answer: uncertainty as insensitivity

Bender’s standpoint introduced the panel of the 1970 conference and stimulated various answers. Among these, Taylor’s commentary is particularly insightful: it is the testimony of the core principles of the main approach of the adjustments inherited from Birge. Taylor drew on his own experience: in 1969, he had performed an adjustment with Parker and Langenberg, in a slightly specific context, as the recent discovery of the “Josephson effect” had opened new theoretical and experimental considerations about the physical constants⁷¹.

Taylor challenged Bender’s claim that the main objective of adjustments was to provide unbiased values of the physical constants. In fact, he considered the production of final recommended values as “the *least* important result of a constant adjustment”⁷². From his viewpoint, adjustments were certainly useful, but “perhaps not for the most obvious reasons”⁷³. Instead, he insisted on the epistemic virtues of the adjustments as a *process*. He emphasized that the main interest of the adjustment might not be found in the final values themselves, but in the overall procedure:

The challenge to prove the constants adjuster incorrect, thereby forcing a new adjustment, is much more stimulating than the challenge to obtain a number which will just be thrown into the “pot”⁷⁴

According to Taylor, the procedure had several other merits than just producing a set of recommended values, among which (i) triggering a critical analysis of recent experiments and theoretical developments, (ii) leading to commonly accepted standardized methods of uncertainty analysis, (iii) supporting the discovery of experimental and measurement errors, and eventually (iv) providing a general consistency test of the physical theories in use. Another member of the conference panel, Thomsen, backed this viewpoint, stating that “(t)he critical review is, of course, the most important aspect; the adjustment itself is a more routine procedure”⁷⁵. Taylor almost discarded the importance of the recommended values themselves. He insisted on the fact that there was no specific category of user that would really need these values with the level of accuracy provided. As he stated, “most workers do not really require the last few decimal places in the values for the constants” while, at the same time, “those scientists who really need to use the last decimal place will not be content just to take numbers out of the table but will go to the original article”⁷⁶. This led Taylor to the following conclusion:

Since the majority do not particularly care what numbers they use, adjustments

⁷¹Taylor, Parker and Langenberg (1969)

⁷²Taylor (1971), p.495 (Taylor emphasizes).

⁷³Taylor (1971), p.495.

⁷⁴Taylor (1971), p.496.

⁷⁵Thomsen (1971), p.503.

⁷⁶Taylor (1971), p.496.

should be geared to the small but more important minority who do and who need the most useful and stimulating numbers they can get their hands on.⁷⁷

Taylor could not adhere to Bender's proposition because he did not share his wish for a compromise between sets of discrepant data, and more compellingly because this proposition was generating overly cautious measurement uncertainties. To Taylor,

[Expanding errors] is an admission of defeat – [...] that the only way out is to assume that all the data is bad. It therefore throws away information by making quantities more unknown than they actually are.⁷⁸

With this position, Taylor was drawing on ideas already expressed by Cohen and DuMond in their 1965 adjustment report:

The idea that an overestimate of error “for safety” is somehow more laudable or virtuous than an effort to be as accurate as one can with as little bias as possible in either magnifying or minimizing the standard deviation estimate, is somehow deplorably prevalent. We ask, for whom is such an overestimate “safe”? Certainly not for the general scientific community who wish to use the result. For them it is a concealment of the true facts regarding the results of the measurements. For the author of the result it may appear to be “safe” in the sense that at some later date his measurement might be less likely to be contradicted by later work; but even for him this unworthy timidity may be an illusory safety for, because of the unwarranted exaggeration of his error estimate, a crucial discrepancy, which might otherwise reveal some basically important new fundamental fact, may have been buried and lost forever as far as his reputation is concerned. For courageous men of science there is only one “safe” refuge, the plain, unvarnished truth as to their methods and results.⁷⁹

Eventually, having anomalies in the evolution of the values of the constants through time was not considered as a critical issue. As Taylor, Parker and Langenberg stated:

