Models: Measuring or Cognitive Instruments?*

Florian J. Boge

Institute for Philosophy and Political Science (IfPP), TU Dortmund University
& Lamarr Institute for Machine Learning and Artificial Intelligence

Abstract: A number of authors (Morgan, 1999; Boumans, 2005; Morrison, 2009; Massimi and Bhimji, 2015; Parker, 2017) have argued that models can be quite literally thought of as measuring instruments. I here challenge this view by reconstructing three arguments from the literature and rebutting them. Further, I argue that models should be seen as cognitive rather than measuring instruments, and that the distinction is important for understanding scientific change: Both yield two distinct sources of insight that mutually depend on each other, and should not be equated. In particular, we may perform the exact same actions in the laboratory but conceive of them entirely differently by virtue of the models we endorse at different points in time.

1 Introduction

Models are a vital part of scientific inquiry, and they are in good use, among other things, in the evaluation of physical measurements. However, there is a view which additionally holds that models, especially those implemented as computer simulations,

> [n]ot only [...] allow us to interpret so-called measurement outputs, but [...] that the models themselves can *function as measuring instruments* [...]. (Morrison, 2009, 35; my emphasis)

Should we really think of models as measuring instruments, and if so, why exactly? Similar views have been advanced by Morrison and Morgan (1999), Morgan (1999), Boumans (2005)

*The version of record of this article will be published in *Journal for General Philosophy of Science*. A link/doi will be provided when available.

Massimi and Bhimji (2015), andParker (2017), but it is not always straightforward to discern the actual *arguments* in their favor.

The present paper has two parts. I will first reconstruct three arguments to the effect that models function as (or 'are') measuring devices from the more recent literature and rebut them in turn (Sect. 2–4). In Sect. 5, I will then, in the second part, turn to an alternative, according to which models are *cognitive* rather than measuring instruments, and argue that this helps us understand scientific practice, and especially scientific change, much better than if we lump models and measuring instruments together.

2 The argument from indispensability

2.1 Argument: Models are indispensable for measurements

Probably the first argument for models being measuring instruments has been offered by Morrison (2009), who distinguishes a "rather straightforward" (35) use of models as measuring instruments, when these are physical models (or model-*systems*), from a less straightforward one, when they are mathematical. Although the former case is interesting its own right (e.g. Currie, 2020; Dardashti et al., 2017), I here focus on the latter case.

To illustrate her point, Morrison (2009) considers a physical pendulum for measuring the gravitational acceleration, g, as an example (also Morrison, 1999). Recall that, if we use Newtons second law and a small angle approximation, we retrieve the relation $g = 4\pi^2\ell/T_0^2$, where T_0 is the (fixed) period over which the pendulum oscillates back and forth and ℓ the chord length. However, various simplifications, approximations, and idealizations are necessary to arrive at this result. To retrieve any realistic measurement of g, one will have to introduce all kinds of further modeling steps that correct for the numerical artifacts thus induced (Morrison, 2009, 49). It is thus impossible to measure g without these models, as much as it is impossible without the chord and bob. Hence:

> using the physical pendulum as an experimental apparatus requires us to have in place a very sophisticated model of that apparatus that *renders it a measuring instrument.* (ibid.; emph. added)

The italicized phrase is ambiguous, for what does 'it' refer to? It could mean that having the sophisticated model in hand renders the apparatus a measuring instrument. But given that Morrison is interested in models being measuring instruments, I take 'it' to mean the model here. Thus, Morrison effectively offers the following argument, which I call the *argument from indispensability*:

P1_i Due to the necessity of idealization, simplification, and approximation in applying theoretical formulae in physical measurements, models are indispensable for getting realistic measurement outcomes.

P2_i If something is indispensable for getting realistic measurement outcomes, it is (or 'functions as') a measuring instrument.

C_i Hence, models are (or function as) measuring instruments.

This argument is fairly straightforward, and we could take it as the sole target of the paper. However, as I shall demonstrate, ideas advocated by Parker (2017) and Boge (2021b) give rise to further arguments to be addressed independently.

2.2 Rebuttal: Indispensability does not imply identity

What is the plausibility of the argument from indispensability? I believe that this argument can be rebutted, when a distinction between three types of measurement by Parker (2017, 280–3) is taken into account:

- 'direct measurement', which means using an instrument indication, without any "explicit symbolic calculation", to "assign[...] a preliminary value to the parameter that is ultimately of interest" (280);

- 'derived measurement', in which "measurement outcomes are calculated or derived from [...] directly measured values for other parameters, using reliable scientific principles or definitions" (281); and

- 'complex measurement', "which involves making multiple direct and/or derived measurements and then using their results together, with the aim of obtaining a measurement outcome that is more informative than could be achieved using just a subset of those results" (283).

In the pendulum case, a 'direct measurement' would be one of ℓ with a measuring rod: We normally take the scale on the ruler to be well-calibrated and the ambient conditions of the measurement to be sufficiently non-distorting, so that the little bars on the ruler are approximately equidistant and similar to a reference-length for a comparison between chord and ruler to be informative. Every measurement involves some stage that should count as 'direct' in this sense: There always is a step wherein the numerical readout of a physical device, or some interface that is connected to further devices, is used to deliver a – usually preliminary – value for a quantity of interest. However, for reasons of measurement-uncertainty, one will normally take several such direct measurements – of ℓ, say – and then suitably average over these, making pretty much all scientific measurements 'derived'. Furthermore, due to the need to rely on further models and corrections, the measurement of g, based on the former two measurements, will be 'complex', as is true of many (if not most) scientific measurements.

Notice that the derived nature of measurement holds if by 'averaging', we mean nothing more than computing the arithmetic mean: if $\bar{m}$ is the reported value of the measurement, and m_i are individual, direct measurements, then $\bar{m}$ is derived from the m_i by means of the prescription $\bar{m} = \frac{1}{N}\sum_{i=1}^{N} m_i$. However, there are of course numerous ways (e.g. MacMahon et al., 2004) of averaging individual measurements, m_i, which are often in an even more obvious sense ways to derive or calculate a measurement value "using reliable scientific principles or definitions".

Furthermore, most quantities scientists are interested in are even harder to get at, and there is no way to ascertain their values just by suitably averaging over relevant direct measurements. For instance, there simply can be no top-quark mass-ometer, as top quarks decay within some 10^{-25} seconds. Just as much may there never be a reliable attitude-ometer in psychology, as external indicators are insufficient for determining psychological states. The unreliability of lie detectors serves as a case in point. In other words: many scientific measurements are necessarily *complex* in Parker's sense.[1]

Thus, some amount of modeling is usually indispensable for scientific measurement. But at the same time, this undermines the very basis for thinking of models as literal measuring instruments: Given that already derived measurement requires symbolic manipulations, scientific measurements, in general, include both, physical activities *and* inferential steps. Hence, not every action within a measurement procedure needs to count as a measurement activity properly so called. Why not, in other words, simply view the models used in the context of measurements as fully residing on the *inference*-side of activities contained within scientific measurements? $\mathrm{P2}_i$, it seems, rests on shaky grounds.

