Studies in History and Philosophy of Science xxx (2017) 1–13

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

Calibration: Modelling the measurement process

Eran Tal

McGill University, Canada

A R T I C L E I N F O

Article history:
Received 4 April 2016
Available online xxx

Keywords:
Measurement
Calibration
Models
Prediction
Uncertainty
Metrology

A B S T R A C T

Calibration procedures establish a reliable relation between the final states ('indications') of a measurement process and features of the objects being measured ('outcomes'). This article analyzes the inferential structure of calibration procedures. I show that calibration is a modelling activity, namely the activity of constructing, deriving predictions from, and testing theoretical and statistical models of a measurement process. Measurement outcomes are parameter value ranges that maximize the predictive accuracy and mutual coherence of such models, among other desiderata. This model-based view of calibration clarifies the source of objectivity of measurement outcomes, the nature of measurement accuracy, and the close relationship between measurement and prediction. Contrary to commonly held views, I argue that measurement standards are not necessary for calibration, although they are useful in maintaining coherence across large networks of measurement procedures.

1. Introduction

The reliability of commonplace measuring instruments, such as clocks, thermometers and balances, is easily taken for granted. The facility with which one operates and reads the indications of such instruments suggests an unproblematic and almost direct contact with the object or event being measured. The study of metrology – the science of measurement – reveals a different picture. Establishing a reliable relation between the indications of an instrument and relevant features of an object turns out to be a complex endeavor even for seemingly simple instruments. This endeavor is known as calibration. In its full generality, calibration consists of a variety of activities that are mostly carried out behind the scenes, e.g. by instrument designers, instrument manufacturers, accredited laboratories that provide calibration services to instrument manufacturers, and national and international bureaus of standards that coordinate the activities of accredited laboratories. By the time a typical measuring instrument reaches its end-user, its calibration is complete, and no more than minor tuning is required.

What is calibration, and how does it establish the reliability of measuring instruments? In what follows I will offer an epistemological analysis of metrological calibration practices. Despite the wide variety of tasks that fall under the label 'calibration', I will show that calibration activities have a common inferential structure and shared epistemic goals. Specifically, I will argue that calibration is a special kind of modelling activity where the system being modelled is a measurement process. The primary aim of calibration is to identify which parameter values in the model coherently and accurately predict the final states of the measurement process. By 'model' I mean an abstract and approximate representation of a local phenomenon, a representation that is used to predict (and sometimes also explain) aspects of that phenomenon.[1] By 'modelling' I refer not only to the act of constructing models, but also to the iterative process of acquiring background knowledge, extracting predictions from a model, testing those predictions empirically, and modifying both the model and the concrete system it represents to achieve a better fit.

The term 'calibration' is used in a variety of ways across different contexts and disciplines. Some of these uses bear little more than a loose resemblance to the metrological use. In Bayesian statistics, 'calibration' sometimes refers to the fit between the predicted probabilities of events and the observed success rate of those predictions (Seidenfeld, 1985, p. 275). In climate simulation, 'calibration' means tuning the values of free parameters in the computational model to fit known data (Steele & Werndl, 2013, p. 610). It is common parlance to refer to the adjustment of the blank ('zero') indication of a household scale as 'calibration'.[2] None of

E-mail address: eran.tal@mcgill.ca.

[1] Models are constructed from theoretical, statistical or other assumptions, but function autonomously from the background theories that informed their construction. This conception of models follows closely the views expressed in Morrison and Morgan (1999), Morrison (1999), and Cartwright (1999).

[2] From the viewpoint of the VIM, this terminological confusion is to be avoided: "Adjustment of a measuring system should not be confused with calibration, which is a prerequisite for adjustment" (JCGM, 2008a, 3.11, Note 2). Calibration operations establish a relation between indications and outcomes, and this relation may later be expressed in a simpler manner by adjusting the display of the instrument.

https://doi.org/10.1016/j.shpsa.2017.09.001
0039-3681/

these uses of the term coincide with the meaning of 'calibration' discussed here, although they share with it some broad features.[3]

Other uses of the term 'calibration' partially overlap with the metrological one. For example, the term is sometimes used synonymously with 'scaling' or 'gradation'. In this sense, calibration is the empirical activity of detecting correlations among the indications of instruments, or between the indications of an instrument and a set of reference systems that are associated with fixed values. Values are then assigned to the indications of the instrument being calibrated so as to match previously known values, often along with a rule for extrapolating between (and beyond) those known values. Hasok Chang uses 'calibration' in this sense when he describes procedures for making different kinds of thermometers produce similar temperature values (2004, pp. 59, 79, 115, 127–9). These procedures of scaling and gradation fall under the concept of 'black-box' calibration discussed below. As I will show, black-box calibration is only the simplest form of calibration and is not representative of the general category. Moreover, even this simple form of calibration already involves the abstract and idealized modelling of the measuring instrument, the reference systems, and the circumstances of the comparison.