[The graphical representation of the evolution of our knowledge of the fundamental constants] illustrates our contention that no set of fundamental constants should be taken as Gospel truth. While we may hope that the present adjustment brings us closer to that truth, realism compels us to recognize that further significant changes in our knowledge of the constants may well take place. Goethe might well have been speaking of this when he wrote “Es irrt der Mensch, solang' er strebt.” [*man errs as long as he strives*]⁸⁰

The final quote, from Goethe's *Faust*, reveals how they conceive of the virtues of the adjustments. Instead of focusing on the up-to-date, immediate closeness-to-the-truth accuracy of the recommended values, the general adjustment process is rather oriented towards the possibility of a progress *in the future*. As the authors said so themselves in their 1969 adjustment report:

⁷⁷Taylor (1971), p.497.

⁷⁸Taylor (1971), p.496.

⁷⁹Cohen and DuMond (1965), p.541.

⁸⁰Taylor, Parker and Langenberg (1969), p.478.

Measurements of the fundamental physical constants to ever greater levels of accuracy are important, not just because they “add another decimal point” and provide us with a more consistent set of constants to work with, but because they *may lead* to the discovery of a previously unknown inconsistency or the removal of a known inconsistency in our physical description of nature.⁸¹

By this statement, the authors expressed that it was not worth to value safety over precision: were measurement uncertainties to be artificially expanded for the sake of unbiased results *in the present*, identifying the unknown errors would be made trickier.

This is what leads me to the main claim of this article. While Bender was concerned with a safe assessment of the accuracy of the recommended values at the time they were published, Taylor and others were conceptualizing accuracy as something that was to be reached through the constant improvement of results. The conception drawn by Taylor and others leads to an understanding of uncertainty and accuracy that differs from the traditional account of measurement. The aim of an individual physicist is not to guarantee the accuracy of a measurement result: the problem of unknowability makes it an illusory and impossible requirement. The issue is rather shifted to the continuous quest of identification and correction of measurement errors. Although the true value of the measurand, and with it the amplitude of the final error itself, remain ultimately unknowable, scientists are still entitled to believe that they can find explanations to the anomalies and identify causes of previously unknown errors. This constitutes an actual scientific progress and it is fair to claim that the elimination or the correction of such errors would then constitute an increase in accuracy. This eventually relates to a general epistemological concern already underlined by Hon⁸²: it is much more interesting to trace back an experimental anomaly to its source in order to identify an error, discover new phenomena or modify the physical theories, with a real cognitive input, than to simply cover the anomaly. With this account, physicists continue to attach a value to truth, but acknowledge the problem of unknowability and formulate their own answer while carefully avoiding scepticism: here, the positive content of knowledge is embodied in what has been learned from error correction, which itself gets scientists closer to the truth, even if they cannot evaluate how close they are from the truth.

This study buttresses an idea that Nadine de Courtenay and I recently defended⁸³. As we argue, accuracy should not be contemplated from a static and individualistic point of view, as is usually done, but rather reconsidered from a dynamical and collective perspective, in tune with an ideal of scientific progress. Acquiring knowledge about the constants and securing an increasing accuracy in measurement cannot be achieved by an individual scientist (or even a local group of scientists) at present; it is a task that can only be carried out by a *community* of scientists *in the long-run*. This task requires an adequate social and epistemic structure

⁸¹Taylor, Parker and Langenberg (1969), p.377 (my emphasis).

⁸²“From an epistemological perspective, it is worth to inquire into the source of the error and not so much to examine the mathematical features of the error and the means of calculating it away – the causal feature being of a higher interest here than the pragmatic one”, Hon (2003), p.183.

⁸³Mirowski (1994), p.580. The collective issues related to the assessment of uncertainty in measurement have been raised in de Courtenay and Grégis (2017)

of scientific activity because the proper functioning of the process cannot ever be taken for granted. It is necessary to make sure at the collective and institutional levels that unknown errors are being actively tracked down (for example by promoting diversity in the measurement methods and avoiding methodological biases such as the “bandwagon effect”⁸⁴). This illustrates in particular the crucial role of an institution such as the CODATA in the contemporary history of science⁸⁵. The collective issues related to the assessment of uncertainty in measurement, and more generally the social aspects of measurement in science constitute a promising but still largely unexplored addition to the recent developments of the epistemology of measurement. As also argued in de Courtenay and Grégis (2017), social epistemology may in particular prove especially relevant to the philosophy of measurement.