One might question the distinction I have drawn between inferences and physical activities. For example, when a charged particle passes through a drift chamber, it ionizes gas molecules within the chamber, the ionized particles drift to the chamber wires, producing a signal that travels down the wire to an analog-to-digital converter. That digital signal then travels through a series of transformations in a computer before it becomes part of a readout accessed by any physicist. By that time it is just one point relied on in the reconstruction of a track. But

[1] Skimming through *Nature*-papers from the past 20 years further supports the point: Ley et al. (2006) measured a correlation between weight-loss through dietary changes and increases in certain bacteria in the colon. For that, they had to extract and measure bacterial RNA in a complicated procedure from human feces, which involved a bio-informatics tool for sequencing, and at the same time measure the weight loss of test subjects. Gross et al. (2011) measured matter-wave quadratures within a Bose-Einstein condensate through an analogue to homodyne detection in quantum optics, which requires models of both measuring devices and the condensate. Among other things, comparison to a reference measurement was necessary to ensure that the data were produced by a targeted sort of collision within the condensate. Wright and Jackson (2022) measured the predictivity of childhood temperament for later-in-life personality. For that, they had to individually score and then integrate temperament and personality by means of different questionnaires, which requires models of 'personality' and 'temperament'. And so on.

where does the physical instrument end and the inference begin? Is it at the point that the analog signal reaches the analog-to-digital converter? Why there and not further upstream or downstream?

First, it is important to note that I have not claimed that the measurement ends somewhere and that the inference begins there: Measurement involves both inferences and physical activities. That does not make the tools used for inference measuring-devices. And I see no problem in sorting the cognitive, inferential activities made by appealing to, and in preparation of, a computer simulation run in the course of a measurement from the physical actions performed in the course of that same measurement.

To see the difference, we should ask for the *purpose* of programming the computer to implement a given model. Is the purpose to interpret the computer's behavior as *directly indicative* of the measured system's behavior, as in the manual calibration of an instrument? Or is the purpose to *infer* something about the system's behavior, based on the *modeling assumptions*? I think the answer is clear, and I suggest to only count activities on such devices directly indicative of the measured system's behavior 'measurement activities properly so called'.

Of course, instrument-calibration[2] also involves modeling assumptions, but these are modeling assumptions as to how the measuring device will react to the measured system in the ambient conditions (Tal, 2012). In contrast, the model informing the computer in the scenario discussed above only concerns the behavior of the targeted system. Any considerations connected to the computer's behavior in devising the program will relate to its ability to accurately inform us about the *consequences of the model*, not its immediate reaction to the measured system. I believe this makes all the difference in the world.

However, instead the objection might be that the physical activity carried out by the computer performing the transformation is *itself* already part of the inferences drawn, thus blurring the distinction in a different way.[3] That would be a mistake: For who should we say is drawing the

[2] I shall have to say more about this point later on.

[3] I owe the above example and this objection to an anonymous referee.

inferences here? The computer? Unless we believe that digital computers running simulations are thinking beings – which I do not –, such a claim would be a mere metaphor. The computer mentioned in the above example may, in other words, be an instrument, but it is neither a measuring instrument, nor something that draws inferences: If anything, it is an instrument *supporting* certain inferences by *human researchers*.

3 The argument from physical involvement

3.1 Argument: Computer simulations are directly involved in measurement-procedures

We have already leaped ahead a bit, for like Morrison later on in the paper, Parker is specifically interested in computer simulations (CSs) and whether they "can be embedded in measurement practices in such a way that simulation results constitute measurement outcomes" (Parker, 2017, 274). I take it that, if the output of something 'constitutes a measurement outcome', then that something can be considered a measuring device.

The arguments Parker offers are different from Morrison's though. First off, Parker (2017, 280) considers a 'measurement outcome'

> a selective representation of the system under measurement, inferred from one or more instrument indications; it consists of values for a single parameter (a best estimate plus uncertainty range) or, in some cases, values for multiple parameters (each with a best estimate plus uncertainty range) that together can represent system states, trajectories, fields, flows, and so on.

Here, an 'instrument indication' is "the physical state of an apparatus used in measuring, such as a pointer position or a digital display showing a numerical readout" (ibid.). However, as explained above, there are several ways in which one can move from an instrument indication to a selective representation of a system under study: Some measurements are direct, some derived, some complex.

A surprising claim by Parker is now that values retrieved via models can not only count as derived or complex measurement results, but that

> it is possible for computer simulations to be embedded in [...] measurement practices in such a way that simulation results constitute raw instrument readings or even measurement outcomes. (Parker, 2017, 285)

This is an even stronger claim than Morrison's, for the result of the CS is here construed as the *indication of an instrument*, so it may even count as the result of a *direct* measurement.

What is the argument in support of this? To establish this claim, Parker refers back to an example first considered by Van Fraassen (2008):

> Suppose we are interested in measuring the temperature of a very small cup of hot tea at time t_0, and we insert a mercury thermometer at that time; we wait a short while for the mercury to stop rising in the tube and take a reading. But thermodynamic theory tells us that the thermometer itself will affect the temperature of the tea and hence the reading obtained. To arrive at a more accurate temperature estimate for t_0, our measurement process will need to include a step that corrects the thermometer reading for this interference. (Parker, 2017, 285)

This can be done using some calculations; however, it can also be done using the output of a CS, and so, "[i]n those cases, simulation results can be direct measurement outcomes" (ibid.).

Let us unpack this claim a little. Recall that an instrument indication is "the physical state of an apparatus used in measuring, [...] showing a numerical readout". Thus, the computer involved in the CS may satisfy this definition, and its display may count as an 'instrument indication' in Parker's sense. Furthermore, since there will be no further "explicit symbolic calculation" involved, this process of reading off the corrected temperature may count as a 'direct measurement' in her sense.

Thus, Parker seems to endorse the following argument:

$P1_p$ State X of physical apparatus Y is an instrument indication *iff* (if and only if) X is interpretable as a (numerical) readout and Y is used in the process of measuring.

$P2_p$ Process P is a direct measurement *iff* during P, an instrument indication is used to assign a preliminary value to a parameter of interest.