Another context in which 'calibration' is used in a roughly metrological sense is laboratory experimentation. In their analyses of experimental physics, Harry Collins and Allan Franklin define calibration as "the use of a surrogate signal to standardize an instrument" (Collins, 1992, p. 105; Franklin, 1997, p. 31; See also Franklin, 1986, pp. 175–180). A surrogate signal is a well-understood system that resembles in relevant respects the system one would like to detect or learn about. For example, an electrostatic signal may serve as a surrogate for a gravitational wave. If the experimental apparatus correctly detects or measures relevant properties of the surrogate signal, this is taken as evidence for the ability of the apparatus to detect or measure the target phenomenon. Provided that the notions of surrogate signal and standardization are interpreted broadly enough, the Collins-Franklin definition covers many cases of metrological calibration. At the same time, Collins and Franklin did not attempt to provide a full-fledged epistemology of calibration. Their accounts analyzed the grounds for making detection claims in novel experimental situations where 'ordinary' measuring instruments can already be assumed to be calibrated. Topics such as the nature of measurement accuracy, the functions of measurement standards, and the role of coherence in calibration fell beyond the purview of their accounts.[4]

Recent accounts of calibration in the philosophical literature have been thoroughly informed by contemporary metrological discussions (Boumans, 2007, 2012; Frigerio, Giordani, & Mari, 2010, pp. 143–4; Mari & Giordani, 2014; Soler et al., 2013; Soler, 2015). This article will build on these works and extend their scope, with the aim of providing a general outline of an epistemology of calibration. My point is not terminological. I will not argue for any particular use of the term 'calibration', but develop an account of the ways in which the reliability of measuring instruments is established. As I will show, understanding calibration as a modelling activity whose target is a measurement process leads to helpful insights for the wider field of epistemology of measurement. Measurement outcomes turn out to be model-based predictors of instrument indications, and measurement accuracy a special kind of predictive accuracy. The objectivity of measurement outcomes turns out to be grounded on considerations of coherence among models of multiple measurement processes, and measurement standards turn out to be highly useful, but not exclusive, tools for securing such coherence. The increasing use of numerical methods and computer simulations as part of calibration is also easily accommodated by the model-based view.

I will begin by introducing preliminary concepts and discussing the official metrological definition of calibration (Section 2). Next, I will explore three ways of modelling a measurement process. These are not mutually exclusive, but progressively more complex and general modes of representation, each containing the previous ones as limiting cases. The simplest, 'black-box' calibration, neglects most the complexities of the measurement process and the measurement standard (Section 3). 'White-box' calibration represents the measurement process in detail, but neglects the complexities of standards (Section 4). Finally, coherent calibration represents a measurement process as part of a larger network of intercomparisons among instruments, some of which may be measurement standards (Section 5). Section 6 will situate the model-based view with respect to recent accounts of calibration, and Section 7 will conclude the discussion.

2. Preliminary concepts

2.1. *Instrument indications vs. measurement outcomes*

The first step in elucidating the inferential structure of calibration is to distinguish between instrument indications and measurement outcomes.[5] An indication (or 'reading') is a property of a measuring instrument in its final state after the measurement process is complete. Examples of indications are the numerals appearing on the display of a digital clock, the position of an ammeter pointer relative to a dial, and the pattern of diffraction produced in x-ray crystallography. Note that the term 'indication' in the context of the current discussion carries no normative connotation. It does not presuppose reliability or success in indicating anything, but only an *intention* to use such outputs for reliable indication of some property of the object or event being measured. Note also that indications are not numbers: they may be symbols, visual patterns, acoustic signals, relative spatial or temporal positions, or any other sort of instrument output. However, indications are often represented by mapping them onto numbers, e.g. the number of 'ticks' the clock generated at a given period, the angular displacement of the pointer relative to the ammeter dial, or the spatial density of diffraction fringes. These numbers, which may be called 'quantified indications', are convenient representations of indications in mathematical form.[6] A quantified indication is not yet an estimate of any property of the object being measured, but only a mathematical description of a state of the measuring apparatus.

A measurement outcome (or 'measurement result') is a knowledge claim attributing one or more parameter values to the object or event being measured, a claim that is inferred from one or more instrument indications along with relevant background knowledge. Although measured parameters need not be quantitative,[7] this article will deal solely with quantitative parameters, or

[3] All three uses just mentioned refer to a procedure that involves testing the compatibility between one or more parameter values and empirical data. Indeed, under specific circumstances, the term 'calibration' in all three uses just mentioned may refer to one or another *step* in a full metrological calibration.

[4] Franklin's historical examples shed some valuable light on these issues, but he does not attempt to offer an overarching account of the inferential structure of calibration, nor does he engage with the metrological literature on this topic.

[5] This subsection rehearses the discussion found in Tal (2017), pp. 235–6.

[6] The difference between numbers and numerals is important here. Before it is quantified, an indication is not a number, though it may be a numeral (i.e. a symbol representing a number).

[7] For example, parameters like shape and chemical formula are used to classify objects on a nominal scale.