Calibration: A Conceptual Framework Applied to Scientific Practices Which Investigate Natural Phenomena by Means of Standardized Instruments

**Léna Soler · Frédéric Wieber · Catherine Allamel-Raffin ·
Jean-Luc Gangloff · Catherine Dufour · Emiliano Trizio**

Abstract   This paper deals with calibration in scientific practices which investigate relatively well-understood natural phenomena by means of already standardized instrumental devices. Calibration is a crucial topic, since it conditions the reliability of instrumental procedures in science. Yet although important, calibration is a relatively neglected topic. We think more attention should be devoted to calibration. The paper attempts to take a step in this direction. The aims are two-fold: (1) to characterize calibration in a relatively simple kind of scientific practices; (2) to provide conceptual and taxonomic tools of broader scope that help to get a better understanding of calibration in more complex cases and other kinds of scientific practices. For this purpose, we first provide indications for why a conceptual framework is needed. Second, a bibliographic survey of works dealing with calibration is attempted. Third, we introduce different tools to enable a better understanding of calibration. Fourth, we turn to the elaboration of what we call a "simple exemplar" of calibration, illustrated through the case of the calibration of an equal-arm balance. Fifth, the tools previously introduced, and the framework of the simple exemplar, are applied to a more complex case of calibration: calibration procedures in X-ray experiments. This serves to show the work accomplished by the simple exemplar and to

Catherine Dufour—deceased.

L. Soler (✉) · F. Wieber
CNRS, Laboratoire d'Histoire des Sciences et de Philosophie, Archives Henri Poincaré, UMR 7117,
Université de Lorraine, 91 Avenue de la Libération, 54000 Nancy, France
e-mail: l_soler@club-internet.fr

C. Allamel-Raffin · J.-L. Gangloff
IRIST (Institut de Recherches Interdisciplinaires sur les Sciences et la Technologie),
Université de Strasbourg, 7 rue de l'Université, 67000 Strasbourg, France

C. Dufour
Institut Jean Lamour (IJL), UMR 7198 CNRS, Université de Lorraine, Parc de saurupt, CS 50840,
54011 Nancy Cedex, France

E. Trizio
Seattle University, 901 12th Avenue, Box 222000, Seattle, WA 98122-1090, USA

emphasize features of more complex cases of calibration. Eventually, we revisit and specify the nature, status, scope and value of the proposed framework.

1 Introduction

This paper[1] deals with calibration with a special focus on calibration in a *certain kind of scientific practices*, namely practices which investigate relatively well-understood natural phenomena by means of already standardized instrumental devices—for short, UNSI practices (U as Understood, N as Natural, SI as Standardized Instruments).

1.1 A Crucial But Neglected Topic

From an epistemological point of view, calibration is very important, since it essentially conditions the reliability of instrumental procedures which play a major role in science. The outputs of measurement apparatuses are no more than marks deprived of a determined physical significance, unless the apparatuses in question have been correctly calibrated. New devices or techniques must successfully pass calibration tests (or tests akin to calibrations) in order to be recognized as sound candidates for further development, including broad diffusion in the scientific community and in diverse non-scientific social groups.

Yet although highly essential, calibration is a relatively neglected topic (see Sect. 2). This is probably because in many configurations at least, calibration is treated as a preliminary unproblematic procedure that precedes the 'main show'—namely, the characterization of objects of interest *by means of* what has been previously calibrated in this purpose. Under closer examination, however, calibration is rarely a straightforward matter. The study of calibration practices shows that *even* in the at-first-sight *less problematic* configuration, that is, in ordinary, relatively routine scientific practices, calibration procedures are often delicate, complex, and not obvious activities. This *a fortiori* holds for calibrations involved in more creative, even "extra-ordinary" or "revolutionary" scientific practices, to borrow a famous expression of Kuhn.

1.2 Aims and Structure of the Paper

For the reasons given above, we think that more attention should be devoted to the topic of calibration. The present paper attempts to take a step in this direction.

The aims are two-fold: (i) to characterize calibration in a *particular*, *relatively simple* kind of scientific practices, namely UNSI practices; (ii) to provide conceptual and taxonomic tools *of broader scope* that also help to provide a better understanding of what can

---

[1] This paper is the product of a collective research project called "PratiScienS", launched by Léna Soler in 2007 in Nancy (France). PratiScienS stands for "Rethinking science from the standpoint of scientific practices". The aim of the project is to evaluate what we have learned about science from the "practice turn" in the science studies. One axis of the PratiScienS program is to characterize the procedures, reasoning, schemes of action, through which practitioners of the experimental sciences establish robust achievements, and calibration is part of these procedures. For further information about PratiScienS, http://poincare.univ-lorraine.fr/fr/operations/pratisciens/accueil-pratisciens.

be calibration in *more complex* cases and in *other types* of scientific practices (different from UNSI practices).

The article is structured as follows. First, we give indications about why we need such a framework (Sect. 1.3) and how the framework has been constructed (Sect. 1.4). Second, we give a survey of the works interested in science that have dealt with calibration and we locate our own work in the pictured landscape (part 2). Third, different tools, elaborated for the purpose of a better understanding of calibration, are introduced: (i) a four-question frame (Sect. 3.1.1), which invites to specify the target, presuppositions, aim and procedure of calibration; (ii) a rough mapping of scientific practices (Sect. 3.1.3), which enables us to *situate* the primary object of this article, calibration in UNSI practices (Sect. 3.1.2), with respect to calibration *in other kinds* of practices, and on this basis, to motivate our choice to start with calibration in UNSI practices (Sect. 3.1.4); (iii) a frame for the conceptualization of measuring devices (Sect. 3.3), needed to get a fine-grain grasp of the classic target of calibration—which is, classically, a measuring device. Our strategy is also explicated (Sect. 3.2), in particular the "strategy of the simple exemplar" (Sect. 3.2.2), which consists in the construction of an exemplar in the Kuhnian sense, intended to work as a benchmark and a compass for the analysis of instances of calibration. Fourth, we turn to the elaboration of the simple exemplar of calibration in UNSI practices, illustrated through the relatively simple and familiar case of the calibration *of an equal-arm balance* (part 4). Fifth, a more complex case of calibration in UNSI practices is examined (part 5): calibration procedures in X-ray experiments. The tools previously introduced, and the framework of the simple exemplar, are put to work for the analysis of this case. This leads to indicate the kind of work accomplished by the simple exemplar, and to emphasize some features of more complex cases of calibration that are not captured by the simple exemplar. Concluding, we revisit and specify nature, status, scope and value of the proposed framework.

1.3 A Conceptual Framework to Find a Way into the Jungle of Calibration Activities and Definitions

In order to motivate the project of building a conceptual framework for calibration,[2] it is required to say a word about the process that led our research team to conclude that such a framework was needed, and more generally about the broader research program of which the present work is part.

It is through Allan Franklin's work that we first met calibration as an object of interest in philosophy of science. Franklin provides a repertory of what he calls "experimental strategies", and calibration is one of these strategies. At the beginning, our aim was to consider Franklin's repertory of experimental strategies in a critical perspective, and to give a refined characterization of recurrent schemes of action through which scientists institute robust results (i. e., what they take to be reliable results). To advance progressively with respect to this aim, we decided to start with calibration. Is not calibration both a preliminary inevitable step, required in any experimental activity aiming at the establishment of robust results, and a not too complicated activity, at least compared with subsequent scientific investigations by means of previously calibrated devices? We thought this to be the case, so we took calibration as the starting point of our inquiry on experimental schemes of action. But we rapidly discovered that what was at stake was not so

---

[2] We are grateful to Sally Riordan (one of the reviewers of this paper, who agreed to be identified after the reviewing process), to have insisted on the need to motivate the project more explicitly, so as to highlight its value better.

simple (to say the least), and at the same time, much more interesting than we anticipated. As a result, calibration became an object of research per se.

For our inquiry on calibration, we relied on two kinds of materials: on past and present case studies available in the literature; but foremost on direct observations of present ongoing science as it is actually practiced. Such direct observations of 'science in the making' was possible, since two members of our team were in a position to provide such resources. A first set of examples of actual practices was provided by Cathy Dufour, who had been trained as a physicist and worked, for more than 20 years, in an experimental laboratory of condensed matter physics in Nancy, the Institut Jean Lamour (hereafter IJL), more specifically in the "Nanomagnétisme et électronique de spin" (nanomagnetism and spintronic) team.[3] A second set of examples came from ethnographic studies conducted by Catherine Allamel-Raffin in the field of astrophysics and pharmacology.[4]

Relying on these materials and interacting with physicists, astrophysicists and pharmacologists, we realized that the kinds of activities called "calibration" were not a set of obviously homogeneous activities. This heterogeneity was strikingly reflected in the instances of current scientific practices on which we relied as material of the present inquiry. The term "calibration" is used differently by pharmacologists, astrophysicists and solid state physicists. Moreover, some activities we were intuitively inclined to identify with calibration were not viewed as calibration by some practitioners.

To give just one example, the pharmacologists we studied did not consider the creation of a standard curve (in order to measure quantities of cytokine) as a procedure *of calibration*, but, rather, as the first part of their experiment.[5] Whereas according to the intuitions of the members of our research team, such a procedure was, without discussion, a case of calibration—since to perform a standard curve amounts to establish a reference scale (here with respect to quantities of cytokine), thanks to which the outputs of subsequent measurements can be converted into determined values of quantities of cytokine. Astrophysicists with whom we interacted operated with the same intuitions as ours. The procedure which, in astrophysics, can be seen as the equivalent of the standard curve in pharmacology starts with the choice of a set of celestial objects whose physical properties are already well-known. The objects in question are used to determine the relation between their known properties and the outputs of astrophysical instruments, so that when the same instruments are applied to *unknown* celestial objects, their outputs can be converted into reliable information about the phenomena under interest. Astrophysicists call "calibration sources" the well-known celestial objects which serve as points of reference in this procedure, and they identify the whole procedure *with a calibration*.[6]

---

[3] Very sadly, Cathy Dufour, one of the co-authors of the present paper, a crucial contributor to our research group, and also a very dear friend, prematurely passed away in March 2011 at the age of 46 years. This reference to her contribution is an occasion to pay homage to her.

[4] Pharmacological observations were conducted at the laboratory "Inflammation et environnement dans l'asthme" (EA 3771), Faculté de Pharmacie de Strasbourg, France. Astrophysical studies, at Center for Astrophysics (CfA), Harvard, Cambridge, USA.

[5] At the laboratory under scrutiny, scientists produce the standard curve by means of the ELISA test (Enzyme Linked ImmunoSorbent Assay). This left them with a correspondence between determined quantities of cytokine and instrumental outputs. They explicitly deny that this procedure is a calibration, and say: "When we produce a standard curve, it's in fact the first step of an experiment, it's included in the experiment, it's a part of an experiment" (interview made by C. Allamel-Raffin on 3rd March 2008, "Inflammation et environnement dans l'asthme" (EA 3771), Faculté de Pharmacie de Strasbourg, France, our translation from the French).

[6] Is this disparity of intuitions about calibration between pharmacologists and astrophysicists the manifestation of the fact that calibration is 'science dependent'? We leave the question open.

We also realized that in the science studies literature,[7] the few works focused on calibration rarely attempted to define it, and did not always rely on similar conceptions of calibration (see part 2). All in all, we rapidly felt lost at the heart of a dense and rich jungle—and we bet that anyone aiming at an understanding of what calibration is and what work calibration accomplishes in scientific practices would be subject to the same experience. In such a situation—which is perhaps typical of what holds for a relatively unexplored field—some order has to be introduced. This is precisely the function of the conceptual framework we are going to present in this article. Our ambition is to propose a framework able to work as a compass for researchers interested in the topic of calibration. We are confident that the framework will help these researchers to find *one* viable way— certainly not the only tractable path but at least *one among other possible ones*—into the 'jungle of calibration'. At the end of this article (Sect. 6), once the main pieces of the framework have been presented, and once the framework will have been put to work in relation to concrete practices of calibration performed at the IJL, we will articulate more precisely what we mean by a conceptual framework, what insights it has brought so far, and what insights it could bring applied to other cases not explored here.

1.4 Genesis of the Framework

A word about the genetic process through which our conceptual framework has been generated *in concreto* will help to get a first sense of its nature and its potential value.

We started with a set of examples of activities candidate to calibration, i.e., activities that could possibly be identified with calibration for a reason or another—because the scientists with whom we interacted considered them as calibration; or because scholars in the science studies treated them as calibration; or because our own intuitions, as analysts of science, led us to assimilate them to calibration. As mentioned Sect. 1.3, most of these examples were provided, on the one hand by the activities performed by C. Dufour and her colleagues at the IJL, and on the other hand by the astrophysical and pharmacological practices observed by C. Allamel-Raffin as an ethnographer of science. In order to find a path into the jungle of calibration, we attempted, starting from this diversified set of examples of practices candidate to calibration, to grasp their very nature, to identify significant differences between them, and to examine the possibility of some common core features between them. In other words, we attempted to construct a conceptual framework which would help to compare, classify and provide a better understanding of the activities under scrutiny.

Numerous successive versions of conceptual frameworks have been built in the course of our inquiry. The prism of each version shed a new light on our examples, but at the same time, new kinds of difficulties appeared, with which we had to cope through modified conceptualizations. So, the framework we are going to present is the result of a long and sinuous maturation process, constituted of a back and forth movement between our examples of calibration and multiple attempts to understand and classify them. Yet, the work invested in the clarification of the encountered difficulties and in the construction of more and more satisfying solutions is difficult to perceive in the end-product. Contemplating the final result apart from the process of its constitution, this final result might appear straightforward: the

---

[7] All along this paper, we use this expression "science studies" in a broad sense, intended to include all studies that take science in general, or one particular science, as the object of their inquiry, whatever approach, method and orientation is preferred. This includes philosophy of science, history of science, social studies of science, conceptual analysis such as the ones provided by the International Vocabulary of Metrology (see part 2), and any imaginable hybrid of these.

characterization of calibration provided in part 4 might appear obvious, and thus perhaps useless or uninteresting. We will come back to this point at the end of this paper, Sect. 6.3.

2 Calibration in the Science Studies: A Survey

Since the topic of calibration is in such preliminary state, we think it useful to offer a view of how calibration has so far been dealt with in the science studies (see note 7). This will enable us to introduce the main references which, according to us, provide interesting elements regarding the topic of calibration—even if calibration is not the primary object of interest of most of these works.

We map this new terrain by addressing three questions: the target of calibration; the definition of calibration; and the perspective in which calibration is called for. The answers attempt to take into account and relate many different domains and strands: history and philosophy of science, philosophy of experiment, social studies of science, philosophy of biology, philosophy of economics, metrology… as well as descriptive *versus* normative approaches.

2.1 The Target of Calibration

Let us start with the issue of the target of calibration—which is, as we will see Sect. 3.1.1, one element of the frame through which we propose to analyze calibration. The question is: when we scrutinize both the science studies literature and scientists' ways of talking, what kinds of item appear to be the object of calibration activities?

2.1.1 The Traditional Target: Material Instrumental Devices

Most of the time, and classically, the target of calibration is *a material instrumental device*, and often a *measuring* instrument.

Let us give a selection of examples of this type *in the historical, philosophical and sociological studies of science*. The target of calibration corresponds: to a spectrometer in Franklin (1997) (as well as to many other complex experimental devices); to thermometers in Carnap (1995 [1966]), Sibum (1995), and Chang (2004); to electron microscopes in Rasmussen (1993); to gravity-waves detectors in Collins (1985, 2004); to a chromatographer in Livengood (2009); to machines dedicated to the measurement of body composition in O'Connell (1993); to CO–CO_2 analyzers in Mallard (1998); to a measuring instrument in general in Buccianti et al. (2009).

In relation to the claim that the target of calibration is commonly identified with a material instrumental device, let us report a testimony issued from the studies of 'science in the making' that nourished our research on calibration. After having discovered disparities in the way the scientists used the term "calibration", we tried to clarify the underlying conceptions of calibration involved. As a result, it appeared that for contemporary pharmacologists (or at least the one we met), the fact that the target *is a material instrument* worked as a *necessary condition* to identify a given type of measurement *with a calibration*. As one pharmacologist told us: "We talk about calibration *only when* we calibrate *an instrument* such *as* a pipette, a pH meter, a plethysmograph, and so on".[8] It is because the creation of a standard curve does not have a material instrument as its target, that

---

[8] Interview made by C. Allamel-Raffin on 3rd March 2008, "Inflammation et environnement dans l'asthme" (EA 3771), Faculté de Pharmacie de Strasbourg, France (our translation from the French).

pharmacologists of this lab do not identify this procedure *with a calibration procedure*, as reported in Sect. 1.3.

2.1.2 Stretching the Target of Calibration

Calibration is, widely and classically, thought of as one way of controlling *material instruments*. However, besides the latter widespread traditional use of the calibration vocabulary, *other, extended uses* appear to hold. Below is a brief survey of *other kinds* of target found in the literature. The target of calibration can be extended to:

a. Immaterial conceptual devices such as *mathematical procedures*: see Boumans (2004), who assimilates mathematical procedures used in the social sciences to measuring "non-material instruments"; or see Franklin (1997), who is ready to extend the target of calibration to the procedures of analysis—including computer analysis, cuts in the data, etc.—which mediate the conversion of raw data issued from the experimental device into claims about natural phenomena.
b. *Models*, for instance *economic* models (Boumans 2007), or *biochemical* models (see Ramsey (2007), who deals with models in protein chemistry).
c. *Simulations* (Boumans 2006; Werker and Brenner 2004).
d. Material but not artifactual objects, such as *living beings*: populations of organisms in the laboratory (see Skipper (2004), who discusses the calibration of populations of organisms in artificial selection experiments); or human beings, for example scientists considered as instrument-readers, taking into account the possibility that different practitioners might read in different possible ways (Woodward 1989), or ordinary people considered as measuring systems in relation to quantities related to human perception (see e.g. Rossi et al. (2005)), who discuss calibration in the context of panel or jury testing), or to take another last example, scientists considered as more or less trustful sources of scientific information that have to be calibrated (see e.g. Andersen 2013).

Authors who apply calibration to extended targets of the latter kinds are often well-aware that their use involves a 'stretching' with respect to the classical target identified with a material instrumental device; their decision to stretch the target is often deliberate and explicit. For instance, Franklin talks about an "extended sense" of calibration or of calibration "more broadly construed" (Franklin 1997, 33) when he turns from the calibration of material apparatuses to the 'calibration' of data analysis. Still more explicitly, Ramsey (2007, 310) writes: "Usually, we think of calibration as something we do to instruments. (…) However, calibration can involve any scientific apparatus, not just scientific instruments". Under the general notion of "scientific apparatus", Ramsey includes, in addition to material instrumental devices (i.e., "scientific instruments"), *other kinds* of 'devices', such as equations (for example in quantum chemistry), models, and populations of organisms (Ramsey makes here a reference to Skipper (2004)).