$P3_p$ When a CS is used to immediately correct for the error of a given apparatus, the physical state of the computer on which the CS is run is interpretable as a numerical readout used in the measuring process.

C_p Therefore, the result of the CS is an instrument indication and the process is a direct measurement.

Call this the *argument from physical involvement*. As stated, this is an argument for the CS-result being an instrument indication and for the process being a direct measurement, not for the CS or the model underlying it being a measuring instrument. However, it immediately supports the following *subsidiary argument*:

$P1_s$ If the state of a physical apparatus Y is an instrument indication in direct measurement process P, then Y can function as a measuring instrument.

$P2_s$ According to the argument from physical involvement, the results of some CSs are instrument indications in some Ps.

C_s Therefore, some CSs can function as measuring instruments.

It should be noted that Parker (2017, 285) qualifies her agreement with Morrison (2009) as follows: "computer simulation models can function as measuring instruments [...] only in the sense that a computational instrument/apparatus (such as a [...] programmed computer) can be part of a measurement process." Thus, $P1_s$ should be given such a modest reading. However, since the argument from physical involvement was reconstructed entirely from premises Parker endorses, and since it feeds directly into the subsidiary argument, it is unclear how much really separates Parker's views from Morrison's.

3.2 Rebuttal I: Not every apparatus used in measurement is a measuring instrument

The line of reasoning supported by the argument from physical involvement and the subsidiary argument is a little more subtle, as it involves simulations implemented on a physical computer, but I believe there are at least two reasons to reject it.

The first concerns $\mathrm{P1}_p$ and the attached notion of a measuring instrument that it supports. Recall that Parker defines an instrument indication as "the physical state of an apparatus used in measuring, such as a pointer position or a digital display showing a numerical readout". I had based $\mathrm{P1}_p$ on this definition. As shown above, this directly yields the relevant consequence (C_s), when conjoined with other claims Parker endorses. Yet, it remains unclear whether the CS, or even the physical computer running it, satisfies *all* conditions something needs to satisfy to count as not just as 'an apparatus used in measurement' but a *measuring instrument*.

Following the thread of Sect. 2.2, I suggest that only *some* physical devices used in actual measurements should be thought of as delivering *instrument indications*. Concretely, I suggest that this should be devices *directly* involved in *direct* measurement: The ruler in the measurement of the rod is a primitive example; a Geiger counter measuring radiation a less primitive one; and a particle detector measuring debris from scattering particles a fairly involved one.[4] What differentiates these from a computer running a simulation in order to correct some expected error though? Answering this requires a more careful analysis of 'measuring instrument'.

[4] One could define the notion differently, but for the purposes of what I am arguing for, I presuppose Parker's definition, with the added specification given above. Said specification has precisely the implication that, in the example on p. 5, only the untransformed digital signal may count as an instrument indication; the computer-transformed signal, based on algorithms that in turn are based on physics background-knowledge, involves an inferential step: That the 'correct' signal is the one as delivered by the model-based transformation. It should hence not count as an "instrument indication".

For starters, Heidelberger (2003, 147) suggests that instruments can play a *representative* role:

> In a thermometer, for example, the different states of heat accessible to our sense of heat are transformed into different states of the instrument itself [...] that are accessible to sight. [...] The changes the instrument undergoes can be taken as representative of the changes of the measured phenomena.

Thus, a measuring instrument suitably *correlates* with the system under study in such a way that the correlation can be exploited to retrieve the desired information about the measured system. However, I submit that the correlation between measured system and measuring device can only be so exploited when it is appropriately *causal*.

A little thought experiment may elicit the right intuitions. Imagine a freak physicist who knows the states a given measured system will take on in advance, due to clairvoyance. She then adjusts a certain device's readout synchronously with the state changes so that the readout correlates perfectly with the system's state. This device would, nevertheless, make for a poor measuring device, as the correlation would not be brought about in the right ways. Hence, changes in the device which correlate with changes in the system under study must be brought about by the system causing the device to change, in order for the device to fulfill a representative role.

Giere (2009) has actually argued, against Morrison (2009), that it is exactly this causal connection which distinguishes CS epistemically from experiment:

> The epistemological payoff of a traditional experiment, because of the causal connection with the target system, is greater (or less) confidence in the fit between a model and a target system. A computer experiment, which does not go beyond the simulation system, has no such payoff. (Giere, 2009, 61)

Against Giere, Massimi and Bhimji (2015, 74) have objected that on three possible readings of 'causal contact', CSs can be relevantly in contact with the target system; namely, as either

(i) calibration of quantities by direct contact, (ii) the tracking of interactions between variables by quasi-direct contact, or (iii) the inference to an entity's existence (say, a novel particle) by means of indirect contact, when the general experimental context is taken into account.

I will turn to issues of calibration below, as this is directly connected to a third argument for models being measuring instruments. For now note, however, that is unclear that (i)–(iii) are all possible readings of 'causal contact'. For instance, Boge (2019, 11–12) points out that none of these readings

> even makes contact with the real bite of Giere's argument. [...] Any unexpected 'recalcitrance' occurring in a CS must [...] either be attributed to unrecognized (maybe unintended) elements of the models used, or to unrecognized properties of the computer itself (failures of the hardware or the operating system, limitations of the programming language...). It cannot be contributed by the target system.

Thus, the problem with Massimi and Bhimji's claims is that the causal contact in question is insufficiently direct, and not suitably attached to properties of the relevant device, for the CS to count as a measuring instrument: In order for a device to count as a measuring instrument, changes in the instrument – especially highly *unexpected* changes – should be recognizable as being brought about by properties of the system measured. In a CS, this is hardly possible, due to the fact that the CS realizes a theoretical model (or a model derivative therefrom) on a digital computer: Any unexpected change will be attributable to either errors in implementation (or maybe even compilation) of the model, or to errors of hardware. The connection between measured system and simulation model is simply too remote and too highly mediated (also Boge, 2020) for the model to count as a proper measuring instrument.

If the above is correct, however, we immediately also see that a part of $P1_p$ from the argument from physical involvement rests on shaky grounds: The mere use of the computer on which the simulation is run in the measurement is insufficient to establish the simulation's status as a measuring instrument.

3.3 Rebuttal II: Computer displays are not numerical readouts

I claimed above that there are two main lines of attack against the argument from physical involvement, so consider P3_p now, the claim that the computer display can be interpreted as a numerical readout in the relevant sense. There is a lengthy debate about the status of CSs; in particular, whether they are merely automatically executed arguments, or could in some sense count as experiments (Morrison, 2009; Parker, 2009; Massimi and Bhimji, 2015; Beisbart, 2012; Beisbart and Norton, 2012; Boge, 2019, 2020).