In that context, measurement uncertainty does not represent the (unknown and unknowable) metaphysical accuracy of the measurement results, but it rather indicates the *sensitivity* with which the results are supposed to enable the testing of hypotheses, theories, or with which they can be compared to each other⁸⁶. Taylor explained it himself with the following metaphor:

Our philosophy in the 1969 adjustment was to provide the best possible cutting tool for those workers who needed it most and who were prepared to use it in full knowledge that care was necessary to avoid cutting themselves.⁸⁷

“Sensitivity” is used here in a specific sense, as recently defined by Pierre-Hughes Beauchemin⁸⁸ and Kent Staley⁸⁹ in their respective studies in high-energy physics, namely the capacity of a result to actually tell something meaningful about the physics under examination. When measurement uncertainties are large, the result is quite insensitive to the physics of interest and will therefore produce very little new knowledge. As Staley demonstrated, there is a trade-off between security and sensitivity: this trade-off is instantiated in the case of the adjustments. Bender and Taylor’s approaches may indeed be interpreted as two different epistemic strategies to manage security, by appealing or not to the selection of input data. To put it in the words of Staley and Beauchemin, Taylor’s strategy favours sensitivity over security.

5 Measurement uncertainty: objective and epistemic stances

I conclude this discussion by sketching a systematic account of measurement uncertainty, drawing on the previous case study. I will present two typical characterizations of measure-

⁸⁴Mirowski (1994), p.579.

⁸⁵As Mirowski pointed out, Birge’s work revealed the “ineluctable social character of error and precision”, Mirowski (1994), p.576. Mirowski observes that his “historical sketch of the estimation of the physical constants has been one of the construction of more and more elaborate institutions and social structures for the production and allocation of quantitative error”, Mirowski (1994), p.580.

⁸⁶As Terry Quinn explains, the question the adjusters most often raise is not so much “how far away from the truth are we?” but “how far away from our models are we?” (private communication). Understanding measurement uncertainty in terms of sensitivity goes in that direction.

⁸⁷Taylor (1971), p.497.

⁸⁸Beauchemin (2017)

⁸⁹Staley (2017)

ment uncertainty, discuss their limits, and suggest improvements – in a way that may reconcile them.

5.1 The objective stance

As was seen before, measurement uncertainty is traditionally interpreted as the evaluation of the “final error” affecting a measurement. What I designated as the traditional account of measurement in metrology can be summarized as such:

P.1- Measurement uncertainty is a (metaphysical) accuracy claim: it is an evaluation of the closeness of the measured value to the true value of the target.

This counts as an “objective” stance because it concerns the relationship of the result with regard to the true value of the target quantity, independently of the cognitive agent (no reference to belief or knowledge).

(P.1) is a qualitative statement but it may be associated to a quantitative criterion by appealing to probabilities – typically by using the frequentist interpretation of probabilities in which probabilities are long-run relative frequencies of events:

C.1- (success rate criterion) Uncertainty evaluation procedures are valid if and only if, for a high number of measurements, the frequency with which the final measurement errors are contained within the respective uncertainty bounds is greater than the probability associated with the uncertainty statements.

In that case, what matters is that the frequency of correct or incorrect claims has to be controlled in some way. In particular, measurement uncertainties are underestimated when the results turn out incorrect too many times. This reflects what Bender or also Henrion and Fischhoff were expressing in their criticism on the adjustments. And indeed, according to this interpretation, a high “surprise index” is not acceptable.