2.2 The Definition of Calibration

Among the works which devote substantial developments to calibration—including those in which calibration is not the primary subject of the inquiry—very few of them provide an explicit definition of calibration. In what follows, we start from two available explicit

definitions, and extract their common core. Next, we reconstruct, from works that do not define but use the term of calibration, another definition which involves aspects not encompassed by the previous explicit definitions.

2.2.1 Franklin's Definition of Calibration

The most widely used definition of calibration is the explicit definition given by epistemologist Allan Franklin (Franklin 1986, 1990, 1997).

As already mentioned in Sect. 1.3, Franklin gives to calibration the status of an "experimental strategy", used by scientists to legitimate the reliability of experimental achievements. In such a perspective, Franklin defines calibration as "the use of a surrogate signal to standardize an instrument" (Franklin 1997, 31). The signal is a "surrogate", in the sense that it is used as a substitute for the unknown phenomena that experimenters aim to investigate. The "surrogate signal" is a signal of *already-known* properties. The logic of the strategy is the following: "If an apparatus reproduces known phenomena [i.e., the known characteristics of the surrogate signal], then we legitimately strengthen our belief that the apparatus is working properly and that the experimental results produced with that apparatus are reliable" (Franklin 1997, 31).

Franklin's definition in terms of the use of a surrogate signal works as a reference. It is explicitly mentioned and exploited in many works interested in calibration, including works in which the target of calibration is extended beyond material instrumental devices (see notably Boumans 2004, 2007; Harris 2003; Ramsey 2007; Rasmussen 1993; Skipper 2004; Woodward 1989).

2.2.2 The VIM's Definition of Calibration

A second explicit definition, not frequently mentioned in the literature of philosophy, history and social studies of science—with the notable exception of the papers of Boumans, and of Tal (2011)—is offered in the *International Vocabulary of Metrology* (VIM). This definition is interesting to consider, since the VIM is the result of complex negotiations between metrological experts issued from diverse fields and trained in relation to diverse backgrounds, and since the central aim of the VIM is to provide a uniform vocabulary of international scope about measurements of all kinds in any scientific discipline.

The last edition of the VIM (2008) defines calibration as an "operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication" (VIM 2008, 28). An "indication" is the "quantity value provided by a measuring instrument or a measuring system" (VIM 2008, 37).

The 2008 definition of the VIM involves two steps (say step 1 and step 2). In the step 1, a "measurement standard" is used. A measurement standard is an object of known properties. Step 1 consists in performing different measurements with the same measuring instrument, using different measurement standards characterized by different known values, and to record the "indications" (measured-values) delivered by the instrument in each case. The result is often presented as a curve which plots the measured-values obtained as a function of the values associated with the measurement standards. Since the values associated with the measurement standards are predetermined and known in advance, the measured-values provided by the instrument can be compared to them. Taking into account

the uncertainties, the first step therefore enables to estimate the difference between the taken-for-granted values of the measurement standards and the values actually obtained. Then, in step 2, this information can be used, in *subsequent* measurements performed on some *unknown* phenomena under investigation, in order to *convert the indications* delivered by the instrument *into correct values* of the quantities under interest.

2.2.3 A First Sense of Calibration: Calibration by Means of an Object of Known-Properties

Let us now compare Franklin's and the VIM's definitions of calibration. The "measurement standard" mentioned in the VIM's definition is equivalent to Franklin's "surrogate signal". True, the VIM probably primary targets measurement standards that have been produced and certified by professional metrologists (let us call them "*metrological* measurement standards"), whereas the idea of a "surrogate signal" can be understood in a broader sense, including the use of any phenomenon assumed to be sufficiently well-known to play the role of a taken-for-granted reference with respect to some purpose. As Boumans writes: "Franklin's (1997) discussion of calibration as one of the epistemological strategies for the establishment of the validity of experimental results is useful for our discussion of accuracy because *this strategy does not presuppose a standard*" (Boumans 2004, 230, our italics). In this quotation, "standard" is used in the sense of what we have called "*metrological* measurement standard", that is, an object inserted in a metrological traceability chain ordered in terms of degrees of precision. To the extent that Franklin's formulation does not necessarily imply a *metrological* measurement standard, it is more general, and this might be one reason why authors like Boumans and others, who attempt to stretch the target of calibration beyond material instruments of the natural sciences, frequently rely on Franklin's definition.

Beyond this difference, however, Franklin's "surrogate signal" and the VIM's "measurement standard" play the same role. In both definitions, the point is to make measurements on an 'object' of already-known properties (in a broad sense of the term 'object', not necessarily identified with a material solid). Furthermore, the same logic underlies Franklin's definition and step 1 of the VIM's definition, namely: since the 'correct results' of the measurements performed by means of the instrumental device under calibration on the 'object' of already-known properties is pre-determined, it is thus possible to evaluate the extent to which the result *actually* obtained differs from the taken-to-be-correct result. On this basis, the reliability of the instrument can be assessed.

True, Franklin's definition is not formulated into two steps. But it could have been, as it is obvious from Franklin's examples. In particular, Franklin writes that calibration can "provide a numerical scale for the measurements or for a numerical correction" (Franklin 1997, 33). Applying a numerical correction to the measured-values delivered by the apparatus under calibration in subsequent measurements would amount to perform an operation of the type described in the step 2 of the VIM's definition.

We conclude that the two definitions under discussion present a common core. This common core is: to compare the *known-in-advance value* of a property of an 'object' with the *measured-value* obtained for this 'object' by means of the measuring instrument under calibration; next, to use the result of this comparison to assess the reliability of the instrument and to correct the indications subsequently obtained when the same measuring instrument is applied to new, unknown objects. We call this conception of calibration the "calibration by means of an object of known-properties".

2.2.4 A Second Sense of Calibration: Calibration in Relation to Measurement Scales

A second sense of calibration—related to, but sufficiently different from the previous one—can be *reconstructed* from several works which are led to talk about calibration but *do not provide any explicit definition of calibration*. This sense of calibration has to do with *measurement scales*.

This second sense of calibration can be found in Carnap's famous 1966 book, *An Introduction to the Philosophy of Science*, chapter 6 (Carnap 1995 [1966]). The aim of this chapter is to analyze quantitative (or numerical) concepts. For that purpose, Carnap examines how "empirical criteria are established" (Carnap 1995 [1966], 53), which give meaning to quantitative concepts and enable to assign numerical values to them. He takes measurements of temperature with thermometers as an illustration. Calibration is mentioned in this context, in relation to scales *of temperature*.

Having explained how the magnitude "temperature" can be numerically specified through the construction of a given scale of temperature—for example the centigrade scale—, Carnap turns to the comparison of different scales. Considering the Fahrenheit, absolute (Kelvin) and centigrade scales, he writes: "The centigrade and Fahrenheit scales may be thought of as variants of the absolute scale, *differing only in calibration* and easily translated to the absolute scale" (Carnap 1995 [1966], 67, italics added). The three scales differ in calibration in the sense that "other states of substances are chosen for the zero and 100 points" (Carnap 1995 [1966], 66). Thus here, calibration concerns the choices that define one particular scale of temperature (choices of two fixed points, etc.). Calibration refers to some (partly conventional) decisions related to the concrete procedure through which some determined numbers are attributed to a given magnitude (here the temperature).

Leaving Carnap's particular text and taking into account other writings, we can say that *in relation to scales*, calibration—the act of calibration, or equivalently the verb "to calibrate"—denotes *two types* of actions.

i.   In some contexts, calibration names the *genetic process of construction of a given scale* and the *choices finally retained* (for the zero, etc.), by some scientists or groups of scientists, in a given episode of the history of science. In such case, to calibrate means to elaborate a new scale and to take the decisions that go with such task. We find this usage in some of Chang's and Sibum's historical works, which, as Carnap's, discuss thermometry. Chang shows how difficult it can be to determine the two fixed points that are needed to define a scale of temperature, as well as the precise form of the scale (see Chang (2004), notably chapters one and two). Sibum directs the attention on the uniformity of the inner diameter of the thermometer tube, which can be a problem in practice, and he analyses how Joule ingeniously dealt with this problem (Sibum 1995). Chang and Sibum explicitly use the vocabulary of calibration in their discussion. Although they do not define calibration in this context, we can reconstruct their position as involving sense (i).

ii. In some other contexts, calibration stands for *the use of an available (already established) scale*. In such case, to calibrate means to use this taken-for-granted scale as a reference for certain purposes. The just-cited works of Chang and Sibum involve this sense as well. For example, Chang mentions "the fact that Wedgwood did calibrate his thermometric scale with the Fahrenheit scale" (Chang 2004, 127). In the same vein, Sibum writes: "In order to compare his measurements with those of contemporary researchers he [Joule] very often brought foreign thermometers into his laboratory (…) to calibrate them against his own" (Sibum 1995, 101). Some instances

of what Rasmussen names "calibration against different instruments or methods" (Rasmussen 1993) can also be interpreted as occurrences of sense (ii).

2.3 Three (Not Mutually Exclusive) Perspectives on Calibration

Why are scholars of science studies interested in calibration? For what purposes, in what perspective, do they mention calibration? We distinguish three perspectives or trends: epistemological, socio-historical, and analytic. The three perspectives are conceptually distinct, but from a logical point of view, they are not necessarily mutually exclusive, and from an empirical point of view, they are sometimes dealt with in one and the same work.

2.3.1 An Epistemological Perspective Aiming at the Adjudication of Reliability Issues

A first perspective is *epistemological* and deals with the justification of reliability. The central issue is "the importance of calibration in establishing the reliability of scientific knowledge" (Rasmussen 1993, 229). In our categories, the issue is the reliability of the target of calibration—material apparatuses and their outputs, and by extension, as we have seen (Sect. 2.1.2), possibly many other things such as models, simulations, data analysis, and so on. The issue is to determine if, and to what extent, calibration is one way to *warrant*, *justify* or at least *support*, reliability claims about the target (and when the target is an instrument, reliability claims about the data obtained with this instrument).

This perspective seems to be the most often represented one. It must be added that reliability is always involved, even when reliability issues are not explicitly dealt with. In this perspective, the emphasis is on the *function* of calibration. The primary question is: what, exactly, is the power of calibration with respect to the justification of the reliability of its target? Which in turn leads to investigate in more or less details questions such as: what is calibration? What are the criteria to conclude that a given target is indeed well-calibrated? What kind of work do calibration procedures accomplish? In brief, the aim is to evaluate the *nature* and the *force* of the support calibration provides, in terms of reliability, to the target of calibration.

This epistemological perspective is the one within which Franklin has introduced calibration in the science studies. In this vein, calibration figures among the "*epistemological strategies*" (see for example Franklin 1997, 1, italics added) listed by Franklin in his "epistemology of experiment". Relying on Franklin's proposal, several other scholars have analyzed calibration in the same perspective. Boumans (2004), Ramsey (2007), Rasmussen (1993), Skipper (2004), Tal (2011), or Woodward (1989) can be considered as examples. These authors have discussed calibration as a potential way to support the reliability of diverse kinds of material or immaterial scientific instruments, in relation to different areas of the natural and the social sciences.

The same epistemological perspective focused on justificatory issues about reliability, also constitutes the background of the famous Franklin–Collins debate about the so called "experimenters' regress" (Collins 1985, 2004). In relation to calibration, the debate revolves around the question of knowing whether or not the "experimental strategy of calibration" (in Franklin's terms) is able to stop the experimenter's regress. Franklin argues in favor of a positive answer, at least in typical scientific configurations, whereas Collins advocates a negative answer. Franklin and Collins disagree about the power of calibration procedures, as well as about the kinds of factors (characterized as

"epistemological" *versus* "social") involved in scientists' judgments about whether or not a given instrument is indeed correctly calibrated and hence reliable.

To the extent that the aim is to settle justificatory issues, this perspective can be characterized as normative. Accordingly, the above-mentioned works can be characterized as *in part normative* (only *in part*, because the epistemological conclusions of these works often *rely on historical accounts* of scientific episodes that are assumed to be *descriptive in status and descriptively accurate*).

2.3.2 A Socio-Historical Perspective Aiming at the Description of the Standardization of Instruments at a More or Less Broad Scale

By contrast to the normative, epistemological perspective primarily directed toward justificatory issues, we can identify a second perspective, animated by a more 'purely' descriptive spirit, within which calibration is studied as a socio-historical object. The aim, in this socio-historical perspective, is to provide an descriptively accurate account of the (often complex) routes through which instrumental devices, techniques and objects of all sorts have, historically, acquired the status of well-calibrated and hence reliable items, either at a local level (at the scale of one individual or small groups of scientists), or at a global, possibly international level.

A relatively important corpus of socio-historical studies dedicated to the *local* conception and realization of new instrument devices, which, at a point or another of their developments, are led to mention calibration, fall within this second perspective. In this vein, we can mention, for instance, Bourget et al. (2002), Chang (2004), Chang and Yi (2005), Heering (2005), Licoppe (1996), Livengood (2009), Schickore (2007), or Sibum (1995). In many of the studies of this corpus, however, calibration is not the central topic: it is considered incidentally, or among other equally or most important issues (for example, the theoretical principles on which the explanation of the instrument rests). We could perhaps also include, as examples of the socio-historical perspective, the 'historical part' of the epistemological studies mentioned Sect. 2.3.1—providing we are ready to assume that the descriptive-historical stratum can be separated, and considered apart, from the epistemological-normative contribution.

Other works conducted in the socio-historical perspective involve calibration in relation to *standardization at a broader scale* and *professional metrology*. Calibration is a central task in professional metrology. Calibration in metrology has to do with the conception, realization, standardization and spreading at a large scale (ideally an international scale) of instruments, measurement standards, and units. Mallard (1998) or O'Connell (1993) can be cited as works conducted in a socio-historical perspective, which mention calibration in relation to professional metrology and large-scale standardization—O'Connell talks about "the creation of universality" (in the title of the article), Mallard about "how results and discoveries that emerge in a culturally laden context can be made universal" (Mallard 1998, 573).

2.3.3 An Analytical Perspective Aiming at a Systematic Conceptualization of Calibration

Finally, a third perspective on calibration can be introduced: the *analytical* perspective. Considered at the most ambitious level, the primary aim of the analytical perspective is the achievement of a systematic, fine-grained and comprehensive conceptualization of calibration.

As far as we know, only two writings explicitly and deliberately examine calibration in an analytical perspective: the VIM (2008), and one article written by authors involved in the elaboration of the VIM (Buccianti et al. 2009), which intends to explain and specify some aspects of the definition of calibration provided by the VIM.[9] Admittedly, a few number of the works conducted in the *epistemological* and *socio-historical* perspectives provide *some insights* about what calibration *is* (or may be). However, none of these works go very far in this direction, and in any case, none of the corresponding characterizations can be considered as systematic, fine-grained and sufficiently complete conceptualizations of calibration (which is not a reproach to these works, since to achieve such a characterization is not their aim). Carnap's perspective, in Carnap (1995 [1966], chapter 6), can be characterized as analytic, but as we have seen Sect. 2.2.4, the subject matter of Carnap's chapter is not calibration, and in this chapter, calibration is not the target of any specific analysis. Since the VIM is the only available attempt to provide a systematic conceptualization of calibration, and since this is also our aim in the present article, let us specify the distinctive features of the analytic perspective of the VIM. This will help, in the next section, to situate and motivate further our own project.

First, the analytic perspective of the VIM is a *normative* perspective, intended to standardize the vocabulary and the concepts at an international scale, as it is clearly stated at the beginning of the document. "This Vocabulary is meant to be a common reference for scientists and engineers—including physicists, chemists, medical scientists—as well as for both teachers and practitioners involved in planning or performing measurements, irrespective of the level of measurement uncertainty and irrespective of the field of application. It is also meant to be a reference for governmental and intergovernmental bodies, trade associations, accreditation bodies, regulators, and professional societies." (VIM 2008, 1). Calibration is defined in order to standardize what calibration has to be in different socio-techno-scientific contexts. In this horizon, second, the aim of the VIM is to list and define the "basic and general concepts" of metrology, and metrology is understood as the "science of measurement and its application" (VIM 2008, vii). The concepts of metrology that have to be listed and defined are "basic and general", in the sense that they are intended to be trans-disciplinary valid. Such project appears realizable because the VIM takes "for granted that there is no fundamental difference in the basic principles of measurement in physics, chemistry, laboratory medicine, biology, or engineering" (VIM 2008, vii). Accordingly, the definition of calibration proposed by the VIM is very general—and could be specified in regard to more specific classes of scientific configurations. Moreover, third, the definition is not very developed, and could be developed further—which is an obvious consequence of the fact that calibration is only *one* of the multiple concepts that the VIM attempts to define. Furthermore, fourth, the VIM is only concerned by measurements which use *already standardized* instrumental devices, measurement standards and units. Consequently, the VIM's definition could be completed to encompass calibration in more innovative practices.

2.3.4 Situation of Our Work on Calibration with Respect to Others

We have distinguished three perspectives, associated with different purposes, according to which the topic of calibration is discussed in the science studies literature. Let us now situate our own work with respect to these three perspectives. On this basis, we shall be in a

---

[9] Mari and Sartori (2007) also contains some insights about calibration in an analytic perspective.

position to motivate it further (in addition to what has already been developed in Sect. 1.3) and give a more precise idea of its object.

In the present article, our perspective is *analytic*—even if the ultimate aim of our research program about calibration is epistemological (see Sect. 1.3). Thus, our approach differs from epistemological discussions centered on justification of reliability and from socio-historical studies intended to describe how instruments have been stabilized in particular historical episodes. Our approach intends to provide a systematic conceptualization of calibration. Our analytic aim, expressed at the more general level, is to achieve a systematic, detailed and reasonably complete characterization of calibration. The existing conceptualizations are, according to us, far from being sufficient. We have indicated why in the previous section, focusing on the only available systematic characterization of calibration, namely the characterization of the VIM.

In this paper, we take a step in the direction of achieving our general aim. To tackle the problems one by one, we restrict ourselves, as a starting point, to a relatively simple and not too problematic case, calibration *in UNSI practices*—practices in which scientists use already well-mastered devices for the purpose of the investigation of relatively well-known types of phenomena. This circumscribed object of study—calibration in UNSI practices considered in an analytic perspective—is *different from what is most of the time studied in the literature under the heading of "calibration"*. Calibration in UNSI practices is different from calibration in professional metrological practices (understood as standardization at the international scale), and different from local practices of establishing new measurement scales or new instruments. Calibration in UNSI practices, as it will appear, encompasses instances of Franklin's and the VIM's definition, as well as instances of calibration against a given established scale. But beyond this partial intersection, first, our analytic perspective differs from Franklin's epistemological one, and second, the characterization of calibration we offer in this paper is more elaborated than the definition of the VIM.

In part 3, our perspective will be further clarified, through the presentation of some of the strategic options we have adopted and of the tools we have elaborated for the purpose of the investigation of calibration.