The second kind of view actually subsumes two distinct views. On one of these, CSs obtain an experimental status exactly in the sense of the view at stake here (Morrison, 2009; Massimi and Bhimji, 2015); so presupposing this in order justify a premise that is then used to justify the view itself would be a *petitio*.

This leaves the other two views: That (I) CSs are just arguments executed by computers (Beisbart, 2012; Beisbart and Norton, 2012), or (II) that CSs are experiments literally performed on the computer (Boge, 2019; Parker, 2009), whence the 'simulation' in 'computer simulation' obtains the same basic meaning as in other simulation studies (also Dardashti et al., 2017). However, on option (I), the output of a CS is a conclusion, not a numerical readout. Hence, on this view, P3_p is false.

On option (II), however, the ability to infer something about the system *of interest* from an experiment on the computer depends on the ability to map the computer's final state back to the system of interest, via a chain of models and approximate mappings (Boge, 2020): Usually, theoretical models implemented in a CS need to be discretized, they then need to be translated into code, complied into a language that instructs the physical computer how to behave, and then actually run on said computer. In all these stages, information may be lost or distorted, due to lumping effects or partial failures of isomorphism between, say, state transformations within a discretized, mathematical model and a piece of computer code (Boge, 2020, 27–29), or even due the need to include physically unrealistic assumptions to compensate, e.g., truncation or floating-point errors (Lenhard, 2007, for a pertinent example). Due to such factors, one might end up representing "two non-trivially different target systems depending on the chosen programming language, code execution, and the like." (Durán, 2024, 157)

However, the above description actually concerns only the simplest kind of CS: Often, several models will be integrated by appeal to dedicated integration modules (Durán, 2020), and inferences back to the target then become even more complicated. For instance, it might only be possible to validate the total, integrated model as a whole, so that it becomes unclear where to locate potential errors. And the integration of different modules may require the introduction of botched-together pieces of coding, not at all motivated by one's conception of the target (Lenhard and Winsberg, 2010, 2011; Lenhard and Küppers, 2006; Boge and Zeitnitz, 2020).

Based on all this, inferences made from the behavior of the computer to the behavior of the targeted system are subject to the same uncertainties as mentioned above: If something radically unexpected happens, researchers are likely – and likely right – to interpret this as an indication that something on the path from theoretical modeling to physical implementation has gone wrong. The crucial point being that the result of a CS, as displayed on the computer screen, thus cannot possibly count as a readout in a measurement process *on the system of interest*: It either counts as a readout concerning the state of the computer itself, or as a piece of inferred information about the system of interest, as facilitated by the mapping-relations between the models mentioned above—with all the *modeling-related* uncertainties as induced by measures for model-integration, the need for discretization and the physical limitations of the computer in realizing a given model. Hence, $P3_p$ comes out as false on (II) as well.

Insofar as these are all the relevant alternatives, $P3_p$ equally rests on shaky grounds. Moreover, when we reject the premises of this argument, we obviously also have no serious grounds for accepting $P2_s$, and so for buying into the subsidiary argument.

4 The argument from calibration practices

4.1 Argument: Model tuning serves the same purpose as instrument calibration

A couple of authors have recently drawn attention to the fact that some models have free parameters that are usually 'calibrated' to data, in similar ways as this must be done with the flexible parts of a physical measuring instrument. The first claim to this that I am aware of is in (Hasse and Lenhard, 2017).

Actually, Hasse and Lenhard struggle to make sense of calibration or 'tuning' practices in modeling-intensive contexts, such as thermodynamics, as there are "similar terms with (only) slightly differing connotations" (102) around. 'Tuning', they claim, "has a slightly pejorative meaning", whereas "[c]alibration [...] makes models look a bit like precision instruments" (ibid.). This is reason enough for them to reject the term 'calibration' and to opt for 'adjustment' instead.

However, Lenhard and Hasse's observations contrast with those made by Boge (2021b, 25), who believes that the interpretation of parameters in models suggested by the term 'calibration' is "on the right track":

> The delicate status of parameters exhibited [in some measurements] leaves us in no better position than to regard [certain] simulation models as reliable [...] instruments, insofar as they establish a[...] connection between data and [experimental results]. (ibid.)

Assuming, for the moment, that this supports yet another argument for models' functioning as measuring instruments, we may reconstruct the following *argument from calibration practices*:

P1$_c$ If the adaptation of the parameters θ_M of some model M serves the same purpose as the adaptation of an instrument I's parts to a measurement process P, then the function of M in P is the same as that of I.

P2$_c$ In some measurement processes P, the adaptation of the parameters θ_M of some models M serve the same purposes as the adaptation of an instrument I's parts to P.

C$_c$ Therefore, some M in some P can function as measuring instruments.

What are the reasons for thinking that P1$_c$ and P2$_c$ could possibly be true? First, P1$_c$ seems almost analytic, since 'purpose' and 'function' certainly have partially overlapping meanings. Yet one must be careful here: The 'purpose' refers to the adaptation, the 'function' to both model and instrument. Hence, this premise is certainly not trivial.

A more convincing reason to accept it may be advanced on the basis of Parker's notion of complex measurement: Since both models and physical instruments are clearly involved in all such measurements, it seems plausible that they can function in the same ways: As devices to get at a measurement result. And if the purpose of some adjustment performed on both is the same, this may well justify accepting their identical functioning in the overall measurement. So let's assume for now that P1$_c$ is at least sufficiently plausible (I will return to this point below).

Elaborating the reasons for accepting P2$_c$, however, requires more effort. I here draw on the case study presented by Boge (2021b), namely models used as the basis for CSs in the context of High Energy Physics (HEP) (also Morrison, 2015). High-energy physicists searching for traces of new particles, or measuring the properties of known ones, make use of various models that are often not intimately connected to HEP's fundamental theory (the 'Standard Model'). Furthermore, all of these models have various free parameters that, as in the thermodynamics case, need to be calibrated to data. However, in contrast to Hasse and Lenhard (2017), Boge (2021b) suggests to take this notion of 'calibration' rather seriously, based on distinct similarities between relevant practices.

In particular, some HEP researchers (Corcella et al., 2018) have suggested to calibrate certain models *in situ* in order to increase the accuracy of relevant measurements. The idea is to use the *very same data* on the basis of which a measurement result for some quantity, such as a particle's mass m, is supposed to be established to also adjust the parameters of the models used in the measurement. Obviously, this implies a threat of circularity: It must be done in such a way that the value to be measured does not influence the adjustment of the model, so that an *assumed* value for quantity m is not ultimately smuggled in through the use of the model. However, this problem is addressed in (Corcella et al., 2018) by assessing the sensitivity of the relevant 'calibration observables' to the quantity m and picking them such that their dependence on m is negligible.