Admittedly, criterion (C.1) is falsifiable, since the “surprise index” is relevant in indicating violations of the criterion. However, the index does not demonstrate the actual success of the measurements – because, once again, of the problem of unknowability. One may only infer from a low surprise index that the results are consistent with the hypothesis of no unknown systematic error, but it does not prove this hypothesis. The problem of unknowability, in fact, makes criterion (C.1) impossible to verify, and in turn makes proposition (P.1) inoperative. Lately, certain metrologists and statisticians have tried to circumvent the issue by advocating simulating uncertainty evaluation processes by calibrating them on well-known “dummy” targets for which the true value is fixed by definition⁹⁰: this calibration would then justify further confidence in the outcomes of the process when applied to unknown targets. This solution, however, essentially concentrates on statistical methods and cannot handle the problem of possible unidentified and unknown errors.

So eventually, criterion (C.1) does not achieve what it was intended for and cannot be used to make sense of proposition (P.1). This is where Taylor’s account of accuracy shows its merits:

⁹⁰See in particular [Willink \(2013\)](#).

it acknowledges that unknown errors are unavoidable and that the epistemic virtue of the adjustments resides mainly in the possibility to unravel them step by step, thus continuously improving the knowledge about the physical constants.

Of course, uncertainty analysis draws on an evaluation of sources of error, but only of the ones that have been identified at a certain point (variability of the experimental data, experimental biases, etc.). Proposition (P.1) is acceptable only under the additional condition – again unverifiable – that all sources of errors have been identified. This is precisely what the *Guide to the expression of uncertainty in measurement*, the current reference metrology guide about uncertainty evaluation, states:

The exact error of a result of a measurement is, in general, unknown and unknowable. [...] 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 [*the target quantity*] and that its combined standard uncertainty is a reliable measure of its *possible* error.⁹¹

As a result, I suggest a way to improve the proposition (P.1) as:

P.1bis- *Measurement uncertainty is a (model-based) inference about the summed amplitude of the identified sources of error.*

This is where my account somewhat meets Tal’s considerations on uncertainty as an instance of predictive accuracy: here, measurement uncertainty is an inference in a given assemblage of models and theories. It should be noted that identifying the *source* of an error is not necessarily understood here as knowing its *cause*: for example, an uncertainty budget generally includes statistical errors, which are generally not explained⁹², but which can be identified in the statistical dispersion of data.

Proposition (P.1bis) aims at improving the characterization of measurement uncertainty in the objective viewpoint, but other scientists have adopted quite a different perspective, focused on belief and knowledge.

5.2 The epistemic stance

As I argued, the main pitfall of the objective stance is its focus on “truth” and “true value”. A recent trend in metrology consists in characterizing measurement uncertainty in terms of knowledge and (rational) belief – adopting an epistemic stance. The metrologist and statistician Walter Bich sums it up by stating that “uncertainty can be viewed as the logical reciprocal of state of knowledge”⁹³, a view shared by many scientists, especially in some statistical circles⁹⁴. The position might be summarized as such:

⁹¹Joint Committee for Guides in Metrology (JCGM) (2008), p.51 (emphasis in the original text).

⁹²As Bogen and Woodward (1988) have rightfully explained.

⁹³Bich (2012), p.2155.

⁹⁴I cannot avoid mentioning here that behind this discussion lies a very classical debate about the interpretation of probability. The objective stance tends to be popular for frequentist statisticians, whereas Bayesian statisticians will

P.2- Measurement uncertainty is a knowledge claim: it expresses what the experimenter⁹⁵ believes (alternately: knows) about the target quantity.

This mostly corresponds to the definition suggested for the upcoming fourth edition of the *International Vocabulary of Metrology*, still in preparation⁹⁶. Again, one could associate a probabilistic criterion to proposition (P.2), stipulating that an uncertainty statement (attached to a given level of credibility represented by a probability) is adequate if the experimenter is ready to bet on the validity of the result with a gain calibrated on the corresponding probability. Such an idea was already formulated during the 1970 conference⁹⁷. In this case, the fair estimation of a measurement uncertainty is not related to the result being correct or not. Instead, a scientist underestimates the uncertainty of her result when she does not actually believe what the result tells – this is very different from the objective stance.