3 Tools and Strategy

3.1 A First Set of Tools

3.1.1 Four Questions to Investigate Calibration

In order to characterize calibration, it proves helpful to investigate four questions.

1. **The *target T* of calibration**: What kind of thing can be the *object O* of a calibration?
2. **The *presuppositions Ps* of calibration**: what is *taken for granted*, which delimitates what is *not* granted and has to be checked and controlled?
3. **The *aim* of calibration** applied to the object O under presuppositions Ps.
4. **The *procedure* of calibration**: the nature of its structural elements and the kind of logical stages through which the aim of calibration is achieved.

The question of the target of calibration has already been introduced above (Sect. 2.1), and applied as an analyzer to the literature related to calibration (see Sect. 2.2). As a result, it appeared that the target of calibration can be identified with heterogeneous kinds of objects, such as material apparatuses and immaterial mathematical procedures. This heterogeneity

suggests the possibility that *different targets* might go with *different types of calibration activities*. In any case, we cannot presuppose without examination that calibration activities are *not* target-dependent. A complete analysis of calibration should examine, for different types of calibration targets, if it makes a difference, and if so, should specify the kind of difference it makes. In this paper, however, we **restrict** ourselves **to** the most common case: the case in which the **target of calibration is** *a material instrumental device*.

The four questions are part of the framework we propose in order to achieve a fine-grained grasp of calibration. We will see below how the four questions apply to the case of calibration in UNSI practices, and what the answers are in this case (see part 4). However, the intended scope and fecundity of the four-question frame go much beyond the particular case of calibration in UNSI practices. We will come back to this point in the conclusion of this paper, Sect. 6.3.

3.1.2 What Are UNSI Practices?

When attempting to introduce some order into the network of activities that can be identified with calibration, it proves useful to distinguish *types of scientific practices* with respect to which calibration *takes different forms* and is *more or less problematic*. In this article—to repeat—we concentrate on calibration in *one particular type* of scientific practices, abbreviated as UNSI practices. It is thus needed to specify in more details what UNSI practices are. Since the specification of what UNSI practices are requires the delimitation of what they are *not*, *other types* of scientific practices will inevitably have to be introduced into the picture.

UNSI practices investigate well-understood domains of natural phenomena with already standardized instrumental devices. More precisely, UNSI practices are characterized by three structural distinctive features.

i. **The primary target of the inquiry is** *the natural world*. The aim is to acquire *knowledge about nature* (hence the "N" of UNSI, as Nature). This differentiates the practices in question from scientific practices *directed toward the means* of scientific inquiry—instrumental devices and techniques—, such as professional metrological practices or practices devoted to the local invention of a new type of instrument.

ii. The natural phenomena under inquiry are explored **by means of** *already standardized instrumental devices* (hence the "SI" of UNSI, as Standardized Instruments). In other words, the means of the inquiry are already *well-designed* and *widely entrenched* in some scientific communities. They are already theoretically well-understood and practically well-mastered. Practitioners of UNSI practices are *users* of socially well-mastered instrumental devices. This distinguishes UNSI practices from practices in which actors are *designers* of *new* instrumental devices.

iii. The kinds of **natural phenomena under inquiry** are relatively **well-understood** (hence the "U" of UNSI, as Understood). This distinguished UNSI practices from innovative practices dedicated to the investigation of newly explored and poorly understood phenomena. In UNSI practices, the natural phenomena under scrutiny do *not* pertain to a *completely new* domain. This excludes practices interested in phenomena hitherto experimentally undetected, the existence of which is discussed and contentious. In the practices we study, the phenomena involved have already been the object of an experimental investigation: their very existence is not questioned, and their main global features are already known. The aim of scientists is to go further in terms of precision and details.

3.1.3 A Rough Mapping of Scientific Practices, as a Basis for a Typology of Calibrations

The present article focuses on calibration in scientific practices characterized by the three features (i), (ii), and (iii)—called UNSI practices. *Other types* of scientific practices have been included in our attempt to delineate UNSI practices Sect. 3.1.2—as well as in Sect. 2.2 although more implicitly—which do not possess these three features altogether. A systematic typology of scientific practices would be needed, because calibration arguably takes different forms in each of these other practices and in UNSI practices (see next section). Here, we shall only provide a rough mapping.

Four types of scientific practices can be contrasted with UNSI practices. Two of these types have the *means of scientific inquiry* as their primary target. One corresponds to **professional metrological practices**, which aim at establishing a standardized instrumental stratum of international scope (and which can in turn be partitioned in several, more or less innovative sub-types, see Mallard 1998). The other corresponds to the **local invention of a new instrument** outside of professional metrology. The latter practices, like professional metrology, struggle to increase knowledge about the means of scientific investigation; but contrary to professional metrology, they are not essentially concerned by international standardization. Typically, their aim is to achieve a *particular, locally stabilized* instrumental prototype. The two remaining scientific practices take *natural phenomena* as their primary target. One corresponds to the **investigation of *new, poorly-understood phenomena* with *well-mastered* instruments**. The other corresponds to the investigation of ***new, poorly-understood phenomena* with *poorly-mastered* instruments**.

3.1.4 Situating Calibration in UNSI Practices with Respect to Calibration in Different Practices: A Comparatively Simpler and Less Problematic Case

For our present purpose, two points are important: first, calibration in UNSI practices *differs in significant respects* from calibration *in the four other types* of scientific practices; second, calibration in UNSI practices corresponds to *one of the simplest, less problematic configuration*. A full demonstration of these claims would require a complete characterization of calibration in the different types of practices involved, and a systematic comparison between them. Such 'demonstration-of-walking-by-walking' is obviously not an available option, given the limited scope of a paper. However, the two points just-mentioned seem rather obvious, *at least from certain angles of comparisons*. Let us give indications about the angle of comparison with respect to which they are the most obvious.

According to the second feature of UNSI practices (see Sect. 3.1.2, (ii)), calibration in such practices is calibration from the standpoint of scientific practitioners who are *users* of *already standardized*, *theoretically well-understood* and *practically well-mastered* instrumental devices. Focusing on this feature, it is clear that calibration in UNSI practices is *notably different* from calibration in practices which intend to design *new* instruments or to investigate the natural world with *poorly-mastered* instruments. In UNSI practices, the instrumental piece taken as the target of calibration is already socially stabilized: many things about this target are *already established*, *shared and taken for granted* (the 'things' in question will be specified Sect. 4, especially Sect. 4.2). Consequently, to calibrate in UNSI practices will have to do with controlling the *conformity* of a given instrument *with what is already collectively taken-for-granted* about this type of instrument. Whereas to calibrate in professional metrological practices which aim to establish a *new* instrumental 'universal' standard, or to calibrate in practices dedicated to the local stabilization of a *new* instrumental prototype, or to calibrate in practices devoted to the exploration of natural

phenomena with *poorly-mastered* instruments, whatever it might mean in details, will have to do with the *invention* and *collective acceptation* of an instrumental *novelty*.

Consequently, calibration in UNSI practices will be less creative, more predetermined, more pre-codified, and hence potentially less problematic and less controversial, than calibration in practices involving new instrumental standards or poorly mastered instrumental devices. Note, however, that this does not imply that calibration in UNSI practices is a completely straightforward matter that amounts to automatic routines and does not require ingenuity.

Finally, let us make one important last remark about the relation between calibration in UNSI practices and calibration in professional metrological practices. Although calibration in professional metrological practices will be left aside in the remainder of this article, whereas calibration in UNSI practices will be at the center, it is important to keep in mind the *essential dependence* of the first with respect to the second. Calibration activities in UNSI practices *use*, and most of the time *take for granted*, the standards established by professional metrologists. Thus the metrological calibrations *come first, both logically and chronologically*. Calibration in UNSI practices *presupposes* metrological calibration: the former takes the results of the latter (standards of all sorts) as already instituted and unproblematic. These remarks point to the *relative and recursive structure* of calibration activities. Calibration activities in UNSI practices exhibit a relative and recursive structure, which ultimately depends on the work of measurement experts (professional metrologists) who define and build the primary standards. Even if practitioners ignore the primary standards, and more generally ignore a large part of the metrological chain in which the instruments they use in their current scientific activities are inserted, *in concreto*, metrological calibration constitutes the background of calibration in UNSI practices.

3.2 Strategic Decisions

When we scrutinize scientific practices with the aim of understanding calibration, it appears—as illustrated above, see Sects. 1.3 and 2—that activities candidate to calibration are diversified and at first sight not obviously homogeneous. In such a situation, partly conventional decisions have inevitably to be taken, which are both conceptual and terminological: they simultaneously institute a certain cutting of the object under scrutiny (calibration) and posit some linguistic conventions (decision to name *this* calibration *and not that*). Such decisions can be motivated to a certain extent, but they are not universally compelling decisions. The intuitions can vary from one analyst to another, and even one and the same analyst might be subject to deep and evolving hesitations (we had multiple occasions to experience this in the course of our collective research on calibration). To cope with such a situation, we have adopted a number of strategic principles which have guided several inaugural constitutive decisions. One of these constitutive decisions has already been mentioned in the previous development, namely, the decision to distinguish types of calibration depending on types of scientific practices. In what follows, we present some other decisions, in relation to the strategic principles we have adopted.

3.2.1 From the Simple to the Complex

To find a path in the complicated and dense jungle of activities at first glance candidate to calibration, we followed a first usual strategic principle: go from simplest cases to more and more complex cases. In other words, we first attempted to identify and characterize relatively simple *isolable* configurations, and then, informed by the previous investigation

and armed with the tools elaborated in the process, we turned to more complicated and more problematic configurations.

This guiding principle led us, after having mapped the territory of calibration into different areas—calibrations in the five types of practices distinguished Sect. 3.1.3—, *to start with calibration in UNSI practices, as the first step of our inquiry on calibration*. Given the general features described in Sect. 3.1.2, calibration in UNSI practices predictably corresponded to one of the simplest and less problematic configuration we can find.

3.2.2 The Strategy of the Simple Exemplar

This, however, is not the end of the story, because *even focusing restrictively on UNSI practices*, multiple, not always homogeneous activities are still being identified with calibration by scientists (see Sect. 1.3 for an example) and/or by philosophers, sociologists and historians of science.

To cope with such a situation, we have followed a strategy that might be called the "**strategy of the simple exemplar**". The strategic maxim is the following: as the first step of the investigation, construct a simple exemplar of calibration, and then, in the next steps, use the simple exemplar as an analytic tool. The construction of the simple exemplar relies on the four-question frame (see Sect. 3.1.1). The final product—the simple exemplar as a result—provides determined answers to the questions of the target, the presuppositions, the aim and the procedure of calibration. Thereby, the simple exemplar offers a first grasp of what calibration in UNSI practices is or can be.

Let us clarify further the status and the role of the simple exemplar of calibration. The simple exemplar is an exemplar *in a Kuhnian sense*, which means that it has the value of a *striking, often encountered, prototypical configuration* (it is a "paradigm" in Kuhn's narrow sense of the term—the large sense corresponding to the disciplinary matrix). Moreover, the content of the exemplar which constitutes the starting point of the inquiry must be *simple*—that is, sufficiently simple to offer a telling scheme able to manifest *salient features* of what is at stake—*without any pretention to exhaust the topic*. So the simple exemplar that will be elaborated in part 4 does not pretend, *neither to be sufficient* to characterize the diversity of practices candidate to calibration in UNSI practices, *nor to provide a set of necessary and sufficient conditions* for identifying an activity of UNSI practices *with calibration*. As any Kuhnian exemplar, the simple exemplar is nothing more, but nothing less, than a *reference point* which works as a *compass* and an *analyzer*. The simple exemplar works as a compass and an analyzer, since it helps to discuss other, less simple and more problematic candidates to calibration, the latter being analyzed *in reference* and *by contrast* to the simple exemplar. Part 5 will give substance to the previous claims and will clarify their content further, through the consideration of some activities of calibration involved in X-ray experiments. In the conclusion of this paper, we will come back in a more systematic way on the kind of work that the simple exemplar is expected to perform (Sect. 6.2).

3.2.3 The Strategy, in Brief

To recap, our overall strategy is: to start with the most simple kind of scientific practice (UNSI practices); then to construct, for this kind of practice, a simple exemplar of calibration which offers a first schematic grasp of the kind of activity under scrutiny (in this respect, the simple exemplar is part of our characterization of the *content* of calibration);

subsequently, to exploit the simple exemplar as an analyzer for the discussion of more complex, more problematic and possibly less prototypical candidates to calibration (with respect to this function, the simple exemplar is part of our *analytical framework*).

3.3 Tools for the Analysis of Instrumental Devices Taken as the Target of Calibration

In our attempts to give precise answers to the four questions of the target, the presuppositions, the aim and the procedure of calibration in UNSI practices, it appeared that conceptual distinctions about the instrumental devices taken as the target of calibration were needed. So first of all, let us introduce the distinctions we finally retained (mentioned in bold in the text below). The developments will be illustrated by means of the example of an equal-arm balance such as the one of Fig. 1.

For short, we call a measuring instrumental device a "measurer". For example, a balance is a measurer, whereas an X-ray source is not a measurer. The function of a measurer is to evaluate quantities of a certain kind. In the case of the balance, the corresponding kind of quantity is "mass". The result of the evaluation of a quantity with a measurer usually takes the form of a numerical value associated with a measurement unit and a specified uncertainty.

3.3.1 Types and Tokens of Instruments

Two dimensions of any instrumental device need to be distinguished for the characterization of calibration, namely the **type** and the **token**.

The classical type-token distinction is actually applicable to any kind of reality, far beyond the particular case of instrumental devices. But applied to a *measurer*, this leads to distinguish what we call the **measurer-type** and the **measurer-token**. The measurer-type is a *conceptual* object, a *certain kind* of instrumental device. For example: a scale of the type "balance". The measurer-token is a *particular instantiation* of a given type. In our case, since we concentrate on *material* instrumental devices, the measurer-token is a singular *material* object: *this* particular balance here and there. For example, one of the precision balances constructed by Fortin for Lavoisier.

3.3.2 The Mesurandum and the Instrumental Outputs

When practitioners use a measurer-token of a certain type (for example of the type "balance"), they want to evaluate with a certain precision, with respect to certain aims, the values of certain kinds of quantities (for example mass) in some particular conditions.

What practitioners want to evaluate in a particular measurement sequence involving a particular measurer-token, we call it the "**mesurandum**". The mesurandum corresponds to what practitioners intend to measure. The specification of the mesurandum *minimally* requires the specification of the *kind of quantity* intended to be quantified in the measurements under interest (in our example, mass). But most of the time, the specification of the mesurandum *moreover* involves the specification of *multiple other elements*, such as the targeted degree of precision, or some environmental conditions. For instance, if the aim is to determine the length of metallic objects, the mesurandum will include the mention of the external temperature.

The mesurandum must be distinguished from what we call the "**instrumental outputs**". The instrumental outputs refer to the measurement results considered at the 'less

interpreted' level. In the case of our balance, the instrumental outputs would correspond to the different numbers inscribed on the different standard masses used when determining the mass of a given object, and to the position of the pointer in this configuration. Beyond the particular case of the balance, instrumental outputs might be pointer deviations, numbers inscribed on digital counters, graphs… In a word, the instrumental outputs correspond to measurement results as they would be described by the layman.

When performing a measurement with a measurer-token of a certain type, strictly speaking, we do not directly obtain the mesurandum. We obtain the instrumental outputs, which are then *convertible, more or less directly*, in the numerical value of some quantities of certain kinds (in the case of the balance, the conversion of the instrumental outputs in the mass of the weighted object is rather direct and simple).

3.3.3 The Measurer, as a Means to Convert Humanly Performed Operations into the Value of a Mesurandum Through a Scientific Scenario Based on a Fundamental Scientific Principle

As a result of the previous analysis, a measurer can be viewed as a means to convert certain humanly performed operations into a definite value of a mesurandum. In our example, these operations are, typically: to place a material object O on one pan of the balance; then to add standard masses on the other pan until the equilibrium is restored. The conversion of such operations into one value of the mass of object O involves some scientific theories (in our case: elements of mechanics). More exactly, the passage from these operations to a value of the mass is achieved through a certain **scientific scenario** based on some fundamental **scientific principles**.

In the example of our equal-arm balance, **the scientific scenario**, roughly characterized, goes as follows. When a massive object O is placed on one pan, it exerts on the end of one arm of the beam a vertical force F_1 whose magnitude is proportional to the mass of O. This produces a determined displacement of the beam. In order to restore the equilibrium position of the beam, standard masses are then placed on the other pan of the balance. They exert a vertical force F_2, proportional to the total mass of the standards used, on the end of this other arm of the beam. If the two arms of the beam are equal, the equilibrium horizontal position of the beam is restored when the magnitude of F_2 is equal to the magnitude of F_1. The mesurandum, that is, the mass of the object O, is then determinable from the instrumental outputs, namely, here, from the numbers inscribed on the different standard masses
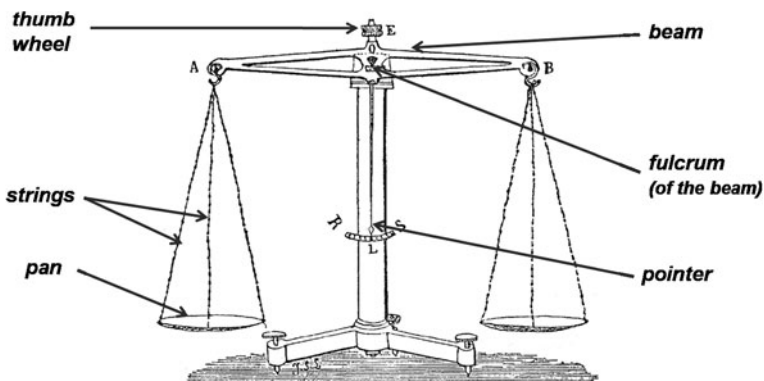
Fig. 1 A balance (a weighing scale with beam arms of equal length)

that have been put on the second pan in order to restore the equilibrium, joined to the remaining slight deviation of the pointer from the zero position on the graduated scale.

The **fundamental scientific principle** involved in the previous explanatory scientific scenario is often called the "lever principle". More technically, it is the principle according to which the static equilibrium of rotation is realized when the sum of the moments is zero.

3.3.4 A Scale of Types: From the Generic Type to the Model of a Measurer

A fundamental scientific principle such as the lever principle is—beyond the particular example of the balance—actually what defines, at the most fundamental level, a ***generic type of measurer***. It is what *individuates* a measurer *as one* determined *generic type* different from *another* generic type. For example what individuates the generic type "balance" and differentiates it from another type of weighing instrument such as, for instance, the generic type "electronic spring scale" (which characterizes, say, a digital kitchen scale).