Why think such a procedure offers a reason to take the notion of calibration quite seriously? Technicians highlight field (or *in situ*) calibration on physical instruments as advantageous in several respects: It means adjusting an instrument to the very conditions in which it is used (Cable, 2005). Thus, any error introduced by the mismatch between calibration- and measurement-conditions can be excluded. The *in situ* calibration of a simulation model essentially serves the same purpose: Both are strategies for reducing the error on the measurement result by making sure that the given device – model or physical instrument – is well-adapted to the overall measurement context.

Actually, even when the calibration is *not* done *in situ*, parallels arise. In this case, a model- or instrument-user must argue that the calibration-conditions are relevantly similar to the measurement conditions, so that any error introduced by the difference can be neglected (or is at least known to be small).[5] As Boge (2021b, 19) points out, high-energy physicists have a whole range of different 'tunes' for the very same model, and the suggested use of these tunes varies with the type of measurement they want to perform. So just as a group of physicists will take into account the *kind* of measurement they want to perform with a certain physical instrument, taking into account different ambient conditions, relevant models need to be calibrated differently depending on the measurement-purpose.

In sum, there is a close parallel between the purpose of calibration on models used for mea-

[5] Boge (2021b) actually applies Tal's (2012) account of calibration to these simulation models to offer a deeper analogy, but I'll here forego discussion.

surement and that of physical instruments used for the same purpose. This establishes P2_c as plausible: In some measurements, parameter adjustments on models and the adjustments of the parts of physical instruments serve the same purposes.

That the same purpose of a certain adjustment can justify that the devices adjusted (model or physical device) fulfill the same functions also delivers a plausibility argument for premise P1_c. Call this the *argument for identical functioning.* In more detail, it may be stated as follows:

P1_f If there is a class C_P of measurements P such that models M and instruments I are both used to get at a central result therein, then if the adaptation of the parameters θ_M of some such model M serves the same purpose as the adaptation of an instrument I's parts to a measurement process P, the function of M in P is the same as that of I.

P2_f In complex measurements, models M and instruments I are both used to get at a central result.

C$_f$ Therefore, if the adaptation of the parameters θ_M of some models M serves the same purpose as the adaptation of an instrument I's parts to the measurement process P in a complex measurement, then the function of M in P is the same as that of I.

The argument from calibration practices relies on this argument as a support.

4.2 Rebuttal: Shared purposes do not imply same functioning

So what about the argument from calibration practices? The main problem I see with it lies in its support by the argument for identical functioning; specifically, in premise P1_f. First note that this is equivalent to the following:

P1_f' If there is a class C_P of measurements P such that models M and instruments I are both used to get at a central result therein *and* the adaptation of the parameters θ_M of some such model M serves the same purpose as the adaptation of an instrument I's parts to a measurement process P, then the function of M in P is the same as that of I.

Hence, to rebut $P1_f$' (and, a fortiori, $P1_f$), it suffices to show that both antecedent conditions can be met but the consequent condition remains false.

To see this, consider an arbitrary complex measurement, such as that of a particle mass, m, in HEP. Here, both models and physical instruments clearly serve the purpose of getting at the central result (the particle mass, with specified uncertainty margin). Furthermore, as already argued above, both instrument calibrations in the traditional sense and model calibrations performed on simulation models will be used to reduce the uncertainty margin, so they too serve the same purpose.

However, is it correct to say that models and physical instruments hence have the same function here? I believe that this is unjustified. Refer once more to the dual nature of complex, or even just derived measurements: They involve both physical actions taken on the system and inferential steps for getting at the desired result.

Thus, a similar line of argument is possible against the central implication in $P1_f$' as was possible against $P2_i$: Just because models and physical instruments share a common goal, and just because, relative to that goal, certain adaptations of them serve the same purpose, this still doesn't mean that they fulfill the same functions *if that goal is itself complex.*

Specifically, when 'complex' is interpreted to mean that the given goal involves integrating information from distinct sources – such as model-based inferences and physical devices – we can clearly see that goals in derived and complex measurements *are* complex. Hence, models and measuring devices may be used to get at a central result and adaptations (calibrations) performed on them may serve the same purpose, but they may fulfill different functions nevertheless.

For example, consider the measurement of the the top quark mass, m_t. In order to retrieve a credible result, physicists use measurements from different experiments at the Large Hadron Collider (LHC), from different decay-channels therein, but also from other colliders, such as the Tevatron at Fermilab. These are then combined in such a way as to obtain an unbiased estimate that takes into account all uncertainties contained in individual measurements. In such an effort, the following three categories may be distinguished (ATLAS, CDF, CMS and D0 Collaborations, 2014, 5–6):[6] uncertainties (i) "from the limited understanding of the detector response to (and the modelling of) different types of jets", (ii) "related to the [Monte Carlo (MC) simulation] modelling of the [...] signal", that is, "from the specific choice of the MC generator and the associated [...] models used", and (iii) "systematic uncertainties stemming from detector resolution effects, reconstruction efficiencies, and the b-tagging performance in data relative to the MC", including "effects related to normalisation and differential distributions of backgrounds events, and the modelling of the data-taking conditions in the MC simulation relative to the data".

At first glance, this categorization may seem to *defy* the claim I am arguing for: Uncertainties stemming from different sources (models and instrumentation) are mixed up in (i)–(iii). Thus, models and instruments seem to be treated on a par. But a closer look reveals that appearances are deceptive, as these categories still distinguish, within themselves, between uncertainties that arise purely from models and such uncertainties that arise purely from the physical properties of the detector: Detector resolution effects as created by the finite size of the different cells of the detector are clearly distinguished from, e.g., artifacts created by the choice of Monte Carlo simulation model used for determining this impact. Thus, the given categorization cuts across—but does not invalidate—the categorization of uncertainties as induced by their source in either instrumentation or modeling.

To see this clearly, observe that changing the impact of resolution effects would require an *action*, i.e., 'going out into the world': It would require *manufacturing* smaller cells that do not give rise to the same sort of complication as larger ones. The effect of this action would manifest itself in the *physical contact* with the debris from the scattering process of interest. In contrast, changing the impact of potential modeling artifacts would require a mere *thought process*: It would require thinking up new ways of conceptualizing the debris or the scattering, but apart

[6] I eschew a discussion of the physics terms here; introductions at varying levels of detail can be found in Boge and Zeitnitz (2020); Boge (2021b); Ritson and Staley (2020); Staley (2020).

from using (say) pen and paper, would not require actions within the real world. Using pen and paper for conceptualizing and drawing conclusions in the mind would *not* alter the physical contact with the system in an interesting sense.