Proponents of the epistemic stance praise it because beliefs or knowledge are accessible to us, and can be revised with further information. As a consequence, this interpretation is not endangered by the problem of unknowability. But again, it does not correspond to what practice shows. Scientists certainly do not found their trust only on measurement uncertainty, which, I have argued, only provides an internal evaluation of reliability. A recent example completes this observation. When, in 2011, the international collaboration “OPERA” announced the possible observation of faster-than-light neutrinos⁹⁸, many scientists, and even members of the OPERA collaboration themselves, did not really believe in the result. The most common guess was that this anomaly was very likely to be due to an unknown error in the experiment chain, which was confirmed one year later⁹⁹. Sergio Bertolucci, director of research at the CERN, told after the identification of the errors that “although this result isn’t as exciting as some would have liked, it is what we all expected deep down”¹⁰⁰. Questions of trust about the measurement were disconnected to the fact that the measurement uncertainty was very low – which it was – because it was possible that some unknown error had been overlooked. The overall conclusion is that measurement uncertainty can be said to represent a state of belief only when it is independently believed that no unknown systematic error affects the measurement. The problem of unknown error, which had apparently been evacuated by the epistemic formulation, actually finds a way back.

I would like to press the idea that measurement uncertainty does not describe what scientists believe, but rather what they are ready to believe *as long as* they have confidence in the

favour the epistemic one. However, I do not wish to enter into questions related to the methodology of uncertainty calculation, and I choose here to leave these kinds of debates in the background. More details on the statistical debates are given in [de Courtenay and Grégis \(2017\)](#).

⁹⁵“The experimenter” may refer to an individual or to the community at large: variations of this conception can be conceived according to whether one adopts a personalist or intersubjective position about uncertainty.

⁹⁶“Possible new definition of measurement uncertainty [...] parameter (or parameters) characterizing how well the (essentially unique) true value of the measurand [*the target quantity*] is believed to be known”, [Ehrlich \(2014\)](#), p.S150.

⁹⁷See in particular [Thomsen \(1971\)](#).

⁹⁸[Adam, T. et al. \(2011\)](#)

⁹⁹[CERN \(2011\)](#)

¹⁰⁰[CERN \(2011\)](#)

result as a whole (which includes, for example, trusting the experimenters who published it). I therefore suggest two (qualitative) alternatives to proposition (P.2):

***P.2bis- (in terms of belief)** Measurement uncertainty expresses what the experimenter should rationally believe about the target quantity if she supposed that there was no remaining unknown error.*

***P.2ter- (in terms of knowledge)** Measurement uncertainty expresses what the experimenter knows that she does not know, based on the sources of uncertainty that she has identified. The measurement uncertainty makes no claim about the possible unknown unknowns - ie unidentified sources of error.*

Both propositions (P.1) and (P.2) are fundamentally compromised by the general problem of unknown errors. Alternate propositions (P.1bis), (P.2bis) and (P.2ter) aim to acknowledge this problem. They narrow the gap between the objective and the epistemic stance, since “objective” uncertainty is not directly related to accuracy and “epistemic” uncertainty does somehow relate to measurement error. The amended propositions are connected, depending on the hypothesis that there remains no additional unknown error, and eventually, they may even be understood as two faces of the same coin.

6 Conclusion

The adjustments of the physical constants are a social and epistemic platform where scientists collectively discuss issues about error, uncertainty and accuracy. They thereby offer an interesting viewpoint on the problem of unknowability. I showed how tensions appeared between selection and combination, safety and precision, and ultimately between (metaphysical) accuracy and measurement uncertainty. The path followed by Taylor and other actors¹⁰¹, favouring precision (as well as challenging decisions) over the safety of the published results, is a way to try to answer the problem of unknowability while avoiding a sterile scepticism.

In the case of a knowledge-oriented research, Taylor’s conception may hold: accuracy is a guiding ideal, not to be achieved in the actual, but as a dynamical feature of scientific progress. Safety is not an imperative concern when no urgent matter crucially relies on the statements formulated – for example on the last decimals of the physical constants. However, to the long-term concerns of metrologists can be opposed the more urgent uses of measurement results. The dynamical account of accuracy is hardly relevant to an experimenter who would like to test the toxicity of a sample of drinking water, to evaluate the resistance of an aircraft wing, etc. For users who need to act and take short-term decisions – users who are taking *risks* – it is not enough to know that any biased measurement will be likely to be corrected in the future: they need assurance for what they are doing at the given time.