Given the fundamental principle that defines a generic type of measurer at the most general level, a multiplicity of different sub-types can be conceived and realized (this is typically the task of engineers). For example, the lever principle is involved, and differently exploited, in a multitude of different sub-types of measurers that can all be used to assess the mass of objects (each having some specific advantages and disadvantages). A first sub-type corresponds to our example, that is, an equal-arm balance. A second one is the steelyard balance (or Roman balance). A third one is the Roberval balance. And so on… We can conceptualize families of instruments through an indefinite multiplicity of hierarchical levels of sub-types, sub-sub-types, etc. From the highest level of the generic type, to lower levels, the mesuranda, and the scenario through which one describes how the measurer converts human operations into determined values of mesuranda, are defined through *more and more specified* descriptions which use *more and more specific* technical categories.

For our present purpose, we shall focus on the 'lowest-level' of such a scale of types. We call the description of a measurer at this lowest-level the ***model* of the measurer-type** (or, for short, the **measurer-*model***).

3.3.5 The Measurer-Model: Optimal Working and Predicted Deviations

We use the term "model" as it is used when we speak, for example, of the "Ford Model T" (generally regarded as the first affordable model of automobile). On the one hand, the measurer-model is a particular sub-type of the generic type, and hence a *conceptual* object. But on the other hand, the description of a measurer-model intends to be the description of a *real material object*. It is a detailed description of a real measurer-token and its performances, *not* the characterization of an *idealized* measurer. It is situated at a level which is, so-to-speak, the most 'contiguous' to the concrete, material measurer.

As an illustration, in the case of an equal-arm balance, the model specifies, in addition to one version or another of the scenario sketched Sect. 3.3.3:[10] the materials, the form and

---

[10] The scenario provided Sect. 3.3.3 has been introduced at a point of the developments where the notion of model was still not available, but it should be clear from now that it is situated at the level *of the model*. It corresponds to one possible description *of one particular model*—the model "equal-arm balance"—of the type "balance". The scenario is obviously too specific to be associated with the level of the generic type. What holds for the generic type (definition of the mesuranda, description of the scientific scenario) must also hold for *all* the multiple particular models, and is then specified differently for each particular model. But the scenario proposed above does not apply to *any model* of the generic type "balance". For instance, it is not valid as a description of a balance of the model "Roman balance" (which involves only one pan).

the dimensions of the beam, of the strings, of the pans, etc.; the attainable performances, in particular the sensitivity, the maximum capacity of the balance, etc.; the conditions of use, taking into account real-life environmental variations (for example, air currents in the room can affect the balance's operation).

With respect to an analysis of calibration, it is useful to distinguish two aspects within the characterization of the measurer-model: (i) the optimal working; and (ii) the predictable deviations with respect to the optimal working.

i.   The characterization of the **optimal working**, including an associated uncertainty of measurement, tells the users what is at best actually obtainable from a measurer-token generated according to the model. "At best" means: just after the manufacturing of the measurer-token, assuming no defaults of fabrication, assuming normal conditions of utilization, etc. In brief, assuming the conformity of the real token to the specifications of the conceptual model.
ii. The **predictable deviations with respect to the optimal working** are, for example in the case of our balance, a possible drift in the standard masses daily-used, or a possible blunting of the central knife-edge (such a blunting would increase the frictions on the balancing of the beam and would then lead to a decrease of sensibility).

On the whole, the specifications of a measurer-model delimitate the identity of the model, and at the same time define what should be the identity of each individual measurer-token generated according to the model. So the measurer-model provides the normative identity of an (indefinitely large) class of individual measurer-token supposed to be instantiations of this model. It gives a normative characterization of any individual instantiation of the class defined by the model.

3.3.6 A Frame for the Conceptualization of Instrumental Devices

Table 1 offers a synthetic overview of the conceptual distinctions introduced Sect. 3.3, applied to the particular case of the equal-arm balance.

Table 1 The instrument-frame (bold items), applied to the particular case of the equal-arm balance

Generic type	Generic name	Balance
	Measuranda for the measurement of which the generic type is designed	Mass of certain kinds of material objects in a given context with a certain degree of precision
	Fundamental scientific principle	Lever principle (static equilibrium of rotation realized for null sum of the moments)
	Kinds of quantities typically involved in the mesuranda	Mass
Model	**Name and description**	Equal-arm balance (see Sect. 3.3.3 for a description)
	Scientific scenario	See section 3.3.3
	Instrumental outputs	Numbers inscribed on the standard masses of the balance used in a given measurement joined to the residual deviation of the pointer from graduation zero
	Predicted deviations with respect to the optimal working	Drift of the standard masses used Blunting of the central knife-edge

Table 1 is intended to achieve two aims simultaneously. On the one hand, it provides a *substantial* characterization of a *particular* measurer, namely of the model of balance that will be used in part 4 as the central example for the sake of the illustration of the simple exemplar of calibration. On the other hand, it provides a *tool of general scope*, namely a general frame for the conceptualization of *any measurer-token*. The frame corresponds to the bold items of the Table. This frame can be applied to any newly-encountered case of measurer (see Sect. 5, Table 3). It is conceived as a basic starting point, intended to be, when needed, either completed, or accommodated (see Sect. 5, Table 4 for an adaptation to a non-measuring device). To the extent that this frame serves as the basis of an adaptation to instrumental devices that are *not* measurers, its fruitfulness is not restricted to the case of measurers, and we can thus see it as a frame for the conceptualization *of instrumental devices* in general. So we call it **the instrument-frame**.

We are now in a position to construct the simple exemplar of calibration by providing determined answers to the questions of our four-question frame and by using the instrument-frame.

4 Constructing the Simple Exemplar by Means of the Four-Question Frame and the Instrument-Frame

4.1 Investigation of the First Question: The Target T of Calibration

The first question is: what kind of thing can be the object of a calibration?

For the simple exemplar as we define it, the target is a *measurer*. To go further, we need the distinction (introduced above Sect. 3.3.1) between the measurer-type and the measurer-token. According to this distinction, our answer to the first question is: for UNSI practices, and as we define the simple exemplar, **the target T of calibration is a measurer-*token***.

But not all procedures which target a measurer-token are calibrations. So we need to further specify what distinguishes a calibration test from other possible tests of a measurer-token that would not count as calibrations. One important difference lies in some presuppositions about the measurer. This leads us to our second question.

4.2 Investigation of the Second Question: The Presuppositions Ps of Calibration

The question is: what presuppositions Ps are constitutive of the kind of activity calibration is?

Calibration in UNSI practices involves some presuppositions *about the target of calibration*, that is, about the measurer. At this level, the question is: what is taken for granted *about the measurer*, which delimitates what is *not* granted and has to be *checked and controlled* in a calibration procedure? Our answer is: two different presuppositions about the measurer are involved in calibrations in UNSI practices.

The first presupposition is: the measurer as a conceived object is *not problematic*. This means that neither the generic-type nor the model of the measurer-token are being questioned. In other words, all the elements involved in the characterization of the model, including the elements related to the generic-type, are taken to be solid and are viewed as uncontroversial pieces of human knowledge. This is clearly the case for an equal-arm

balance. From a theoretical point of view, such a model of balance is a fully understood object. Nobody questions the validity of the deeply entrenched lever principle which defines the generic-type. Nobody contests the ability of balances of this model to enable reliable measurements of mass. So the **first presupposition** (**P1**) is: the **measurer-model is unproblematic** (for short, say "**unproblematic model**").

Obviously, this first presupposition derives from our initial choice to concentrate on calibration *in UNSI practices*, that is, in practices which use *already standardized* instruments. Or more exactly, to make P1 explicit and to explain (with the help of the concepts of the instrument-frame) what P1 exactly means is one way to explicate important aspects of what is meant when we say that UNSI practices use *standardized* instruments: standardized measurers are unproblematic measurers in the sense explicated in the previous paragraph.

The **second presupposition** (**P2**) is: **the measurer-token is not defective** (for short, say "**not defective token**"). In other words, there is no breakdown, no failure. Globally, the measurer works properly—even if it is not necessarily precisely adjusted. Re-described in our categories, the distance between the measurer-token and the measurer-model is not *too* important. Obviously, this assumption can be questioned in the course of a sequence of actions first conceived as a calibration. But as soon as this assumption is abandoned, the nature of the activity changes, and the activity then involved is, according to our conceptual and terminological decisions, no more adequately categorized *as a calibration*. If practitioners conclude that the measurer-token is defective, typically, the operations that follow correspond to a *repair*, not to a calibration.

4.3 Investigation of the Third Question: The Aim of Calibration

Our next and third question is: what is the aim of a calibration applied to a measurer-token T under presuppositions Ps? Our answer will distinguish the proximate aim and some more distant aims.[11]

4.3.1 The Proximate Aim of Calibration

Expressed at the more general level, the proximate aim of calibration is to master the possible gap between the measurer-token and the measurer-model. More precisely, the aim is to master the distance between:

i.   the instrumental outputs *actually obtained* with *this* individual measurer-token *at a given time in a given context* (in our example: the different numbers inscribed on the different standard masses used to restore the balance equilibrium and the residual deviation of the pointer, which are convertible in the value of a certain mesurandum, here the mass of an object);

ii. the value of the mesurandum that *should have been obtained* in the optimal configuration (that is: *if* this measurer-token as used in this context *actually coincided* with the measurer-model in optimal working).

For short, we will call the difference between these two items **the obtained/optimal discrepancy**. So in UNSI practices **the aim of calibration is to master the obtained/ optimal discrepancy** for a measurer-token at a given moment in a given context, assuming

---

[11] This terminology is borrowed from Hasok Chang's framework for activity-based analysis (Chang 2013). The proximate aim is also called by Chang the "inherent aim".

P1 and P2. This is what we call **the proximate aim** of the activity of calibration, that is, the most immediate aim.

4.3.2 The Distant Aims of Calibration

Beyond the proximate aim, calibration also serves other, **more distant aims**. As a crucial illustration, we can stress that a more distant aim of any calibration is to create the conditions for, and to actually obtain, measurement values that are commensurable with the international system of units. Obtaining this commensurability is a condition for another distant aim, namely, to ensure a universal (international) confidence in measurements performed all over the world in different places and conditions.

Although in the perspective of a more complete and meaningful characterization of calibration, it would be desirable to introduce such distant aims of calibration, calibration in UNSI practices is nevertheless independent of the specification of *most* of its multiple possible distant aims. However, beside the proximate aim, there is one distant aim that cannot be left aside for the understanding of calibration in UNSI practices. It cannot be left aside because it influences the way the proximate aim of calibration is specified and translated into a concrete procedure in practice. This distant aim is to obtain reliable results in subsequent, more or less pre-determined targeted measurements. We call the latter measurements the **end-measurements**.

4.3.3 Reliable End-Measurements as the Essential Distant Aim of Calibration in UNSI Practices

The expression "end-measurements" intends to play on the two-fold meaning of the term "end". First, it intends to suggest the idea of *finality* as involved in a means-end relation: here, the end is to perform reliable subsequent measurements, and calibration is one means with respect to this aim. Second, it intends to suggest the idea of *the end of a story*: in our case, the end of an experimental sequence in which calibration plays the role of the introductive chapter.

A procedure of calibration in UNSI practices is never accomplished just for itself but always with the intention to perform some end-measurements. So a calibration procedure in UNSI practices cannot be adequately thought in isolation, independently of its relation to the end-measurements, because the calibration of *one and the same measurer-token* may vary, depending on the purpose of the end-measurements.

4.3.4 Specification of the Proximate Aim Taking into Account the End-Measurements

We can inject the constraints associated to the end-measurements into the definition of the proximate aim. The proximate aim of calibration would then be: to master the obtained/optimal discrepancy *that matters given the mesurandum of the end-measurements*.

Suppose, for example, that the end-measurements intend to measure, with a precision of ± 1 g, the mass of an object O, and that this mass is expected to fall within an interval of 20–40 g. The calibration procedure of the same balance-token will not be the same as if the end-measurements intended to measure, with a precision of ± 1 g, another object for which the mass is expected to fall within an interval of 200–400 g. As the obtained/optimal discrepancy is not necessarily the same across the range of the balance, the calibration procedure must attempt to determine this discrepancy *for an interval of values which is*

relevant regarding the specific range of values that will be involved in the end-measurements.

More generally, the calibration procedure must master the obtained/optimal discrepancy for a range of values, a degree of precision, etc., which are relevant with respect to the ones involved in the end-measurements. Thus, the **most precise formulation of the proximate aim of calibration in UNSI practices** is: **to master the obtained/optimal discrepancy for a mesurandum which fits with (is as close as possible to) the mesurandum of the end-measurements**.

4.4 Investigation of the Fourth Question: The Procedure of Calibration

Our fourth and last question is: *what kind of procedure achieves the aim of calibration?* What are the structural elements and logical steps of a calibration procedure?

In Sect. 4.3, we wrote that the aim was *to master* the obtained/optimal discrepancy. Now we have to specify what the verb "to master" means here. It means: first, to *evaluate the discrepancy* (which implies most of the time a *quantitative* evaluation of the the gap); second, if needed—depending on the result of the previous evaluation and on the characteristics of the end-measurements—to *correct the discrepancy* or to *take it into account in a way or another*. Thus, a calibration procedure can be divided into two logical moments: first a moment devoted to **calibration *tests***; second, a moment consisting in the application of **calibration *operations*** according to the conclusions of the testing stage.

4.4.1 Calibration Tests

Concerning calibration tests, two species are commonly involved in UNSI practices: blank calibration tests and calibration tests with a measurement standard.

a. *Blank* calibration tests

Blank (or background) calibration tests give the blank indication of the device under test. For example, the blank calibration of our balance would consist in assessing the position of the balance pointer when the pans are empty.

The calibration test is 'blank', in the sense that a measure is undertaken in the absence of any object *of the same kind as the object the end-measurements aim to characterize* (note here one more time the indispensable reference to the end-measurements).

The principle of a blank calibration test is to perform measurements in all points similar to the end-measurements (same measurer-token, same experimental context, same kinds of measured quantities, etc.) *except one*: no measured object of the kind involved in the end-measurements is involved in the calibration tests (in our example: no material object is placed on the pans of the balance). Ideally, the blank calibration test measurement and the end-measurements are in the following relation: *without/with* a measured object of the kind under study *all other things being equal*. If this is actually the case, the blank test gives the background noise, specifically due to the measurer-token, that will be present in the end-measurements.

This noisy contribution of the measurer-token is determined by a comparison between the instrumental outputs obtained as the result of the blank calibration (for example: a position of the pointer corresponding to 1 g), and the prescriptions of the model in optimal working (i.e., a position of the pointer corresponding to the equilibrium zero position, taking into account the precision of the reading). The difference, that is, the signal generated by the balance-token in the absence of any material object placed on

the pans, is something that practitioners must be careful not to count as a contribution of the object under study.

b. Calibration tests with etalons

A measurement standard (in French: "etalon") is an object of already well-known properties. It can be a *metrologically certified* measurement standard or an object with sufficiently well-known and stable properties, which is used by practitioners as a working measurement standard. Since the English word "standard" has a very broad sense which might create ambiguities in the context of our discussion, we prefer to use the French word "etalon". So the second species of calibration test corresponds to **calibration tests with etalons**. Franklin's and the VIM's definitions of calibration, discussed Sect. 2.2, correspond to calibration tests with etalons.

Calibration tests with etalons consist in measuring against the etalon, by means of the measurer-token under test, quantities of already-known values. In the example of the balance, the etalons correspond to certified etalon masses, or at least, to etalon masses the values of which have been previously determined more precisely than the values of the standard masses daily-used as part of the balance-token under test. A calibration test with etalon *of the balance* (that is, a calibration test of the balancing mechanism *plus the daily-used standard masses*), would be: to place an etalon on one pan of the balance; then to place daily-used standard masses on the other pan in order to restore the equilibrium as much as possible; and, finally, to evaluate the difference between the mass actually measured in the test and the already-known mass of the etalon. Since the mass of the etalon is known in advance, the instrumental outputs that *should* be obtained with a measurer-token of this model *if* the token *coincided with the optimal working* are pre-determined. In our example, the position of the pointer and the different numbers inscribed on the different standard masses used to restore the balance equilibrium, which are convertible in a measured value of the mass of the etalon, should coincide with the certified value of the mass of the etalon. If the calibration test leads to a value *significantly different*, the difference corresponds to the obtained/optimal discrepancy (say: +2 g). This contribution of +2 g must be attributed to the measurer-token as used in this context, and *not* to the object under study in the end-measurements.

Ideally, a calibration test with etalon is in all points identical to the end-measurements, *except in one respect*: the measured-object used in the calibration test is a token with already well-characterized properties, known before and independently of the calibration test. Except in this respect, the calibration test is similar to the end-measurements, in that the calibration test (ideally) involves: the same measurer-token; the same experimental context; a mesurandum as close as possible to the mesurandum of the end-measurements, and in particular, a mesurandum which involves the *same kind* of measured-object.

This latter condition, "*same kind* of measured-object", is very important. To understand why, we have to remember (see Sect. 4.3.4) that the aim of calibration is to master the obtained/optimal discrepancy *for a mesurandum which is as close as possible to the mesurandum of the end-measurements*. Applied to a calibration test *with etalon*, this condition imposes on the etalon to be *as similar as possible* to the objects under study in the end-experiments. If the etalon is too different from the object under study in the end-measurements, the information obtained in the calibration test is not applicable to the end-measurements. This point has been stressed by Harry Collins, in the context of his debate with Franklin about the power of calibration with respect to the experimenter's regress. Collins talks about "the assumption of near identity of effect between the surrogate signal and the (…) signal that is to be measured (detected) with the instrument" [in what we have called the end-measurements] (Collins 1992 [1985], 105). Franklin

expresses the point as "the adequacy of the calibration, that is the 'near-enough' identity of the surrogate signal with the desired signal" (Franklin 1997, 75).

Concluding, we can say that a calibration test with etalon in UNSI practices involves, in addition to the two presuppositions P1 and P2 about the measurer-token, **two supplementary presuppositions about the etalon**:

P3: the values certified or assumed about the etalon are indeed reliable (for short, say: **reliability of the etalon**).

P4: the etalon is sufficiently similar to the object under study in the end-measurements (for short, say: **adequacy of the etalon**).

c. A brief comparison between the two species of calibration tests
    The underlying logic of a calibration test with an etalon, as well as of a blank calibration test, is a *differential logic* in which the end-measurements play the role of a reference point. In both cases, a circumscribed contrast is created between the end-measurements and the calibration tests. In both cases, this contrast allows to delimitate differentially what must be attributed, on the one hand to a bias introduced by the measurer-token as used in this context, and on the other hand to the measured-objects under interest.
    The difference between the two cases lies in the fact that in calibration tests *with etalons, the measurer-token is in interaction with a measured-object of the kind under interest* (the measurer-token is employed in conformity to its intended function and normal use, namely, to indicate the mass of weighting objects placed on its pans). In one case or the other (with/without weighting objects on the pans), the possible obtained/optimal discrepancies are not necessarily due to the same causes. Hence the two species of calibration tests provide possibly different and complementary information about the measurer-token and its possible drift with respect to the optimal working as defined by the measurer-model.
    Once calibration tests have been performed, practitioners have to decide, according to the obtained/optimal discrepancy they have found and to some *desiderata* imposed on the end-measurements, if calibrating operations are required or not, and if so, which ones.