It is important to carefully distinguish physics from epistemology here: For instance, in order to evaluate the "experimental uncertainty" due to jet energy resolution, ATLAS uses two methods, one of which involves varying the cut on the differences in azimuthal angle as well as the varying the parameters in a model used to correct for soft radiation effects (see Aad et al., 2013, 10). Thus, physicists do use improved modeling and corrections to alter measurement results, seemingly defying the distinction I am making.[7]

But appearances are deceptive: The impact of resolution effects is *estimated* via multiple methods, some of which are model-based. That doesn't mean that these methods *effect actual changes* in the resolution *or* its contribution. They are just this: ways to *estimate* the contribution. Thus, the central implication in P1$_f$' rests on shaky grounds: Even if usages of models and physical instruments share common goals in measurement, and even if calibration on them may serve a common purpose (uncertainty reduction), this does not establish that they function identically therein.[8]

Here is an important caveat: I am not claiming that things like 'particle jets', or even elementary particles, are lying out there to be passively discovered by hooking them up to well-adjusted devices. All I am saying is that there are two distinct sources of insight into reality that should be kept apart. So we can address whatever *is* out there either by interacting with it, or by making up our minds about it. Of course, both activities depend on each other, but that doesn't make them 'of a piece with one another'. I will make these ideas explicit below.

[7] I owe this objection to an anonymous referee as well.

[8] In fact, Beauchemin's (2017) analysis supports my case: As Beauchemin (2017, 282) points out, "systematic uncertainties [...] quantify the potential variability of an experimental outcome due to all the anticipated variations of the many theoretical assumptions [...] needed to obtain the result." Thus, at least part of the so-called 'systematic' uncertainties explicitly stem from theoretical preconceptions, not from the equipment itself.

5 Not Measuring, but Cognitive Instruments

5.1 Models and measuring instruments reconsidered

The foregoing suggests that models do not function as measuring instruments when employed in measurement contexts. But then what *are* the differences in functioning between models and measuring instruments, and how do models function in measurement? I will here sketch a positive account that explains the functioning of models in measurement and highlights the differences to measuring instruments, properly-so-called. I will then go into the epistemic implications of the resulting distinction.

Recall that, following Parker (2017, 280), a measuring instrument is a device or apparatus whose "physical state [is] used in measuring, [...] as a pointer position or a digital display showing a numerical readout". One key property of measuring instruments thus is that their physical state, at some point in the process of measuring, can be used as a numerical readout, such as a pointer position, an oscilloscope with a scale, a digit displaying a number of particle hits or the sweat level on a test person's skin, and so on. Furthermore, the state should count as a 'readout' exactly when changes in a certain property of the system under study can be expected to effect a corresponding change in the system used as an instrument.

This relation is typically not flawless: Parker's (van Fraassen's) example of the thermometer that needs to be corrected serves to exemplify the point. Stated differently, background knowledge may suggest that the read-out needs to be transformed ex post by means of an error-correction procedure.

So measuring instruments are physical devices whose final state, after the interaction with a system of interest, can be interpreted as a numerical readout which is informative about the state of the measured system, though not generally in a perfect, flawless way. Their *function* thus is to map, by appeal to a physical correlation, the properties of a studied system to numbers one can base calculations on: To provide numerical data. How, in contrast, do *models* function in measurements?

Based on his investigation of calibration practices, Boge (2021b, 25; emphasis added), for instance, argues that models are "*cognitive* [...] instruments, insofar as they establish an inferential connection between data and [quantitative experimental results] without painting a trustworthy picture of the underlying reality."[9] Thus, maybe we *can* view models in measurement as instruments of *some kind*; but since they do not exploit a physical correlation to provide data, maybe they should be seen as 'cognitive' rather than measuring instruments.

So, what exactly *is* a cognitive instrument? Boge (2021b, 25) claims that models used in measurements "establish an inferential connection" between data and measurement results, which sounds somewhat like the (long-rejected) 'inference ticket' account of theoretical claims from the realism-debate (e.g. Stanford, 2016), according to which theories and models are rules of inference rather than claims about nature. This is a semantic thesis that was vigorously disputed already by Nagel (1961, 139):

> neither logic nor the facts of scientific practice nor the frequently explicit testimony of practicing scientists supports [...] construing theories simply as techniques of inference.

Hence, this better not be the whole story. In fact, Boge (2021b, 25) does side with 'sophisticated' instrumentalists, such as Stanford (2006), whence his endorsement of an instrumental status of models in measurement certainly goes beyond the inference-ticket view. But what does this entail?

5.2 Cognitive instruments provide understanding (but not necessarily truth)

An anti-realist position that goes by the label 'cognitive instrumentalism' has recently been developed by Rowbottom (2019a). For Rowbottom (2019a, 1), cognitive instrumentalism includes

[9] This shows that he does not, in fact, embrace the argument from calibration practices.

"the view that science is primarily [...] an instrument for furthering our practical ends". On top of that, it "has three core components", namely that

> [(i)] progress in science centrally involves [...] what it enables us to understand about and do with observable things; [(ii)] scientific discourse may only be understood literally when it is grounded in talk about observable things; and [(iii)] science may reliably progress [...] without discovering new truths [...] about unobservable things. (ibid.)

In short, cognitive instrumentalism involves the view that science progresses by advancing our *understanding of observable things*. The use of 'understanding' here is unproblematic, insofar as several distinct accounts of understanding require only a loose 'tethering to the facts' (Rowbottom, 2019a; De Regt and Gijsbers, 2017; Elgin, 2017), not the full objectivity (and truth) of the models, theories and explanations used to promote understanding. Hence, understanding only of 'observable things' may well be a kind of understanding.

The glaring problem associated with such a position is its reliance on a notion of 'observ*ability*'—a modally laden notion that is supposed to ground all theoretical claims, *including modal ones* (Ladyman, 2000). Furthermore,

> In important areas of science the only perceptual experiences a scientist has when conducting her experiment are with a computer screen. But it cannot be the aim of science to predict our perceptions of computer screens. (Bird, 2022, 137)

Thus, if 'observability' is tied closely to sense-perception, cognitive instrumentalism may provide an implausible view of science.[10]

These objections might be countered by means of a careful analysis of the notion of 'observation'. For instance, one might distinguish 'observation' in a technical sense from 'field

[10] This holds true even though Rowbottom applies questions of observability to properties, not entities: "it's a myth that [...] so-called observable properties are known in experience; rather, with only a few exceptions, the more interesting properties ascribed to entities are introduced via theories [...]. Mass, charge, potential, energy, entropy are just some obvious examples of theory-driven properties." (Psillos and Zorzato, 2020, 7–8)

observation' and 'experiential observation', where the first term means the successful establishment of some claim about a given system by means of causal contact with it, the second an unperturbing act of data-taking, and only the third the recognition of something experienced (Boge, 2024). However, opting for any of the first two readings in defining 'observability' would mean leaving serious anti-realist positions behind.