¹⁰¹The recent example of the proton radius, briefly sketched in [de Courtenay and Grégis \(2017\)](#), illustrated that the debate strives today in almost identical terms, and that as of today – but this might change one day – the general CODATA policy remains attached to Taylor and other’s conception.

Thus, depending on the context, users could benefit from a safe evaluation of accuracy. But at the same time, such a guarantee is illusory – the problem of unknowability will never cease to hold. Under this predicament, how can users trust and rely on forever uncertain results? How should they act with the possibility of remaining unknown errors? The answer most likely rests on acknowledging that these questions go far beyond mere uncertainty analysis. Typically, this requires them to delve into robustness analysis, risk management, quality control, and decision making under uncertainty – among other concerns – in which measurement uncertainty is only a tool. The whole inquiry includes social and pragmatic concerns, for which social epistemology might prove especially useful.

Acknowledgements

The author wishes to thank Nadine de Courtenay, Olivier Darrigol and Jonathan Regier for their reviews on earlier drafts of the paper, Shaul Katzir for his advice on the final stages of this article, and Yosef Cohen for his helpful grammar review – all remaining errors are mine. I am grateful to the participants of the Informal Aspects of Uncertainty Assessment workshop in Cambridge, United Kingdom in May 2016, for their insightful comments when a version of this argument was presented in its embryonic stage. I am indebted to David Newell and François Nez for having offered Nadine de Courtenay and me to be observers at several sessions of the CODATA Task Group on Fundamental Constants. Finally, comments from the anonymous referees and the editors were very valuable. 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

- ADAM, T. *ET AL.* (2011), Measurement of the neutrino velocity with the OPERA detector in the CNGS beam, *ArXiv*, pp.1–24.
- BEARDEN, J. A. and THOMSEN, J. S. (1957), A Survey of Atomic Constants, *Nuovo Cimento Supplemento*, 5, pp.267–360.
- BEARDEN, J. A. and WATTS, H. M. (1951), A Re-Evaluation of the Fundamental Atomic Constants, *Physical Review*, 81(1), pp.73–81.
- BEAUCHEMIN, P.-H. (2017), Autopsy of measurements with the ATLAS detector at the LHC, *Synthese*, 194(2), pp.275–312.
- BENDER, P. L. (1971), Handling of Discrepant Data in Evaluations of the Fundamental Constants, in LANGENBERG, D. N. and TAYLOR, B. N. (eds.), *Precision Measurement and Fundamental Constants*, pp.493–494, National Bureau of Standards. (Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970).

- 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.
- BIRD, A. (2007), What Is Scientific Progress?, *Noûs*, 41(1), pp.64–89.
- BIRGE, R. T. (1919), The Most Probable Value of the Planck Constant h , *Physical Review*, 14, pp.361–368.
- BIRGE, R. T. (1929), Probable Values of the General Physical Constants, *Reviews of Modern Physics*, 1(1), pp.1–73.
- BIRGE, R. T. (1932a), The Calculation Of Errors By The Method Of Least Squares, *Physical Review*, 40(2), pp.207–227.
- BIRGE, R. T. (1932b), Probable Values of e , h , e/m and α , *Phys. Rev.*, 40(2), pp.207–228.
- BIRGE, R. T. (1941), The general physical constants: As of august 1941 with details on the velocity of light only, *Reports on Progress in Physics*, 8(1), p.90.
- BIRGE, R. T. (1945), The 1944 Values of Certain Atomic Constants with Particular Reference to the Electronic Charge, *American Journal of Physics*, 13(2), pp.63–73.
- BIRGE, R. T. (1957), A survey of the systematic evaluation of the universal physical constants, *Il Nuovo Cimento*, 6(supp.1), pp.39–67.
- BOGEN, J. and WOODWARD, J. (1988), Saving the Phenomena, *The Philosophical Review*, XCVII(3), pp.303–352.
- BOND, W. N. (1930), The values and inter-relationships of c , e , h , M_p , m_o , G , and R , *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(67), pp.994–1003.
- BOND, W. N. (1931), The electronic charge, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 12(78), pp.632–640.
- BOUMANS, M., HON, G. and PETERSEN, A. C. (eds.) (2014), *Error and Uncertainty in Scientific Practice*, History and Philosophy of Technoscience Series, Pickering & Chatto.
- 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>
- CHANG, H. (2004), *Inventing Temperature: Measurement and Scientific Progress*, Oxford: Oxford University Press.
- COHEN, E. R. and DuMOND, J. W. M. (1958), Changes in the 1955 Atomic Constants Occasioned by Revision of μ_e/μ_0 , *Phys. Rev. Lett.*, 1, pp.291–292.