4.4.2 Calibration Operations

Two kinds of calibrating actions are commonly involved in prototypical calibrations: *material* operations and *symbolic* operations.

Material operations correspond to concrete manipulations exerted on the individual instrument, which introduce effective and tangible modifications of the measurer-token as a material body. For example, with the balance, following the blank calibration test described in point a) of Sect. 4.4.1: to manipulate a thumb wheel in order to restore the equilibrium of the beam by displacing it slightly with respect to its fulcrum, so that the pointer, which initially coincided with 1 g, finally coincides with zero.

Symbolic operations are diversified, but as a prototypical illustration, they correspond to mathematical corrections applied to the instrumental outputs actually obtained in the end-measurements. In the case of our balance, following the calibration test with etalon described in point (b) of Sect. 4.4.1, it would for example amount to subtracting 2 g to the values actually obtained in the end-measurements.

Material operations transform the measurer-token so as to make the real material instrument as close as possible to the model in optimal working. Symbolic operations do

not intervene in the measurer-token as a concrete physical object: the gap between the real individual instrument and the model in optimal working remains unchanged; the discrepancy is treated intellectually, by an intellectual operation applied to the instrumental outputs actually obtained in the end-measurements.

4.5 A Synthetic Overview of the Simple Exemplar of Calibration in UNSI Practices

Table 2 recaps the main elements developed in part 4.

Table 2 provides a conceptual framework for the analysis of instances of calibration in UNSI practices. The framework will be used below, in part 5, for the analysis of X-ray experiments. This use will show the kind of work accomplished by the framework.

Table 2 The simple exemplar of calibration in UNSI practices

Calibration in UNSI practices: a simple exemplar			
End-measurements	Must be specified case by case because they impact on the details of the aim and procedure		
Target	A measurer-token		
Presuppositions	**About the target** (P1) Unproblematic model (P2) Non-defective token	**About the etalon** (P3) Reliability (P4) Adequacy	
Aim	Master the obtained/optimal discrepancy (for a mesurandum which is as close as possible to the mesurandum of the end-mesurements)		
Procedure	**Tests**	Blank	
		With etalon	
	Operations	Material	
		Symbolic	

4.6 Qualifications, Complements, Indications for Further Research

To close our presentation of the simple exemplar, we would like to stress that the characterization given above could be refined in diverse respects.

4.6.1 Environmental Factors

Our characterization of the simple exemplar did not sufficiently consider the issue of environmental factors. True, our characterization did not completely ignore such factors. In particular, the aim of calibration has been defined as the evaluation—and correction if needed—of the distance from its optimal working of "a measurer-token *at a given moment in a given context*". According to this definition, the mention of "a given context" encompasses possible environmental influences. Moreover, our characterization of the model of an instrumental device also mentioned elements that belong to environmental variables (see Sect. 3.3.5: "the model specifies (…) the conditions of use, taking into account *real-life environmental variations*"). But beyond such simple indications about the importance of environmental factors, we did not go into the details about how to take the environment into account. Such details depend to a large extent on the particular model of measurer under scrutiny, but they should be incorporated in a refined picture of calibration in UNSI practices.

One more word on this refined picture. For standardized models of instruments, the specifications of the model provide information about the environmental factors amenable to influence the instrumental outputs of measurements undertaken with this instrument. When it is known that some environmental factors can affect a measurer, these environmental factors have to be controlled. In the case of the equal-arm balance, environmental conditions such as draughts, vibration, temperature changes or changes in air density can affect the balance's operation. These factors have to be assessed, particularly if a change of location of the balance has occurred. For some of these factors, especially changes in temperature and air density, corrections can be introduced. Take, for example, air density changes. Air density varies with pressure, temperature and relative humidity. A variation of air density leads to an air buoyancy variation, which can affect the values of the masses measured with the balance-token. Hence an air buoyancy correction may be needed, by multiplying the measured-values by a correction factor. This correction factor is calculated by taking into account the density of air during the weighing (which can be estimated by a formula linking air density to air pressure, relative humidity of the air and air temperature), the density of the standard (certified) masses, and the density of the material being weighed.

4.6.2 Repeated Measurements and Statistical Treatment for the Purpose of Evaluating Measurement Precision

In our characterization of the simple exemplar, we have assumed the evaluation of the value of a given mesurandum in the calibration tests (for example the value of the mass of a given etalon provided by a given balance-token in a calibration test with etalon) only required to perform *one single* measurement with the measurer-token. This is, of course, a simplification.

We have ignored the practice of *repeating* the same measurements for the evaluation of *one and the same* mesurandum (for example, the repetition of several same weighting measurements of the same etalon with the same balance-token), and the statistical treatment of the different values obtained (calculation of the mean, of the standard deviation, etc.).

However, even if it had been completed so as to take into account this point, the simple exemplar of calibration in UNSI practices as characterized above would still not correspond to

what might be called a 'full' or 'systematic' calibration procedure. Without giving too many details, we shall nevertheless provide some indications about what 'full' calibration procedures are, and what their relation to the simple exemplar of calibration as described above is.

4.6.3 The Simple Exemplar of Calibration in UNSI Practices in Relation to 'Full' Calibration

A full calibration procedure of one and the same measurer-token includes multiple tests directed toward *different parameters*, such as the measurements *repeatability*, the measurements *error* across the range of the measurer, the *sensitivity*, etc. Using the results of these tests, a global *uncertainty of measurement* is calculated, which is assumed to hold for subsequent measurements, providing a *ceteris paribus* condition can be assumed to hold.

Whereas calibration as described by the simple exemplar is performed *almost each time a determined sequence of end-measurements is planned*, a full calibration procedure is performed *less frequently*. A full calibration procedure is performed when the balance enters service in the laboratory, when a significant change in the laboratory's environmental conditions occurs, when a change in location or position of the balance occurs, when daily calibration tests show a significant change with respect to what is expected according to the model and to the results of the last full calibration procedure, and in any case—even if no contingent event of the previous kind is suspected to have occurred—regularly (but not frequently: for example each year). In contrast to a full calibration procedure, our simple exemplar describes 'daily' or 'before-use' calibration procedures. Relying on the categories previously introduced, the most exact formulation would be: "before a planned sequence of end-measurements". For short, let us say "before-end-measurements", or "daily" calibration procedures.

The relations between full and daily calibration tests are as follows. The information provided at the end of a full calibration procedure—in particular the resulting uncertainty of measurement—is taken to be valid for *any subsequent measurements*, providing that no consequential change has occurred after the full calibration procedure. Thus, this information is taken to be valid both in the subsequent *daily calibration tests* and in the following *end-measurements* with respect to which the daily calibration tests have been undertaken. This information actually constitutes the *background* of the daily calibration tests as described through our simple exemplar. Moreover, it gives precise indications about some features of the results *of the daily calibration tests* that would show that a *new* full calibration *should be* performed. Such indications, expressed in our categories, typically correspond to a maximal allowed obtained/optimal discrepancy, which is specified as a result of the full calibration procedure. If the limit associated with this maximal allowed discrepancy is exceeded in the before-end-measurements calibration tests, a full calibration should be carried out.

A full calibration procedure is "systematic" in a double sense. First, the corresponding sequence of tests is less dependent on the *specificities* of the planned end-measurements than the before-end-measurements calibration tests (as we have seen Sect. 4.3, before-end-measurements calibration tests cannot be adequately thought independently of their relation to the end-measurements). Second, the sequence of tests must be performed *regularly*, even if no event has occurred (such as an accident) that leads to suspect that the measurer-token has drifted from the optimal working.

In order to indicate directions in which the simple exemplar of calibration would have to be completed, we shall now give some fragmentary insights about what may be a full calibration procedure, relying once more on the example of the equal-arm balance. As already pointed out, the full calibration procedure of a measurer-token includes a multiplicity of calibration tests directed toward different parameters which altogether

characterize the quality of that measurer-token. For an equal-arm balance, a full calibration procedure includes a test of repeatability, a test of measurement error across the range of the balance, a test of sensibility, a test of eccentric loading ("off-centre loading")… Let us say a little bit more about one of these tests, perhaps the most important one: the test which aims at evaluating the *measurement error across the range of the balance-token*.

4.6.4 A Calibration Test of the Measurement Error Across the Range of the Measurer-Token as an Estimation of the Accuracy: The Construction of a Calibration Curve

The parameter under test is *the measurement error across the range of the measurer-token*. It is often expressed in the form of a *calibration curve*. It corresponds to an estimation of the *accuracy of the measurer-token* across its range.

The test, illustrated with the case of the balance, goes as follows. Different etalons of increasing mass value are used and a measurement is performed by means of the same balance-token under test for each of these etalons. Then, a calibration curve is constructed. Each measured value is plotted against the corresponding nominal value of each certified etalon (standard mass),[12] and a curve is generated by a fit through the points.

The calibration curve provides two types of crucial information about the measurer-token. The first type concerns *each* point of the curve, *considered one by one*. For each point of the curve, given the corresponding nominal mass value of the etalon, which is known and certified, the calibration curve shows a departure from the measured-value to this nominal value—that is, what we have categorized as the obtained/optimal discrepancy. For each point of the curve, the calibration curve provides an estimation of the difference between, on the one hand the mass-value that should have been obtained with the balance-token, given the maximal precision obtainable with this balance-model in optimal working, *if* the balance-token coincided *with the optimal working*, and, on the other hand, the mass-value that is actually obtained with this balance-token applied to the etalon in this context. Thus, our description of a calibration test with an etalon in the simple exemplar corresponds to *any single point* of a calibration curve.

The second type of information provided by a calibration curve derives from the consideration of the curve *as a whole*. The consideration of the curve as a whole enables an assessment of the *linearity of response* of the balance-token across its range. Such assessment provides crucial clues about the possible *causes* of the observed deviations from the optimal working—as well as about the frontier between a badly calibrated and a defective balance-token. When trying to identify the potential sources of errors—the potential sources of the recorded deviations from the optimal working—a comparison of the slope of the obtained curve with the slope of the curve specified by the model for the optimal working can be performed. Suppose that the curve deviates from linearity more and more as the load increases. Practitioners will suspect, for example, a slight shift of the beam from its optimal position, with the result that the two arms of the balance can no more be considered as being of equal length.

According to the outcome of this calibration test, calibration operations may be needed. These operations can be either material or symbolic (using the categories introduced Sect. 4.4.2). Take the example just given at the end of the last paragraph, according to which a shift of the beam is suspected. In order to accommodate this shift, a *material* calibration operation would consist in displacing the beam with respect to its fulcrum, so as to reestablish the equal-length condition required by the model. The shift could also be

---

[12] More exactly—since for each standard mass, several repeated same measurements are performed most of the time—the ordinates points of the curve are constructed according to the statistical procedure mentioned in Sect. 4.6.2: for each, the mean value of the different measured values is used.

accommodated through *symbolic* calibration operations. In such option, mathematical corrections would be applied to the mass values obtained as the instrumental outputs of the end-measurements. The corrections would vary depending on the range of loading (depending on the mass value of the weighted object).

4.6.5 Concluding Remarks on Full Calibration

The previous elements should be elaborated in order to complete the characterization of calibration in UNSI practices. Our simple exemplar describes before-end-measurement calibration procedures—like checking the zero (blank calibration test) and checking the measured values obtained when an etalon is used (calibration test with etalon)—but leaves aside estimations of precision and uncertainty. Such before-end-measurements procedures are important because they inform about a possible instrumental drift with respect to the optimal working and enable to decide which calibration operations can be performed in order to accommodate the drift. However, they do not exhaust calibration tests in UNSI practices. They are conduced *given the result of a previously performed full calibration procedure*.

To conclude this section, let us stress that in such procedures, the trust we have in the certified etalons used is absolutely crucial, since the reliability of the conclusions of a full calibration procedure *essentially depends* on the certified data (nominal values and uncertainties) attached to these etalons. This echoes the remarks we made at the end of Sect. 3.1.4, where we stressed that calibration in UNSI practices presupposes the results of professional metrologists.

Using our simple exemplar of calibration as a framework, we are now going to investigate a more complex case of calibration in UNSI practices.

5 Calibration in X-Ray Diffraction Experiments in a Nanoscience Laboratory

5.1 General Features and Interest of the More Complex Case

The case deals with X-ray diffraction experiments and the related calibration procedures currently undertaken at the Jean Lamour Institute (hereafter IJL) in Nancy. The IJL has an expertise in the growth and production of very thin (nanometric) films with specific properties. The experiments we are going to discuss concern the characterization of the internal structure of mono-crystalline thin-films by means of an X-ray diffractometer.

These experiments can be assimilated to UNSI practices: we can consider that they investigate relatively well-understood natural phenomena with standardized instruments. This is obvious as far as the *means of the investigation* is concerned: the X-ray diffractometer used at the IJL is an already standardized, widely used instrumental device. This is perhaps less obvious as far as the *phenomena under inquiry* are concerned, since one aim of practitioners at the IJL is to produce material thin-films with *new* interesting properties. To discuss seriously the issue of the 'degree of novelty' of the phenomena under inquiry would require more developments, but here, we will take the point for granted, relying on the fact that the kinds of material samples under study, namely single crystals, are already well-understood materials (with an available well-characterized repertory of kinds of crystalline structures, etc.).

The calibration activities we are going to study are *more complex* than the one analyzed in part 4 relying on the example of the balance, because they involve a more complex

instrumental device *in the sense that* the target of calibration is *treated as* an instrumental system *composed of multiple parts*.

Compared with the simple exemplar of calibration illustrated by means of the balance, this more complex case of calibration is peculiar and instructive in (at least) two important respects. First, the fact that it involves, as the target of calibration, a treated-as-composed instrumental system, directs attention to the relations that might hold between calibration of an instrumental device *as a whole* and calibration *of its parts*. Second, this case points to instances in which the target of calibration is *not* a measurer. Such a situation has important epistemological consequences, and we will examine some of them.

5.2 The End-Measurements

In order to characterize the calibration procedures related to the X-ray experiments undertaken at the IJL, the first thing to do is to describe the end-measurements, since as stressed in Sect. 4.3, different end-measurements may have different impacts on the details of the calibration of one and the same measurer-token.

In the particular case analyzed below, we consider two (related) kinds of end-measurements successively undertaken by practitioners. In the first ones, the aim is to determine the inter-reticular distance d between the crystalline planes of the thin-films synthesized at the laboratory at a given time (say t_1). This is informative concerning the growth process and the structural quality of the thin-film sample that has been produced.

In the second kind of end-measurements, the aim is to check the stability of the sample over time (the possible evolution of the quality of the thin mono-crystalline layer produced and characterized at t_1). This evaluation involves the reiterations of similar measurements at different moments posterior to t_1 (say t_2, t_3, etc.), and the creation of conditions that enable to compare (i.e., to situate on one and the same scale) the values respectively obtained at t_1, t_2, t_3, etc. In the next sections, we analyze the situation in reference to the first kind of end-experiments. The second kind of end-measurements enters into play subsequently (from Sect. 5.5), in relation to the calibration procedure which intends to master the intensity delivered by the X-ray source.

Before performing end-experiments aiming at the characterization of the inter-reticular distance d of a crystalline thin-film, practitioners want to test the reliability of their X-ray diffractometer-token. With respect to these end-measurements and to the robustness of their results (the value attributed to d), the target T of the calibration procedure is *the diffractometer-token as a whole*—since the end-measurements of course use the whole diffractometer-token. In what follows, we first analyze the diffractometer-token involved at the IJL by means of the conceptual tools introduced so far. Then we say a word about the calibration of this diffractometer-token as a whole, before giving a more detailed characterization of one *particular step* of this procedure, directed toward one *particular part* of the diffractometer: the X-ray source-token.

5.3 The X-Ray Diffractometer Used at the IJL: An Analysis by Means of the Instrument-Frame

Table 3 below is obtained by applying the conceptual frame offered by Table 1 to the X-ray diffractometer-token used at the IJL. The result is a synoptic characterization of this diffractometer by means of the conceptual tools previously introduced. A number is associated to some items of Table 3: it refers to further explanations (given below) about the corresponding item.

Table 3 A conceptualization of the X-ray diffractometer-token used at the IJL

Generic type (1)	Generic name	X-ray diffractometer
	Typical **measuranda** for the measurement of which the generic type is designed (**1a**)	Structural properties of crystalline material samples (ex. inter-reticular distances; kind of structure…)
	Fundamental scientific principle (**1b**)	Diffraction of X-ray by a crystal; Bragg's law
	Kinds of quantities typically involved in the mesuranda (**1c**)	Angle θ at which a given intensity is diffracted
		Intensity of diffracted X-ray (I_{diff})
		Diffraction spectra $I_{diff}(\theta)$
Model (2)	Name and description (**2a**)	X'pert Pro MRD PANalytical
		Four-circle diffractometer
		High resolution X-ray diffractometer for thin-film analysis
		Components: see the list below, points (i) to (vi)
	Scientific scenario (**2b**)	See description in (2b)
	Instrumental outputs (**2c**)	Graphs convertible in diffraction spectra $I_{diff}(\theta)$
	Predicted deviations with respect to the optimal working (**2d**)	Decrease of the intensity of the X-Ray beam over time

(1) **About the generic type of the X-ray diffractometer**

(1a) The function of an X-ray-diffractometer is the determination of some structural properties of crystalline material samples, for example the inter-reticular distances between crystalline planes, the kind of crystalline structure (cubic, etc.) or the like. These are the kinds of properties for the measurement of which the generic type have been designed, and the kind of properties that practitioners typically intend to measure (i.e., **typical mesuranda**) when they use an X-ray-diffractometer.

(1b) The fundamental **scientific principle that defines the generic type** is the phenomenon of diffraction of X-ray by a crystal. The phenomenon of diffraction is the result of a particular type of interaction between radiation and matter. An X-ray beam is conceived as a certain kind of radiation constituted of photons (the range of X-ray wavelengths is placed between the ultraviolet region and the region of γ-rays emitted by radioactive substances). When an X-ray beam is directed on a crystal, different physical interactions occur. In particular, some photons of the incident X-ray beam interact with the electronic clouds of the atoms of the crystal and are deflected without a loss of energy. These deflected photons constitute the scattered radiation. The scattered radiation presents certain characteristic properties *which depend on the internal structure of the crystal*. More precisely, some properties of the scattered X-rays (the direction and the intensity of the scattered beam, called the angle of diffraction and the diffracted intensity) are connected to some properties of the internal structure of the crystal (such as, for example, the inter-reticular distance *d* between crystalline planes), through determined and well-known laws. These laws are, for the X-ray diffractometer, the equivalent of the lever principle for the balance. They correspond to the fundamental scientific principle which underlies and define the *type of measurer* that an X-ray diffractometer is.