Instead of fixating on these problems with observability, though, it seems instructive to focus on the account of *understanding* endorsed by Rowbottom. In advancing his own account of 'empirical understanding', Rowbottom (2019a, 116–7) centrally builds on Grimm's (2012, 107–9) notion of subjective understanding, which involves "a kind of legitimate satisfaction that accompanies our experience of 'having made sense of things'" and obtains when "one has grasped a model of how the world works that 'makes sense'". Grimm (2012, 107) contrasts this with a notion of *objective* understanding, which additionally implies that "one's mental model of the world is accurate".

For Grimm (2012, 108), "objective understanding is plausibly something that we desire both for its own sake as well as for the promise it offers of being able to control the world". Hence, objective understanding is both intrinsically and extrinsically valuable: Intrinsically because it is a state of mental satisfaction, and extrinsically because of the abilities it equips us with (in contrast to purely subjective understanding). This is consistent with various recent accounts that hold understanding to be that which science ultimately strives for (e.g. Dellsén, 2016; de Regt, 2017; Elgin, 2017).

However, do we need to assume the accuracy of the modeling assumptions to construe scientific progress in terms of understanding? I believe this is not so: Since empirical understanding implies the ability to successfully grasp the connections between things observed, it will also promote control over aspects of the world: If, without *accurately* representing the goings on *tout court*, a model facilitates successful inferences or enables engineering successes, this will be enough to harvest the desired control accompanying (the not merely subjective forms of) understanding.

In fact, the *legitimacy* of the mental satisfaction associated with the grasping of the relevant model may be rooted exactly in the promoted control over phenomena: "a model can provide mental satisfaction [...] in so far as one grasps it *and* takes it to be sufficient for a desirable end." (Rowbottom, 2019a, 118; orig. emph.)

For instance, it is well known that the predictions made by the appeal to the zodiac do not stand up to empirical scrutiny at all (Carlson, 1985). So an astrological world-model built around the zodiac may provide mental satisfaction, but it is very much unsuitable for relevant actions; it is in a clear sense *illegitimate*. This is different with models such as those underlying the ideal gas law, even though they get the relevant interactions quite wrong, and quantum physics may ultimately require us to reject the notion of 'particles' these models centrally involve (e.g. Halvorson and Clifton, 2002).

In fact, the ideal gas law is well known to be highly inaccurate in many ways, yet it is often cited as (deriving from) a model that clearly fosters scientific understanding (de Regt, 2017; Elgin, 2017; Strevens, 2013): Using the ideal gas law to infer something about the relation between pressure and volume in certain engineering efforts can directly lead to successful construction-principles under certain constrained conditions (e.g. Moses et al., 2024, 250). Thus, for models to fulfill their function as cognitive instruments, and to be sufficient for desirable ends, they need not get everything entirely correct; they may in fact get a great many things wrong.

The bottom line is that models, construed as cognitive instruments, can provide understanding in a sense that facilitates a *basis* for successful action; including inferences (mental actions) but also physical actions (such as engineering efforts). Nevertheless, understanding per se *is not* yet an action. Further, this understanding does not necessarily involve the truth of the modeling-assumptions, or only so to a limited extent: If models are to serve as cognitive instruments, we definitely need to get the connections between things observed right in our heads by means of them, but not necessarily anything beyond that.

Measuring instruments are entirely different: They are physical things acted upon, and at the same time things with which we act upon other physical things: By calibrating and tweaking them we may bring about novel phenomena (also Heidelberger, 2003, 146 ff.); and by causally interacting with whatever we want to measure through them, we may put numbers on the measured system (viz., on its states and properties), as it reveals itself through the interaction.

How is this view connected to the arguments from the paper's first part? As I have argued in the rebuttal of the argument from indispensability, indispensabilty does not commit us to seeing models as measuring instruments, and the arguments given in sect. 2.2 in support of this claim equally support the view advanced here: models are the basis for inferences, measuring devices are the basis for actions on, and interactions with, the world. We need both to measure, but that does not make them 'one and the same'.

Furthermore, as argued in the two rebuttals to the argument from physical involvement, one should not be mistaken to slide from some device, such as a programmed digital computer, being *relevant* to a measurement to it *being* a measuring device. In the most extreme case, that might make a coffee maker a measuring device: Imagine that some brilliant technician has developed a strong caffeine addiction and needs to first pour in some hot coffee before she can operate the measuring devices right. There is definitely some contrived causal connection between the measured system and the coffee maker now, and the coffee maker is physically involved in the measurement. Clearly, that does not make it a measuring instrument though.[11]

In particular, the first rebuttal said that the physical involvement of a CS for correcting errors in a measurement is insufficient for establishing it as providing instrument indications: If the computer delivers a 'readout' here at all, the readout is not an instrument indication; it is an indication as to what to expect on account of the model implemented. Similarly, in the second rebuttal, I argued that even the notion of a 'numerical readout' is misplaced, as it

[11] Incidentally, if we imagine the technician to be the only one capable of operating the machine, this would make the coffee maker even *indispensable* for the measurement, thus yielding a kind of additional reductio against the indispensability argument.

wrongly suggests that the CS is directly informative about the state of the measured system. This neglects the various modeling and re-modeling steps necessary (Durán, 2020), and the (uncertain) inferences thereby supported (Boge, 2020), for getting from computer display to system behavior. These objections are congenial to the view advanced here: If we use a CS to model a system's behavior, this may allow us to understand that behavior better. This may subsequently allow the design of different equipment, experiments, or even theories of that system. But thus evaluating a model of some system with the help of a computer, *based* on information gathered with measuring devices such as particle detectors, is not the same as directly altering the contact with the system by altering the measuring device.

Finally, making reference to relevant case studies, I argued that the sharing of a common purpose – getting at a credible measurement result – does not imply identical functioning for models and measuring devices; even if parameter fits may sometimes make it seem this way. With the views advanced in this section, we are now in a better position to understand that proposal: Both models and measuring instruments may contribute to our access to the physical world within measurements. However they should be seen as complementary, and mutually dependent, sources of insight into the world: models give us cognitive access, measuring devices give us physical access; and only together do these two sources of insight lead us to credible measurement results. I will flesh this idea out further in the next section.