- COHEN, E. R. and DUMOND, J. W. M. (1965), Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965, *Reviews of Modern Physics*, 37(4), pp.537–594.
- COHEN, E. R., DUMOND, J. W. M., LAYTON, T. W. and ROLLETT, J. S. (1955), Analysis of Variance of the 1952 Data on the Atomic Constants and a New Adjustment, 1955, *Reviews of Modern Physics*, 27(4), pp.363–380.
- COHEN, E. R. and TAYLOR, B. N. (1973), The 1973 Least-Squares Adjustment of the Fundamental Constants, *J. Phys. Chem. Ref. Data*, 2(4), pp.634–734.
- de COURTENAY, N. and GRÉGIS, F. (2017), The evaluation of measurement uncertainties and its epistemological ramifications, in CHANG, H., MITCHELL, D. J. and TAL, E. (eds.), *The Making of Measurement*, special issue of *Studies in History and Philosophy of Science Part A*, pp.21–32.
- DELLSÉN, F. (2018), Scientific Progress: Four Accounts, *Philosophy Compass*, p.e12525. Forthcoming.
- DUMOND, J. W. M. and COHEN, E. R. (1948), Our Knowledge of the Atomic Constants F , N , m and h in 1947, and of Other Constants Derivable Therefrom, *Reviews of Modern Physics*, 20(1), pp.82–108.
- DUMOND, J. W. M. and COHEN, E. R. (1953), Least-Squares Adjustment of the Atomic Constants, 1952, *Reviews of Modern Physics*, 25(3), pp.691–708.
- DUNNINGTON, F. G. (1939), The Atomic Constants. A Revaluation and an Analysis of the Discrepancy, *Reviews of Modern Physics*, 11(2), pp.65–83.
- EHRlich, C. (2014), Terminological aspects of the *Guide to the Expression of Uncertainty in Measurement* (GUM), *Metrologia*, 51(4), pp.S145–S154.
- GRÉGIS, F. (2015), Can we dispense with the notion of “true value” in metrology?, in SCHLAUDT, O. and HUBER, L. (eds.), *Standardization In Measurement: Philosophical, Historical And Sociological Issues*, pp.81–93, Pickering & Chatto.
- HELMHOLZ, A. C. (1980), Obituary: Raymond Thayer Birge, *Physics Today*, 33(8), pp.68–69.
- HENRION, M. and FISCHHOFF, B. (1986), Assessing uncertainty in physical constants, *American Journal of Physics*, 54(9), pp.791–798.
- HON, G. (2003), The Idols of Experiment. Transcending the “etc. List”, in RADDER, H. (ed.), *The Philosophy of Scientific Experimentation*, pp.174–197, University of Pittsburgh Press.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008), *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*, Sèvres: JCGM.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2012), *International vocabulary of metrology - Basic and general concepts and associated terms (VIM)*, Sèvres: JCGM. 3rd edition. 2008 version with minor corrections.