In the case of end-measurements *aiming at the determination of the inter-reticular distance d between crystalline planes*, the relevant law is the fundamental law known as the "Bragg's law":

$$2d \sin \theta = n\lambda$$

d is the inter-reticular distance for one stacking of planes; θ is Bragg's angle, that is, the angle between the incident X-ray beam and the reticular planes; n is an integer; λ is the wavelength of the monochromatic X-rays beam (the interval of λ of particular usefulness in crystallography ranges between 0.4 and 2.5 Å).

The Bragg's law relates d, the quantity intended to be determined, to a measured-quantity, namely the angle at which the X-ray beam is diffracted and collected (see Fig. 2).

Relying on Fig. 2, it appears that as far as scientists are able to generate an X-ray monochromatic beam of a certain known intensity I_{inc}, to direct the incident intensity I_{inc} on a crystalline sample according to a certain known angle θ, and to record the intensity I_{diff} of the X-rays diffracted by the sample at an angle 2θ, scientists are able to determine the inter-reticular distance d between the crystalline planes. This is precisely what a diffractometer accomplishes: a diffractometer is a means to convert the diffracted intensity as a function of the angle ($I_{diff}(\theta)$) into structural properties of a crystal.

(1c) From this follows that the instrumental outputs of an X-ray diffractometer must be interpretable in terms of diffracted intensities and angles of diffraction. Thus, the **kinds of quantities typically involved in the mesuranda** associated with measurements with a diffractometer are angles and intensities of X-rays. More precisely, the determination of the mesurandum is achieved through a determination of the angles at which an intensity is diffracted, that is, through the determination of the different $I_{diff}(\theta)$s. This is what is called a diffraction spectra $I_{diff}(\theta)$. From such diffraction spectra, practitioners can infer, through Bragg's law, the values of the mesurandum under interest, in our case, the values of the inter-reticular distance d.

(2)   **About the model of the X-ray diffractometer**

(2a) The **model of the diffractometer-token** currently used at the IJL is a high-resolution four-circle diffractometer, sold under the technical name "X'pert Pro
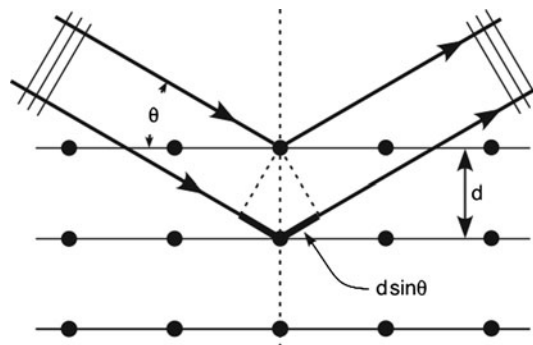
Fig. 2 Bragg diffraction from a cubic crystal lattice. Plane waves incident on a crystal lattice at angle θ are partially reflected by successive parallel crystal planes of spacing d. The superposed reflected waves interfere constructively if the Bragg condition $2 d \sin \theta = n \lambda$ is satisfied

[Title and image reproduced from <https://commons.wikimedia.org/wiki/File:Bragg_diffraction.png>; GNU General Public License]

MRD PANalytical" (PANalytical is the trademark). Figure 3 shows a picture of the corresponding diffractometer-token. Figure 4 offers a scheme with legends of the main components.

Fig. 3 The diffractometer-token used at the IJL
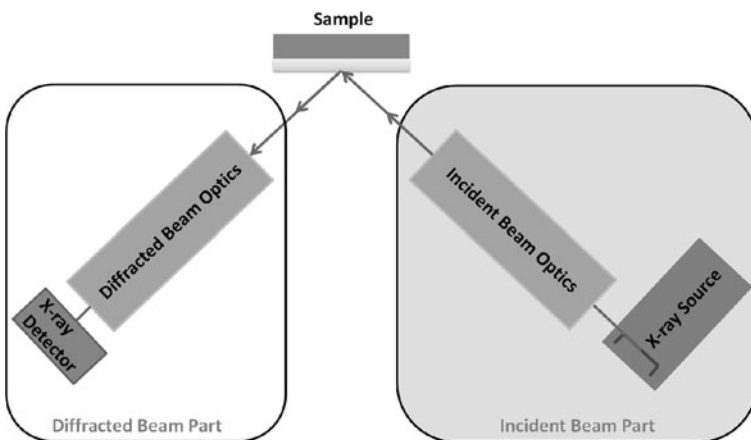

Fig. 4 Scheme of the main components of an X-ray diffractometer

The **components of this diffractometer-model** are (from the right to the left):

i.   The **X-ray source**. Here it is a source-token of the type 'Cu-anode sealed-tube'. The source produces an incident monochromatic X-ray beam with an intensity I_{inc}, which is directed on the crystalline thin-film sample to be analyzed.

ii. The **sample holder** where the crystalline thin-film sample is laid down.

iii. The **X-ray detector**. Here it is a detector-token of the type "PIXcel detector", which is a second generation solid-state detector based on pixel technology. The detector collects the X-ray beam diffracted by the crystalline sample in a given direction, and indicates, for each angle, the diffracted intensity I_{diff}, that is, the number of diffracted X-ray photons.

Other instrumental modules are involved and play an important role.

iv. A **goniometer system**. Its function is the determination of the relative positions of the source, the crystalline sample and the detector. The control of these positions is crucial for the interpretation of the instrumental outputs.[13]

v.   Some **optical devices**. They include: a monochromator, used to produce, as its name indicates, a *monochromatic* incident beam; and a collimator, used to filter the diffracted X-rays in order to collect only the X-rays photons arriving at 2θ (see Fig. 2).

vi. A diffractometer **control software program**. This program carries out all real-time instrument control functions. For example, it drives the goniometer motors, monitors the detector system, etc.

(2b) This model of diffractometer converts certain humanly performed operations (namely here: to put the crystalline thin-film on the sample holder; to switch the X-ray beam on; etc.) into one definite value of the mesurandum under interest (here: the value of the inter-reticular distance d of a given thin-film). The conversion relies on some scientific theories, here the theory of diffraction of X-rays by crystals (which itself involves crystallographic theories and wave theories).

The passage from the above-mentioned humanly performed operations to a determined value d_1 of the mesurandum d is realized through a **scientific scenario**, based on Bragg's law, of the kind described above in point (1b), *specified according to the contextual characteristics of the model*: the monochromatic X-ray beam emitted by the Cu-anode sealed-tube passes through the monochromator; the resulting monochromatic X-ray beam of wavelength λ reaches the thin-film according to the direction θ prescribed by the goniometer; the proportional detector collects the scattered intensity for an interval of angles prescribed by the control software program… Eventually, a graph is produced as the instrumental output of the diffractometer; this graph is then interpreted in terms of a diffraction spectrum $I_{diff}(\theta)$, from which the inter-reticular distance d is inferred.

(2c) The **instrumental outputs** of a X-ray diffractometer are most of the time graphs. This is the case for the model used at the IJL. The corresponding graphs are interpreted as diffraction spectra $I_{diff}(\theta)$, that is, as the number of X-photons diffracted by the crystalline thin-film for different directions of the incident beam. Typically, the graphs present a peak of intensity for the Bragg's angle involved in the Bragg's law.

(2d) The scenario sketched above (point 2b) is one part of the definition of the diffractometer-*model*. Its most detailed version provides the characterization of the

---

[13] Unfortunately, we cannot give the full details here: the analysis of the corresponding calibration procedures would be beyond the scope of this paper.

> **diffractometer in optimal working** (i.e., what is at best obtainable with a diffractometer-token generated according to this model). The other part of the definition of the diffractometer-model involves the **possible deviations with respect to the optimal working** that are anticipated according to the model. For example, a decrease over time of the intensity of the X-ray beam delivered by the diffractometer is predicted.

In the present section, we have applied the frame of Table 1, first constructed in relation to the case of the balance, to *another* instrumental device, namely, the X-ray diffractometer-token currently used at the IJL. We hope that the corresponding analyses have shown how Table 1 enables to achieve a fine-grained characterization of newly-considered instrumental devices, and have been sufficient to convince the reader of the interest of the conceptual tools we have introduced in this purpose.

Having characterized the measurer-token under interest—the X-ray diffractometer-token—we are now going to turn to the issue *of its calibration*. We start with the calibration of the diffractometer-token *as a whole*, using the simple exemplar (as summed up in Table 2) as a tool. At this level, our analysis will not go into the details, but we hope it will show the clarifying power of the simple exemplar as an analytic tool. Our aim is to give a panoramic view of the situation concerning the calibration *of the diffractometer-token as a whole* (Sect. 5.4), before we focus on *one particular calibration step* directed toward one particular component of the whole, the X-ray source-token (Sect. 5.5).

5.4 Calibration of the X-Ray Diffractometer as a Whole: An Analysis by Means of the Simple Exemplar Used as a Framework

Let us now show how the simple exemplar of calibration, first constructed and illustrated by means of the simple case of the balance, can be used as an helpful tool in order to understand *more complex* cases—here, the calibration of the X-ray diffractometer-token used at the IJL in the perspective of ends-measurements aiming at the determination of the inter-reticular distance d of a crystalline thin-film. The simple exemplar is a useful tool, to the extent that it suggests predetermined questions to ask and predetermined answers to discuss (See Table 2). In cases in which the answers that applied to the simple exemplar also apply to the new case under scrutiny, the new case will be quickly and straightforwardly understood, relying on the already available characterization of the simple exemplar. In cases in which the answers that applied to the simple exemplars do *not* apply as such to the new case under scrutiny, this will be instructive as well: this will help to understand the specificity of the new case by contrast with the already well-understood case of the simple exemplar.

Table 2 first invites us to ask what the *target of calibration* is (first question of our four-question frame). We have already answered this question at the end of Sect. 5.2, after having explained what has to be taken into account in order to provide an answer: the target is the X-ray diffractometer *as a whole*, *because* it is the X-ray diffractometer as a whole which is used *in the end-experiments*.

The second question of our four-question frame calls attention to the *presuppositions about the target of calibration*, and Table 2 proposes two answers which apply without any need of accommodation to the X-ray diffractometer: P1 and P2 hold. In other words: (i) the type and the model of the diffractometer are unproblematic (P1)—which, by the way, implies that the same holds for the type and the model *of each component* of the

diffractometer; (ii) the diffractometer-token is assumed to be not deficient—which implies that each of its constituents works properly (P2).

Similarly, the answer to the question of the *aim of calibration* (provided in Table 2), applies as such to the calibration of the diffractometer as a whole: the aim of the calibration procedure of an X-ray diffractometer is to master the possible discrepancy between the diffractometer in optimal working and the diffractometer-token used in a given context in the perspective of certain end-measurements.

The *distinction between calibration tests and calibration operations* also clearly applies here: to master the discrepancy means first, to evaluate, and second, if needed, to accommodate the distance of the diffractometer-token from the optimal working as defined by the diffractometer-model. After examination of the operations currently performed at the IJL, it appears that both material and symbolic operations may be involved.

Table 2 furthermore raises the question whether *calibrations with etalons* are undertaken or not in the practices under scrutiny, and the answer is positive: physicists of the IJL frequently use crystalline samples of known properties in order to evaluate the possible shift between their actual token of diffractometer in a given context and the optimal working as specified by the model. And of course, when such calibration tests with etalons are performed, presuppositions P3 (reliability) and P4 (adequacy) about the etalon are taken for granted.

Finally, Table 2 raises the question of *blank calibration tests* of an RX-diffractometer-token. The issue of blank calibration tests, applied to the diffractometer case, is perhaps not so straightforward at first glance. It will be discussed below in Sect. 5.5.2, point (8).

All in all, relying on the simple exemplar of calibration as summed up in Table 2, it appears that most of the features previously well-characterized and developed in reference to the simple case of the balance, *mutatis mutandis* apply as well to the calibration of the newly considered, more complex case of the X-ray diffractometer. To that extent we immediately achieve, thanks to the framework of the simple exemplar, a detailed characterization of central features of what is at stake in the calibration of a newly encountered case.

There are, however, some specificities of the calibration of the X-ray diffractometer compared to the simple exemplar of calibration. Before we discuss some differences at the level of the *details* of the calibration procedure, let us consider, at the *most general* level, the main difference between the simple exemplar of calibration and the more complex case of the calibration of the X-ray diffractometer. The main difference comes from the clause that the diffractometer is a *complex* instrumental device. To understand this difference, we must specify the status of the complexity involved here. Such complexity must not be understood as an 'absolute' property of the X-ray diffractometer (the absolute property of 'being composed of parts'). Rather, it is a *practical* complexity, *relative* to the practitioners' concrete actions. Actually, the balance itself could, like the diffractometer and like any other instrument, be decomposed in multiple parts—even if we have treated it most of the time as an 'atomic measurer' for the sake of simplification. If the diffractometer is *treated here as* a composed measurer rather than as an 'instrumental atom', it is because at the IJL, experimenters *treat it as composed* in their practices of calibration, *in the sense that they perform different calibration tests directed toward different parts of the diffractometer*. This 'spatial complexity' also implies a *temporal* complexity: the calibration of the diffractometer *as a whole* involves a long sequence of actions, which can be decomposed in multiple *sub-sequences*, or calibration *steps*,[14] more especially focused *on*

---

[14] This temporal complexity is almost always present in real procedures of calibration of a balance. In this respect as in many others, the characterization of the simple exemplar has to be understood as a simplification. This is obvious from the brief considerations in Sect. 4.6.3 about the procedures of 'full' calibration

this or that sub-part of the whole (with possible constraints on the order according to which the steps have to be implemented).

Such spatial and temporal complexity constitutes the main difference between the calibration of the X-ray diffractometer and the simple exemplar of calibration. A complete analysis of the calibration of a treated-as-composed instrumental system, as is the X-ray diffractometer in current uses at the IJL, would require to scrutinize the multiple calibration steps involved, their order and possible inter-relations, and the way they are combined in global judgments of the type "the diffractometer *as a whole* is correctly calibrated (or not)". The construction of more complex exemplars of calibration would be needed to take these aspects into account. In what follows, we provide fragmentary insights on these aspects by considering what we call "co-calibrations" (Sect. 5.6). Note that regarding the aim of taking into account the complexity of calibration of a treated-as-composed measurer-token, the simple exemplar (Table 2) as well as the instrument-frame (Table 1) are helpful to the extent that they can be used to analyze and illuminate each calibration step (see Sect. 5.5 for an illustration applied to the case of the X-ray source).

Having stressed the complexity of the whole sequence of calibration of such a treated-as-composed measurer, we are now going to focus on one particular step of this complex sequence. This step, directed toward the X-ray source, will be analyzed in details, using one more time the conceptual frames given in Tables 1 and 2.

5.5 One Step in the Calibration of the X-Ray Diffractometer as a Whole: A Calibration Procedure Directed Toward the X-Ray Source

In the calibration step under scrutiny, the target of calibration is the X-ray source-token currently used at the IJL as a part of the diffractometer-token. Thus first of all, let us characterize this X-ray source by means of Table 1.

5.5.1 The X-Ray Source Used at the IJL: An Analysis by Means of the Instrument-Frame

The X-ray source, contrary to the instrumental devices considered so far, is not a measurer. To take this into account, Table 1 needs to be slightly adapted. Take the cell entitled: "Mesuranda for the measurement of which the generic type is designed" (second column, line 2). Since the X-ray source is not a measurer, it is strictly speaking not correct to say that it has been designed for the purpose of some determined *measurements* which typically intend to evaluate some specifiable *mesuranda*. True, the X-ray source is designed to deliver an X-ray beam of a given intensity. Thus, the kind of quantity centrally and typically associated with an X-ray source is an X-ray intensity. But this kind of quantity is not well described as a kind of quantity "typically involved in the mesuranda", as Table 1 mentions in its second column, line four. All in all, it would be more adequate to talk about the phenomena for the generation of which the X-ray source have been designed, and about the quantities typically involved in these phenomena. Hence we accommodate Table 1 accordingly: we replace "Mesuranda for the measurement of which the generic type is designed" by "Phenomena for the production of which the generic type is designed", and "Kinds of quantities typically involved in the mesuranda" by "Kinds of quantities typically involved in the phenomena intended to be produced". Next we complete the

---

Footnote 14 continued

of a balance, which clearly show the temporal complexity involved, through the mention of multiple calibration steps (even if in this case, the different steps are directed toward different parameters rather than toward different parts of the balance).

Table 4 A conceptualization of the RX-source-token used at the IJL, by means of a slightly modified version of the instrument-frame

Generic type	Generic name	Sealed-tube X-ray source
	Phenomena for the production of which the generic type is designed	Monochromatic X-ray radiation
	Fundamental scientific principle	Generation of X-rays by irradiation of a material target with a beam of high-energy charged particles (such as electrons)
	Kinds of quantities typically involved in the phenomena intended to be produced	X-ray intensity X-ray wavelength
Model	**Name and description**	Cu-anode sealed-tube
	Scientific scenario	A filament is heated to produce electrons which are then accelerated in vacuum by a high electric field. The high energy electrons are then directed toward a metal target, namely a copper anode. The irradiation of this Cu-anode by electrons generates a beam of X-rays.
	Instrumental outputs	-
	Predicted deviations with respect to the working	Decrease over time of the intensity of the X-Ray beam generated by the source (ageing)

accommodated frame for the X-ray source used at the IJL. Table 4 is obtained as a result. Table 4 gives a synthetic characterization of the target of the calibration step under scrutiny—the X-ray source—according to a frame which is by now familiar enough to be provided without further comments, beyond the slight adaptations just described.

5.5.2 A Calibration Directed Toward the X-Ray Source: An Analysis by Means of the Framework of the Simple Exemplar

Table 5 gives an overview of the results of our analysis concerning the calibration step directed toward the X-ray source. The details of this analysis are provided below, in relation to the different numbers included in the cells of Table 5. In italicized text, we indicate the elements that are peculiar to this case, due to the two facts that (a) the target of calibration corresponds to this *particular* model of X-ray source *rather than another instrumental device* (such as a balance), and (b) that the calibration under scrutiny is performed in the perspective of these *particular* end-experiments *rather than another ones*. In underlined text, we stress the differences manifested by this calibration step with respect to the simple exemplar of calibration. Thus the underlined items direct the attention on aspects with respect to which the simple exemplar of calibration is not sufficient or in need of adaptations.

(1) Our aim is to characterize a particular case of calibration in UNSI practices, so we must specify the first line of Table 2 accordingly (in italics).

(2) As stressed in Sect. 4.3, the details of a particular activity of calibration directed toward one and the same instrument-token might differ according to the **end-measurements**. To that extent, the characterization of a particular activity of calibration must *first of all* examine the end-measurements.
     The calibration step under scrutiny (that is, the calibration step of the diffractometer directed toward the X-ray source) is not systematically undertaken at the IJL. It is

Table 5 Analysis of one calibration step directed toward the X-Ray source-token used at the IJL by means of an elaborated version of Table 2

Calibration in UNSI practices *aiming at the determination of structural properties of mono-crystalline thin-films by means of an RX-diffractometer of the model described in Table 4* **(1)**		
End-measurements (2)	*To assess the stability of a crystalline thin-film sample through time; to compare the structural quality of one and the same sample at different moments (t_n, t_{n+1}, etc.), which implies to compare the height and width of the diffraction peaks obtained for this sample at t_n, t_{n+1}, etc.*	
Target (3)	*The X-ray source-token (see Table 4 for a detailed characterization)* **(3a);** <u>Not a measurer</u> **(3b)**	
Presuppositions (4)	**About the target (4a)** (P1) Unproblematic model (P2) Non-defective token	**About the etalon** -
Aim (5)	To master the obtained/optimal discrepancy *of the intensity delivered by the X-ray source-token* **(5a)** <u>Generalized definition: to master the 'obtained at t_n/obtained at t_{n+1}'</u> difference between *the intensity values*, taking into account, as a reference point, *the intensity value* corresponding to the optimal working **(5b)**; if t_n corresponds to the optimal working, the general definition reduces to the definition given for the simple exemplar, i.e., to (5a)	
Procedure	**Tests (6)** ***Rough description of the calibration tests****: direct measurements of the intensity of the X-ray-source at different moments (t_n, t_{n+1}, etc.)* **(6a)** <u>The calibration tests inevitably require another instrumental device in addition to</u> *the X-ray source* <u>under test, here</u> *an X-ray detector* **(6b)** ***Refined description of the calibration tests****: compare (i) the instrumental outputs obtained at t_2, as the result of direct measurements of intensity performed on the X-ray source-token* <u>*with the X-ray detector-token, in a configuration of the diffractometer as similar as possible to the one involved in the end-measurements (same optics etc.)*</u>*, with (ii) the instrumental outputs that should have been obtained for similar measurements in an optimal working of the X-ray source according to the model of the source, and that have been obtained for measurements performed on this source-token at t_0* <u>*with the X-ray detector-token*</u> **(6c)**	**Blank (8)** <u>*A blank calibration test of the diffractometer but not a blank calibration test of the source*</u> **With etalon**: -
	Operations (7)	**Material**: - **Symbolic:** *mathematical correction: the diffracted intensities of the diffraction spectra obtained in the end-measurements are multiplied by a certain factor* **(7a)** <u>The symbolic operations cannot be applied to the *X-ray source-token* which is the target of the calibration. They are inevitably applied to the instrumental outputs of the *X-ray detector* which plays the role of a measurer in the calibration procedure directed toward the *X-ray source*</u> **(7b)**

required *only* with respect to *some types* of end-measurements (this gives an additional illustration of the crucial importance of the end-measurements, regarding not only the *conception and relevant features* of the calibration activity, but also its *very existence*). The calibration step under interest is required in relation to the second kind of end-measurements mentioned Sect. 5.2. Such end-measurements intend *to* assess *the stability of the thin-film sample through time*. They therefore involve a *comparison* of the structural quality of one and the same mono-crystalline thin-film *at different times* (say t_1 and t_2). When such a comparison is projected, practitioners must care about the *values* of the diffracted intensities, and not just about the fact that *there is* an intensity peak at a certain angle, no matter its value (as it was the case for the determination of the inter-reticular distance at a given moment). Since the diffracted intensities depend on the incident intensity delivered by the X-ray source, a calibration step is required in order to control the intensity of the beam delivered by the X-ray source.

More precisely, to know whether or not the quality of a thin-film has been altered or not since its first production, practitioners must reiterate, at a subsequent time t_2, a structural characterization of the same kind as the one previously performed on the same thin-film at an anterior time t_1. The instrumental outputs of the experiments which provides the basis for such structural characterization are, as we saw, graphs interpretable as diffraction spectra $I_{diff}(\theta)$. So once the end-measurements have been performed at t_2, practitioners are left with (at least) two diffraction spectra, $I_{diff}(\theta)t_1$ and $I_{diff}(\theta)t_2$, that they have to compare. However, the height and width of the diffraction peaks recorded at t_1 and t_2—from which the diffracted intensities at t_1 and t_2 must be derived—are directly comparable (say by superposition of the two graphs obtained at t_1 and t_2 for the same sample) *only under the condition* that, at t_1 and t_2, the X-ray source-token delivered an incident beam having the *same* intensity I_{inc}.

(3) This condition does not necessarily hold. Some accident or other unexpected event may have altered the production of X-ray photons by the source. More fundamentally, as indicated in Table 4 (last line), according to the specifications of the model *of the X-ray source*, a deviation with respect to the optimal working *of the X-ray source* is expected, corresponding to a decrease over time of the intensity of X-ray beam generated by the source ("ageing"). Consequently, with respect to any end-measurements for which the values of the intensities (or height and widths of the diffraction peaks) matter, a calibration procedure must be undertaken at a point or another, in order to control the intensity of the incident X-ray beam. The **target** of this calibration step is the X-ray source-token **(3a)**.

(4) The calibration step in question is undertaken under two **presuppositions about the target of calibration (4a)**: the X-ray source-token is assumed to be (i) of unproblematic type (P1) and (ii) not defective (P2) (we already saw in Sect. 5.4 that this was the case for all the components of the X-ray diffractometer-token).

(5) The **aim of this calibration step** is to master (to evaluate in the calibration tests, and to take into account through the calibration operations) the difference between the values of the intensity delivered by the X-ray source, on the one hand at t_1, and on the other hand at t_2. Such aim can be abbreviated as "to master the obtained/optimal discrepancy of the intensity of the X-ray source-token" **(5a)**, with some qualifications that will be provided below (see point (6)), once the calibration tests will have been specified.

(6) The **calibrations tests** undertaken in order to achieve the previous aim consist in direct measurements of the intensity delivered by the X-ray source at different relevant moments (the measurements are 'direct' in the sense that no diffracting sample is involved in the test) **(6a)**.

Back to the aim of calibration (5).
Having specified the calibration tests, let us come back to the aim of the calibration step directed toward the X-ray source. Two cases can be differentiated.

a. Suppose that t_1 corresponds to the first setting up of a new X-ray source-token just after its manufacturing, and that the source-token presents no default with respect to the specifications of the model (practitioners of course check this conformity when they receive a new instrument). In that case, the X-ray source-token at t_1 coincides with the X-ray source-model *in optimal working*. The direct measurements of intensity realized at t_1 indicate a maximal incident intensity value I_{diff} obtainable with this source-token. The diffraction spectra recorded at t_1 (typically with etalons) constitute a reference point which characterizes the optimal working regime. Referred to this time t_1 at which the X-ray source-token coincided with the optimal working as defined by the model, the aim of the calibration tests and operations directed toward the same X-ray-source-token at a subsequent time t_2 can be characterized, in complete conformity with the definition proposed above for the simple exemplar, as **(5a)**: to master the obtained/optimal discrepancy—here the discrepancy of the intensity of a sub-part of the diffractometer as a whole, the X-ray source. If the X-ray source-token at t_2 still conforms to the X-ray source-model in optimal working, the diffraction spectra obtained at t_2 are *directly* comparable to the one obtained at t_1 in optimal working (by simple superposition of the graphs).

b. However, in practices at the IJL, the comparison is not always directly and explicitly referred to the 'initial' time where the source-token coincided with the optimal working. The comparison often contrasts what holds at a time t_3 with an anterior time t_2 *where the source has already shifted from the optimal working*. In such cases, our initial definition of the **aim of calibration** as the obtained/*optimal* discrepancy (here the decrease of the X-ray source-token intensity) must be **slightly modified in the sense of a generalization and an increased complexity**. In such cases, the gap that practitioners aim to assess in the calibration test is, more generally **(5b)**, an 'obtained at t_n/obtained at t_{n+1}' difference. When t_n corresponds to the optimal working, the latter general definition of the aim of calibration reduces to the definition **(5a)** given for the simple exemplar.

    Actually, this general definition is still a simplification, since in fact, what is at stake is not just two moments t_1 and t_2, but a whole temporal trajectory, which might involve an indefinite number of relevant moments, say t_0 (= optimal working), t_1, t_2, …, t_n. For the sake of simplicity, however, we will reduce the problem of the calibration directed toward the X-ray source to a situation in which (i) *only two* moments are compared (say t_2 and t_0) and (ii) the first moment t_0 *corresponds to the optimal working*. In such a situation, the aim of calibration coincides with the aim defined for the simple exemplar (i.e., to master the possible obtained/optimal discrepancy).

Back to calibration tests (6).
As already stressed in Sect. 5.5.1, the target of calibration, namely the X-ray source, is *not* a measurer **(3b)**: the X-ray source does not deliver instrumental outputs directly accessible as measurement results. The fact that the X-ray source is not a measurer has repercussions at the level of the calibration tests and introduces, at this level, an important difference with respect to the simple exemplar **(6b)**: the calibration tests of the X-ray source inevitably require, *in addition* to the X-ray source under test, *another*

instrumental device which plays the role of a measurer—that is, which records the intensity delivered by the X-ray source-token at a time t_2 close to the end-measurements. In other words, the calibration tests inevitably require an *X-ray detector*. In the calibration tests performed at the IJL, this additional device is the X-Ray detector-token which is part of the diffractometer and is intended to be used in the end-measurements.

Taking that into account, a **refined description of the calibration tests** goes as follows (**6c**). Direct measurements of intensities are performed, at t_2, on the X-ray source-token *with the X-ray detector-token*, in a configuration of the X-ray diffractometer as similar as possible to the one involved in the end-measurements (e.g., using the same detector, the same optics, etc.). The instrumental outputs of the detector obtained at t_2 (which coincide with the instrumental outputs of the *diffractometer* obtained at t_2) are then compared to the ones that should have been obtained for an optimal working of the X-ray source according to the model of the source, and that have actually been obtained at t_0 with this X-ray source-token, this detector-token, these token of optics, etc.

(7) Having described calibration tests, let us turn to the issue of **calibration operations**. If no significant difference is found between the intensity values obtained at t_0 and at t_2, no calibration operation is required. The diffraction spectra obtained in the past at t_0 are directly comparable to the diffraction spectra that will be obtained in subsequent end-measurements at a time close to t_2. A simple superposition of the spectra at t_0 and t_2 enables us to compare the values of the diffracted intensities and, on this basis, to decide whether or not the quality of the mono-crystalline thin-film has been altered between t_0 and t_2.

If, on the contrary, a significant difference appears between the X-ray intensities at t_0 and t_2, then, some calibration operations must be applied. In the present case, they correspond to **symbolic operations** (**7a**) applied to the instrumental outputs of the X-ray diffractometer (or, equivalently, to the instrumental outputs *of the X-ray detector*). More precisely, the diffracted intensities of the diffraction spectra obtained in the end-measurements are multiplied by a certain factor (a proportionality rule is applied). The fact that the X-ray source which is the target of calibration is not a measurer introduces a difference with respect to the simple exemplar as regards the way the symbolic operations are applied (**7b**). The proportionality factor cannot be applied directly to the instrumental outputs of the X-ray source, which is the target of the calibration test, since these outputs are not directly accessible but must be recorded by the mediation of the X-ray detector. The mathematical correction that is needed, in order to take into account the intensity decrease of the X-ray source-token, are thus *inevitably* applied, not directly to the target of calibration, but to the instrumental outputs *of another instrumental device which is a measurer, here the X-ray detector*.

Can we say that once the symbolic operations have been applied, practitioners have calibrated *the X-ray source*? Recall that the source is *not a measurer and* the operations are *symbolic* operations. Consequently, these operations are inevitably applied to the instrumental outputs *of the detector*, rather than directly to the source. Since the corrections are applied to the outputs *of the detector*, we might hesitate to say that practitioners have calibrated *the source*. However, nothing forbids us to say that practitioners have calibrated the source *by applying the symbolic corrections to the outputs of the detector* (or equivalently, by applying intellectual corrections to the

outputs *of the diffractometer*). Actually, this seems a perfectly adequate way of conceptualizing the situation. To be convinced, imagine a structurally similar situation which differs only in that *material* operations are involved rather than symbolic operations. We would have no problem to say that practitioners have calibrated *the X-ray source* (since the material operations would be physically applied directly to the source). The fact that there is no fundamental structural difference between these two cases encourages us to conclude that it is not only permitted but also appropriate to say that practitioners have calibrated *the X-ray source* (or more exactly, the intensity of this source).

(8) Finally, how should the calibration tests involved here be situated with respect to the two species previously distinguished, namely **blank calibration tests and calibration tests with etalons?** Clearly, no etalon is involved in the calibration step under discussion, so the first species is not relevant. But what about the situation of the calibration step here involved with respect to a calibration test *of the blank type?* The answer is not straightforward and implies conventional terminological/conceptual decisions that can be motivated, but are not universally compelling. We experienced divergences at this level in the PratiScienS team during the course of our inquiry. Let us formulate and substantiate the decisions on which we eventually agreed.

First, the calibration procedure described above can be identified with a *blank* calibration to the extent that it conforms to the main elements mentioned in the definition of what a blank calibration is (given in Sect. 4.4.1). The calibration tests under scrutiny are indeed (i) measurements of the same kind as the end-measurements: same X-ray source-token, same X-ray detector-token, same optics (in a word, same diffractometer-token); same kind of measured-quantity (X-ray intensity); and (ii) measurements undertaken in the absence of any object of the same kind as the object the end-measurements aim to characterize (a diffracting crystalline sample). The end-measurements and the calibration tests are supposed to differ only in regard to the condition "with/without a diffracting sample", all other things being equal. On this basis, it seems adequate to describe what is at stake as a *blank* calibration *of the diffractometer*.

Can we also talk about a *blank* calibration *of the X-ray source*? Such does not seem meaningful, despite the fact that the calibration step is indeed directed *toward the X-ray source*. This is because a blank calibration test of an instrument consists in recording the *instrumental outputs delivered by this instrument* in the absence of a measured-object of the kind involved in the end-measurements. Yet an X-ray source does not deliver any directly recordable instrumental outputs. No instrumental output can be recorded with an X-ray source *alone*. As already stressed above (point 6), another instrumental device is required, here a measurer able to record an X-ray intensity. We therefore conclude that it is inappropriate to talk about a *blank* calibration *of the X-ray source*.

Yet, to describe the calibration tests under discussion as *blank* calibration tests *of the X-ray diffractometer* seems to miss an important aspect of these tests, namely that they intend to check properties of *one particular component* of the diffractometer, *the X-ray source*. If we want to emphasize this aspect, we can, *leaving aside the adjective "blank"*, use—as we did so far—the alternative expression "calibration tests *directed toward the X-ray source*". We prefer to avoid any formulation in terms of a calibration test *of* the X-ray source because, although the *intention* is indeed to test *the X-ray source*, *in concreto*, what is tested is never the source *alone*, but always the source *plus the detector* (we will come back to this point in the next section). This is

why we have used the (perhaps at-first-sight pointlessly complicated) expression "calibration test *directed toward* the X-ray source" rather than "calibration test *of* the X-ray source".

To recap, the calibration tests under scrutiny can be characterized either as calibration tests (neither blank, nor with etalons) directed toward the X-ray source, or as blank calibration tests of the X-ray diffractometer. Generalizing beyond the particular case of the X-ray source, it is meaningless to talk about a *blank* calibration *of a non-measuring device*.

5.6 Some General Epistemological Implications for the Calibration of a Treated-As-Composed Instrumental System

Let us draw some general lessons from the particular case of calibration in X-ray experiments at the IJL.

There are calibration tests which differ from the simple exemplar *in that their target T is not a measurer*. This has been illustrated by the case of the X-Ray source, but the point can be generalized. For example, it applies also to the optical devices involved in the diffractometer. In such cases, there is a *dissociation* between the *targeted* object T of the calibration test on the one hand (in our example: the X-ray source-token), and the device which provides the values of the mesuranda on the other hand (in our example: the X-ray detector-token which provides the values of the intensity delivered by the source).

This dissociation at the level of the calibration *tests* has repercussions at the level of the calibrating *operations* when the latter are *symbolic* operations. Indeed, the corresponding corrections (in our example: the application of a proportionality factor) cannot be applied directly to the instrumental outputs of the device which is the target of the calibration test (in our example: the X-ray source), since these outputs are not directly accessible but must be recorded by another device which plays the role of the measurer (in our example: the X-ray detector). The symbolic operations introduced in order to take into account the obtained/optimal discrepancy related to the non-measuring device under test (the decrease of the X-ray source intensity) are thus *inevitably* applied to the instrumental outputs of *another* device *than the target T of calibration*, that is, to a *measurer* (the X-ray detector).

This situation has several epistemological consequences. Indeed, the individual X-ray detector could have shifted itself from the optimal working corresponding to its model. But if the detector-token involved in the calibration test of the source-token is not properly calibrated, then, the calibration test of the optimal/obtained discrepancy of the intensity values of the X-ray source will not be reliable. With such considerations, we begin to meet the *holistic features* of the calibration procedure of a *complex* (i.e., treated as composed) instrumental device. What is actually tested, in the calibration test of a non-measuring device, is not just this non-measuring device-token at t_2 but the set "non-measuring device-token at t_2+ measurer-token at t_2". Thus in situations of this kind, we must distinguish the *intended* target and the *effective* target of the calibration test. The intended target T of calibration (in our example: the X-ray source-token at t_2) does not coincide with the effective target (in our example: the X-ray source-token at t_2 *plus the X-ray detector-token at t_2*). The calibration test of the effective target reduces to a test of the intended target *only if the detector is indeed itself well-calibrated*. So a prior calibration test of the detector-token seems to be required.

However, as a matter of facts, at the IJL, practitioners do not perform themselves the calibration tests and operations directed toward the detector-token. This part of the

diffractometer is, according to its model, much less subject to a significant drift over time than the source. So in routine practices, IJL experimenters do not perform any calibration tests that take the detector as their target. But beyond the particular case of experimental practices at the IJL, the situation just described points to a widespread interesting configuration with respect to calibration practices: something like a *co-calibration* of some instrumental modules involved in a treated-as-composed instrumental system. Let us further elaborate on this point.

Consider a complex instrumental system (say C as complex), treated-as-composed of multiple sub-modules, among which one finds at least one instrumental device which is *not* a measurer (say D as Device) and one instrumental device which is a measurer (say M as measurer). Suppose moreover that the instrumental outputs of C and the instrumental outputs of M coincide. Any calibration test of D inevitably requires, in addition to the non-measuring device under test, an M. This introduces a difference between the *intended* target of calibration (D) and the *effective* target of calibration ($D + M$). Consequently, all the conclusions about the calibration of D depend on whether or not M is itself properly calibrated (and *a fortiori* not defective). How can practitioners deal with such situations?