- KANT, I. (1992), *Lectures on Logic*, The Cambridge Edition of the Works of Immanuel Kant, Cambridge University Press.
- KRAGH, H. (2003), Magic Number: A Partial History of the Fine-Structure Constant, *Archive for History of Exact Sciences*, 57(5), pp.395–431.
- LANGENBERG, D. N. and TAYLOR, B. N. (eds.) (1971), *Precision Measurement and Fundamental Constants*, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- LAWRENCE, E. O. (1926), The Ionization of Atoms by Electron Impact, *Physical Review*, 28(5), pp.947–961.
- MARI, L. (2003), Epistemology of measurement, *Measurement*, 34, pp.17–30.
- MARIAN, D. (2015), The correspondence theory of truth, in ZALTA, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2015 edition. Dernière consultation le 25 janvier 2016.
- MERCIER, J. (1924), De la synchronisation harmonique et multiple, *Journal de Physique et le Radium*, 5, pp.168–179.
- MICHELSON, A. A. (1927), Measurement of the Velocity of Light Between Mount Wilson and Mount San Antonio, *The Astrophysical Journal*, 65(1), pp.1–14.
- MIROWSKI, P. (1994), A Visible Hand in the Marketplace of Ideas: Precision Measurement as Arbitrage, *Science in Context*, 7, pp.563–589.
- MOHR, P. J., NEWELL, D. B. and TAYLOR, B. N. (2016), CODATA recommended values of the fundamental physical constants: 2014, *Reviews of Modern Physics*, 88, p.035009.
- NIINILUOTO, I. (2014), Scientific Progress as Increasing Verisimilitude, *Studies in History and Philosophy of Science Part A*, 46, pp.73–77.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) (1984), *International vocabulary of basic and general terms in metrology*.
- ROSA, E. B. and DORSEY, N. E. (1907), A new determination of the ratio of the electromagnetic to the electrostatic unit of electricity, *Bulletin of the Bureau of Standards*, 3(3), pp.433–540.
- ROWBOTTOM, D. P. (2010), What Scientific Progress Is Not: Against Bird's Epistemic View, *International Studies in the Philosophy of Science*, 24(3), pp.241–255.
- SMITH, G. E. (2010), Revisiting Accepted Science: The Indispensability of the History of Science, *The Monist*, 93(4), pp.545–579.
- STALEY, K. W. (2014), Experimental Knowledge in the Face of Theoretical Error, in BOUMANS, M., HON, G. and PETERSEN, A. C. (eds.), *Error and Uncertainty in Scientific Practice*, pp.39–55, Pickering & Chatto.

- STALEY, K. W. (2017), Securing the Empirical Value of Measurement Results, *The British Journal for the Philosophy of Science*, p. axx036.
- STIGLER, S. (1986), *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA: Belknap Press of Harvard University Press.
- TAL, E. (2011), How Accurate is the Standard Second?, *Philosophy of Science*, 78(5), pp.1082–1096.
- TAL, E. (2016), Making Time: A Study in the Epistemology of Measurement, *The British Journal for the Philosophy of Science*, 67(1), pp.297–335.
- TAL, E. (2017a), Calibration: Modelling the measurement process, *Studies in History and Philosophy of Science Part A*.
- TAL, E. (2017b), A Model-Based Epistemology of Measurement, in MÖSSNER, N. and NORDMANN, A. (eds.), *Reasoning in Measurement*, pp.233–253, Routledge.
- TAYLOR, B. N. (1971), Comments on Least-Squares Adjustments of the Constants, in LANGENBERG, D. N. and TAYLOR, B. N. (eds.), *Precision Measurement and Fundamental Constants*, pp.495–498, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- TAYLOR, B. N., PARKER, W. H. and LANGENBERG, D. N. (1969), Determination of e/h , Using Macroscopic Quantum Phase Coherence in Superconductors: Implications for Quantum Electrodynamics and the Fundamental Physical Constants, *J. Phys. Chem. Ref. Data*, 41(3), pp.375–496.
- THOMSEN, J. S. (1971), Some Aspects of Least-Squares Adjustments of Constants, in LANGENBERG, D. N. and TAYLOR, B. N. (eds.), *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.
- WILLINK, R. (2013), *Measurement Uncertainty and Probability*, Cambridge: Cambridge University Press.