A common practice is to perform calibration tests of C with some etalon. Imagine that the instrumental outputs of C obtained in such calibration tests are interpretable in a way which coincide with the expected characteristics of the etalon. If this is the case, all the sub-modules involved in C (notably D and M) will be considered as correctly calibrated. In such a situation, we can consider that the measurer M serves as a means to test the adequate calibration of the non-measuring device D, and reciprocally, that D serves as a means to test the adequate calibration of M. Such a scheme can be characterized as a co-calibration.

Alternatively, imagine that the instrumental outputs obtained with C in the previous calibration tests are *not* the expected ones. In such a situation, it is often required to find the source of the problem. In the example of the X-ray diffractometer, one possibility would be to use several (say three) X-ray detectors and to test the intensity of the X-ray source with each of them at three moments close to one another. If the instrumental outputs of two of them coincide and the outputs of the third are very different, practitioners will suspect that it is the third detector, and not the source, which is not correctly calibrated. If the instrumental outputs of the three detectors coincide with and correspond to a value of the intensity which is inferior to the value in optimal working, practitioners will suspect that the source has drifted from its optimal working. And so on. In such a situation, we can talk of a co-calibration of the source and the three detectors.

This leads us back to the holistic features of the calibration procedure directed toward a *treated-as-composed* instrumental system. When dealing with a complex instrument C, practitioners have to perform a multiplicity and a long sequence of calibration tests, successively directed, from an intentional point of view, toward this or that sub-part S, S', etc., of the instrument. But some of the testing steps actually performed inevitably involve other elements than just the intentional target of the calibration test, and moreover, the various calibration steps must be performed one after the other. At each step of the whole calibration sequence focused on a particular instrumental module S taken as the intentional target T of the calibration test, practitioners *think as if* all the other modules S', S'', etc., coincided with the optimal working (or more generally with some other determined reference). This happens over and over again, from step to step, when permuting the instrumental module under test and the other assumed-as-optimal modules. A *ceteris paribus* condition must be presupposed all along the temporal process of the calibrating sequence: practitioners must assume that the obtained/optimal discrepancy related to S does not vary during the evaluation of the obtained/optimal discrepancy related to S', and

so on. Of course, this assumption is not a specificity of *calibration* procedures with respect to other scientific procedures. In particular, scientists must assume that the *ceteris paribus* condition holds, not just throughout the calibration sequence, but also subsequently, during the temporal interval which separates the end of the calibration sequence and the end of the end-measurements. But although the *ceteris paribus* assumption is not specific to *calibration* procedures, it has to be stressed that the longer a calibration sequence is, the more the *ceteris paribus* clause can fail to apply in fact.

6 Conclusions

6.1 Concrete Insights on Calibration

In this paper, we have restricted our attention to the situation which seemed at first sight one of the less problematic with respect to calibration, namely, calibration in UNSI practices. In other words, calibration has been analyzed from the standpoint of scientific practitioners who are *users* of *already well-designed* and *well-mastered* instruments and explore *relatively well-known* phenomena by means of such instruments.

In order to clarify the nature of this kind of practice and to grasp its internal logic, we have constructed a simple exemplar of calibration in UNSI practices, illustrated by the case of the equal-arm balance. The simple exemplar in UNSI practices is defined by the following features. The target T of calibration is a measurer-token. The calibration of the measurer-token involves two presuppositions about the measurer-token: (P1) the measurer-model (and thus *a fortiori* the measurer-type) is not problematic; and (P2) the measurer-token is not defective (no breakdown). The aim of a calibration procedure directed toward a measurer-token T under presuppositions P1 and P2 is to master the gap between the values obtained with the measurer-token and the values that should have been obtained in the optimal working defined by the specifications of the measurer-model. The procedure through which the aim is achieved typically involves two steps: first calibration *tests*, which can be blank tests and/or tests with etalons; second, if needed—depending on the result of the tests—calibration *operations*, which can be material and/or symbolic operations. In calibration tests with etalons, two presuppositions are endorsed about the etalon: the etalon is (P3) reliable and (P4) adequate, that is, sufficiently similar to the object under study in the end-measurements.

Next, we turned to a more complex case of calibration in UNSI practices, the calibration of an X-ray diffractometer in a nanoscience laboratory. We choose this example for two reasons. We aimed (i) to show the working power of our conceptual framework for the analysis of different, more complex cases; and (ii) to introduce more complexity into the picture (in agreement with our strategy to go from the simple to the complex) by characterizing *some* features of *more complex* calibration procedures in UNSI practices that are *not* captured by the *simple* exemplar. Two such features have been put forward, and some of their epistemological implications have been analyzed. The first feature is: the target T of calibration can be a treated-as-composed instrumental system (this is certainly the most common case in scientific practices of calibration). As a consequence, the calibration of T is a (more or less) long sequence of multiple calibration steps. This raises the question of the relations between the multiple calibration *steps* and the calibration of the measuring system *as a whole*. The second feature is: the target T of a calibration activity can be a *non-measuring* instrumental token. As a consequence, another instrumental token of the type "measurer" is inevitably involved in the calibration tests of T. Taking into account and

combining the implications of these two features, we have been led to direct the attention toward the *holistic character* of the calibration of a treated-as-composed measurer as a whole, to introduce the idea of *co-calibrations*, and to consider some epistemologically important aspects of co-calibrations, namely, the non coincidence of the *intended* and *effective* target of calibration, and the *ceteris paribus* condition that must be assumed all along a sequence of co-calibration.

The two last paragraphs sum up the main *substantial* results about calibration achieved in the present article. We see these substantial results as concrete insights gained so far about calibration, thanks to the conceptual framework we have elaborated. This conceptual framework itself is, of course, another result of the article, of a more 'formal' or 'structural' kind. As a conclusion, it is now time to come back to this framework and to revisit its main features.

6.2 Revisiting the Conceptual Framework

Now that the framework has been put to work, we are in a position to further specify its nature (what do we mean by a conceptual framework?), its status (what provides to a given characterization the status of a framework?), its scope (what is the domain of application of the different pieces of the framework?), and its value (what kind of work is accomplished by the framework?).

By a conceptual framework, we mean a set of more or less closely interrelated categories and questions, intended to work as conceptual tools and as guides for the analysis of a certain targeted subject matter (here, calibration). This implies that the corresponding categories help to understand and to classify multiple particular cases at first glance susceptible to be instances of the targeted subject matter. These categories must be both sufficiently general to apply to a multiplicity of particular cases (possibly with some adaptations), and sufficiently well-defined to drive efficiently and illuminate the analysis of newly encountered cases. The latter specifications provide answers to the two questions of the framework's *nature* and *expected value*.

The categories of the four-question frame (target, presuppositions, aim and procedure of calibration), the bold part of Tables 1, and 2,[15] can be considered as three conceptual frameworks in the previous sense—or, if one prefers, as three components of the same framework. The three components are of course not independent. The third one (the simple exemplar of calibration) relies on both the first one (since it uses the categories of the target, presuppositions, aim and procedure) and on the second one (since it uses the categories of the instrument-frame, such as the notion of measurer and the type/token distinction).

Further reflections about the relations between the three components enable us to specify what provides a given characterization the status of a framework, and by doing so, to offer a more precise explication of what is meant by a framework, as well as a more thorough explanation of the kind of work the proposed framework is expected to accomplish. What has been called "the framework of the simple exemplar of calibration" in this article (summed up in Table 2) uses what has been called the "four-question frame", and gives determinate answers to each question of the target, the presuppositions, the aim, and the procedure. To the extent that the simple exemplar of calibration gives

---

[15] We could add to this list the mapping of scientific practices sketched in Sect. 3.1.3. This mapping is indeed an integral important part of what we identify with our framework. However, in this paper, only very sketchy elements have been provided.

determinate answers to the four questions, we could be tempted to consider that the simple exemplar of calibration is not rightly characterized *as a framework* (in the common sense of a *form*, that is, of a deprived-of-content *shape* or *structure* intended to be applied to some content related to calibration), but should better be characterized as a *substantial description* of calibration. So is the simple exemplar of calibration a framework in the sense of a form (intended to be applied to some substance), or is the simple exemplar a content (a substantial characterization of calibration)? Our answer is: the frontier between a framework understood as a form, and a substantial characterization understood as a content, is relative and context-dependent; in particular, it depends on the purpose of the analyst.

Let us illustrate. Table 2 indeed provides a substantial (admittedly not exhaustive) characterization of calibration in UNSI practices *relatively to some instances* (the simple instances for which it appears that Table 2 directly applies without any need of accommodation). Relatively to these instances, the simple exemplar 'tells us something' about calibration; it is not just a means in order to access some knowledge about calibration. But *relatively to other*, *more complex instances* of calibration in UNSI practices—such as some calibration procedures in X-ray experiments conducted at the IJL—Table 2 plays the role of a framework in the sense of a form, and works as a guiding and enlightening benchmark, as we hope to have convinced the reader in part 5. The value of Table 2 lies in these two kinds of contributions taken together.

Let us be more precise and also give the recipe for the use of the framework as it stands. Faced with a new instance of activity that could be identified with a calibration (for example: X-rays experiments described in part 5), use Tables 1 and 2 as frames, and examine to what extent they directly apply or must be accommodated.

As far as Table 2 *applies* to the newly scrutinized case, Table 2 provides both a guiding frame and a substantial characterization. More exactly, it provides (i) a guiding frame (for the X-ray example, see Table 5, items in bold), which easily leads to a clear understanding of what is at stake—in the X-ray example, an understanding of most aspects of the calibration *of the X-ray diffractometer as a whole* (see Sect. 5.4), and an understanding of *some* aspects *of the calibration of the X-ray source* (see Table 5, items in italics). To the extent that Table 2 used as a frame drives this understanding, Table 2 offers—or at least works as a guide for the easy acquisition of—(ii) a substantial characterization of the newly encountered case.

As far as Table 2 *does not apply as such* to the newly scrutinized case, Table 2 nevertheless works as a benchmark, in reference and by contrast to which significant differences can be circumscribed and characterized (see Table 5, highlighted in underlined items).

6.3 Scope of the Framework

The scope of our framework still remains to be discussed, that is, the domain in which the framework is expected to make significant contributions. Actually, the scope differs according to the components of the framework.

a. **Scope of the four-question frame**. The intended scope, and the fecundity, of the four-question frame go much beyond the case of calibration in UNSI practices. The broad applicability is straightforward: obviously, the four questions can be applied to any activity candidate to calibration. The fecundity can only be claimed, but not showed, since in the present paper, only calibrations *in UNSI practices* have been discussed.

b. **Scope of the instrument-frame**. The bold part of Table 1 is applicable as such to any socially stabilized measurer, and is a clarifying tool for a fine-grained grasp of any measurer in UNSI practices (as we attempted to show that it was the case for *some* of them, in Sects. 3.3.6 and 5.4). In relation to *some* measurers, complements may be relevant. When the aim is to study the calibration of a *treated-as-composed* instrument, for example, it may be interesting to add some cells to Table 1, dedicated to the specification of the instrumental sub-modules which play a role in the calibration activity. In addition to be a fruitful tool for the analysis of already-standardized *measurers*, Table 1 is not deprived of interest when we turn to standardized *non-measuring* devices. In the latter case, Table 1 serves as a benchmark but needs some *accommodations*—of the kind introduced in Table 4, for the particular case of the X-ray source.

c. **Scope of the framework of the simple exemplar**. Table 2 has the status of an exemplar, which means (as indicated in Sect. 3.2.2) that Table 2 is *not* intended as an *exhaustive* characterization of any domain of calibration activities, nor as a set of *necessary and sufficient* conditions to categorize an activity as a calibration. Table 2 applies to calibration in UNSI practices in the sense that Table 2 is relevant and fruitful *as a tool* for the understanding of calibration in UNSI practices, even if all elements of Table 2 do not directly apply to the calibration instance under study. By our analyses of the calibration of a balance and of an X-ray diffractometer, we hope to have shown that at least in the domain of UNSI practices, the simple exemplar is a powerful analytic tool, which efficiently guides the achievement of a substantial characterization of the particular cases under scrutiny. We can add—but not argue— that the simple exemplar summed up in Table 2 is also very useful for the analysis of calibration in *other kinds* of scientific practices.

6.4 Relation of the Framework to Science in Practice

Eventually, some readers might perhaps wonder "how generating the conceptual framework relies upon the details of experimental practice" and "how the resulting conceptual framework differs from the product of armchair conceptual analysis".[16] These are no easy questions. We are convinced that the framework could not have been provided "by merely thinking about calibration", but it is not so straightforward to argue. True, we can repeat that we started with examples issued from contemporary scientific activities as practiced by our physicist colleague C. Dufour and as observed by our colleague C. Allamel-Raffin as an ethnographer of science (see Sect. 1.3). We can add that faced with difficulties, we discussed these with the relevant scientists involved in the activities of calibration under scrutiny. But this is of course not enough to argue convincingly that the anchorage in 'science as it is performed in practice' made a significant difference at the end of the day, regarding the conceptual framework we finally retained as the most adequate one. To give more convincing arguments, we would have to at least partially reconstitute the genetic "long and sinuous process" through which the final form of the framework has been elaborated (see Sect. 1.4); we would have to focus on hesitations, difficulties, blind spots, and to explain how we finally succeeded in surmounting them. However, such a task is very hard to achieve, not just because the volume of a paper is limited, but also for fundamental reasons. This is because when a new version of framework is indeed better

---

[16] As Sally Riordan, as reviewer of a first version of this article, relevantly wondered.

than the previous one, it leads to a restructuration of the understanding of the whole situation. From this results a neat relief at the moment, which also implies that in the next step, the anterior difficulties are erased and hence easily forgotten, so that they are increasingly difficult to reconstruct in retrospect. We are well-aware that the back and forth movement from practices to successive versions of the framework, and the corresponding important amount of work invested along the path, is difficult to perceive in the end-product. However, at present, what seems to be most important is the value of the end-product for further investigation on calibration; and we are confident that this framework can be valuable in this respect.

We have attempted to provide explicit indications about the fruitfulness of our simple exemplar of calibration, and in particular, to provide explicit insights about the way the simple exemplar is revealing when applied to newly encountered instances of calibration. However, no explicit formulation of this kind will ever be able to replace a first-person-experience. The best way to be convinced of the fecundity of the framework is to put it at work in practice. We encourage readers interested in the topic of calibration to use our framework to analyze the cases in which they are interested. More generally, we hope that the present work will inspire scholars to consider the significant but neglected topic of calibration.

References

Andersen, H. (2013). Epistemic dependence in contemporary science: Practices and malpractice. In L. Soler, et al. (Eds.), *Science after the practice turn in philosophy, history, and the social studies of science*. London: Routledge.

Boumans, M. (2004). The reliability of an instrument. *Social Epistemology, 18*(2–3), 215–246.

Boumans, M. (2006). The difference between answering a 'Why' question and answering a 'How Much' question. In J. Kueppers, J. Lehnard, & T. Shinn (Eds.), *Simulation: Pragmatic construction of reality* (pp. 107–124). Dordrecht: Springer.

Boumans, M. (2007). Invariance and calibration. In M. Boumans (Ed.), *Measurement in economics* (pp. 231–248). London: Academic Press.

Bourget, M.-N., Licoppe, C., & Sibum, H. O. (2002). *Instruments, travel and science: Itineraries of precision from the seventeenth to the twentieth century*. London: Routledge.

Buccianti, R., Cibien, M., Mari, L., & Rebaglia, B. I. (2009). Measurement and calibration: Considerations based on the international vocabulary of metrology (VIM, 3rd Ed.) and related standards, *XIX IMEKO World Congress, Fundamental and Applied Metrology*, September 6–11, 2009, Lisbon, Portugal.

Carnap, R. (1995 [1966]). *An introduction to the philosophy of science*. New York: Dover Publications.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. New York: Oxford University Press.

Chang, H. (2013). Epistemic activities and systems of practice: Units of analysis in philosophy of science after the practice turn. In L. Soler, et al. (Eds.), *Science after the practice turn in philosophy, history, and the social studies of science*. London: Routledge.

Chang, H., & Yi, S. W. (2005). The absolute and its measurement: William Thomson on temperature. *Annals of Science, 62*, 281–308.

Collins, H. M. (1992 [1985]). *Changing order: Replication and induction in scientific practice*. Chicago: University of Chicago Press.

Collins, H. M. (2004). *Gravity's shadow*. Chicago: University of Chicago Press.

Franklin, A. (1986). *The neglect of experiment*. Cambridge: Cambridge University Press.

Franklin, A. (1990). *Experiment: Right or wrong*. Cambridge: Cambridge University Press.

Franklin, A. (1997). Calibration. *Perspectives on Science, 5*(1), 31–80.

Harris, T. (2003). Data models and the acquisition and manipulation of data. *Philosophy of Science, 70*(5), 1508–1517.

Heering, P. (2005). Weighing the heat: The replication of the experiments with the ice-calorimeter of lavoisier and laplace. In M. Beretta (Ed.), *Lavoisier in perspective* (pp. 27–41). München: Deutsches Museum.

Licoppe, C. (1996). *La formation de la pratique scientifique. Le discours de l'expérience en France et en Angleterre (1630–1820)*. Paris: La Découverte.

Livengood, J. (2009). Why was M. S. Tswett's chromatographic adsorption analysis rejected? *Studies in History and Philosophy of Science, 40*(1), 57–69.

Mallard, A. (1998). Compare, standardize and settle agreement: On some usual metrological problems. *Social Studies of Science, 28*(4), 571–601.

Mari, L., & Sartori, S. (2007). A relational theory of measurement: Traceability as a solution to the non-transitivity of measurement results. *Measurement, 40*, 233–242.

O'Connell, J. (1993). Metrology: The creation of universality by the circulation of particulars. *Social Studies of Science, 23*(1), 129–173.

Ramsey, J. (2007). Calibrating and constructing models of protein folding. *Synthese, 155*, 307–320.

Rasmussen, N. (1993). Facts, artifacts, and mesosomes: Practicing epistemology with the electron microscope. *Studies in History and Philosophy of Science (Part A), 24*(2), 227–265.

Rossi, G., Crenna, F., & Panero, M. (2005). Panel or jury testing methods in a metrological perspective. *Metrologia, 42*, 97–109.

Schickore, J. (2007). *The microscope and the eye*. Chicago: Chicago University Press.

Sibum, H. O. (1995). Reworking the mechanical value of heat: Instruments of precision and gestures of accuracy in early Victorian England. *Studies in History and Philosophy of Science, 26*(1), 73–106.

Skipper, R. A. (2004). Calibration of laboratory models in population genetics. *Perspectives on Science, 12*(4), 369–393.

Tal, E. (2011). How accurate is the standard second? *Philosophy of Science, 78*(5), 1082–1096.

VIM. (2008). *International vocabulary of metrology—basic and general concepts and associated terms (VIM)*, Joint Committee for Guides in Metrology.

Werker, C., & Brenner, T. (2004). Empirical calibration of simulation models. In *Papers on economics and evolution,* #0410, Jena: Max Planck Institute for Research into Economic Systems.

Woodward, J. (1989). Data and phenomena. *Synthese, 79*, 393–472.