5.3 Why bother?

I have outlined how I think models can function as cognitive instruments and how this differs from the functioning of proper measuring instruments. But why should we care about the distinction so drawn? For instance, the numbers gathered by means of measuring instruments will normally be used in further inferences and modeling. Thus, they will ultimately also enrich the scope of actions we can perform in the world. And does this not show that both function in the same ways after all?

To appreciate the difference, as well as its epistemological impact, consider how there is a clear dependence of measuring instruments on models: It is only through models that we can grasp what it is that a measuring instrument measures. The point has been made extensively by Tal (2012), who offers a model-based account of measurement: In order to understand what happens when they use a measuring device, scientists need an elaborate model of that device (Tal, 2012, 17–20). However, at the same time, they of course need a model of the stuff being measured: Without any preconception of the phenomenon targeted by a measuring device, it is unclear how any action performed with the device could count as more than idle play, let alone a measurement of some system's properties.[12]

In turn, we saw that models also depend on measurements: If we measure the properties of a system more accurately, this may change the way we conceive of said system entirely. A case in point is Michelson and Morley's (1887) measurement of the speed of light in relation to a purported luminiferous ether: Their null result ultimately suggested that there is *no such thing* as a luminiferous ether. However, if it is correct that models and measuring instruments mutually depend on each other in these ways, they indeed cannot be identified or equated.

As I see it, we have two source of insight into the world that have to function *in concert* in order to yield success in scientific measurements, and to yield genuine insight: physical devices with which to perform actions on the world, some of which may be called measuring instruments; and cognitive devices with which to reason about the world, some of which may be called models (or cognitive instruments). If we change the actions we perform, that may inform the ways we reason about the world, viz., our models; and if we change the way we reason, that may inform how to act successfully upon the world and even what it is that we are acting upon, viz., our use and construction of measuring instruments. Both directions of influence are entirely distinct, even though they may often happen in succession. Hence, again, models and measuring instruments should not be equated with each other.

[12]Exploratory experimentation might be an exception, but it usually also does not operate in a conceptual void, but rather in the context of overthrowing one conceptual framework by another (e.g. Steinle, 1997, S66).

The reader may worry about vicious circularity here, as I have claimed a mutual dependence between models and measuring instruments. However, the process described is rather an instance of what Chang (2004, 45) calls 'epistemic iteration':

> a process in which successive stages of knowledge, each building on the preceding one, are created in order to enhance the achievement of certain epistemic goals.

Notably, the 'epistemic goals' here need not, or at least not immediately, involve truth (Chang, 2004, 46). Furthermore, while Chang advances epistemic iteration as part of his theory of measurement, my reason to appeal to it is different: My main suggestion is that our conception of what it *is* that we actually measure may drastically change with changes in the ways we model the measured systems. Thus, for instance, activities originally interpreted as revealing the properties of tiny lumps of discrete matter are nowadays interpreted as revealing the properties of disturbances in extended fields by physicists, on account of the 'Standard Model' of particle physics (e.g. Zee, 2010, 24).

The point is that our *understanding of* the reality that we measure has often changed over time. This is possible since we *need models* in measurement, *next to* measuring devices and, as was pointed out above, because there is only a loose connection between understanding and truth. So many models known (or at least suspected) to contain various falsehoods have served us well in achieving understanding, in the action-related sense also discussed above, while leaving major room for conceptual improvement and the re-interpretation of results. Further, as has been pointed out by various commentators (Laudan, 1977; Stanford, 2006; Rowbottom, 2019b; Boge, 2021a; Frost-Arnold, 2019; Kuhn, 1962) the conceptual changes to come with different conceptualization and modeling are difficult, if not impossible, to anticipate. If we *equate* models with measuring instruments, instead of acknowledging the mutual dependence between the two, we loose sight of these important epistemological lessons.

6 Conclusions

I have argued that models in measurement are not to be literally construed as measuring devices. To do so, I have considered and rebutted three distinct arguments that may be reconstructed from the literature: The argument from indispensability (after Morrison, 2009), that from physical involvement (after Parker, 2017), and that from calibration practices (after Boge, 2021b). As I have shown, all of these arguments rest on doubt worthy premises: The argument from indispensability essentially confuses indispensability with identity, the argument from physical involvement conflates devices used to physically implement numerical models with physical measuring instruments, and the argument from calibration practices essentially confuses identity and *analogy*.

Further, I have argued that models function as cognitive rather than measuring instruments, in the sense that they provide understanding, and with it, an advanced scope for action. In contrast, measuring instruments are things directly involved in (certain kinds of) physical action. Distinguishing models and measuring instruments in this way allows us the see the mutual *dependence* between action and cognition, instead of lumping everything together. This is important, among other things, as it makes us aware of the fact that, while performing the same kinds of actions on the world, we may conceive of ourselves as measuring entirely different things, depending on the models we endorse.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest: None

Research involving Human Participants and/or Animals: None

Informed consent: Not applicable

References

Aad, G. et al. (2013). Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ tevrecorded in 2010 with the atlas detector. *The European Physical Journal C*, 73(3):2306.

ATLAS, CDF, CMS and D0 Collaborations (2014). First combination of tevatron and lhc measurements of the top-quark mass. *arXiv prepint hep-ex/1403.4427*.

Beauchemin, P.-H. (2017). Autopsy of measurements with the atlas detector at the lhc. *Synthese*, 194:275–312.

Beisbart, C. (2012). How can computer simulations produce new knowledge? *European Journal for Philosophy of Science*, 2(3):395–434.

Beisbart, C. and Norton, J. D. (2012). Why Monte Carlo simulations are inferences and not experiments. *International studies in the philosophy of science*, 26(4):403–422.

Bird, A. (2022). *Knowing Science*. Oxford, New York: Oxford University Press.

Boge, F. J. (2019). Why computer simulations are not inferences, and in what sense they are experiments. *European Journal for Philosophy of Science*, 9(13):30 pp.

Boge, F. J. (2020). How to infer explanations from computer simulations. *Studies in History and Philosophy of Science Part A*, 82:25–33.

Boge, F. J. (2021a). Incompatibility and the pessimistic induction: a challenge for selective realism. *European Journal for Philosophy of Science*, 11(2):46.

Boge, F. J. (2021b). Why trust a simulation? models, parameters, and robustness in simulation-infected experiments. *British Journal for the Philosophy of Science*. https://doi.org/10.1086/716542.

Boge, F. J. (2024). Re-assessing the experiment / observation-divide. *Philosophy of Science*. https://doi.org/10.1017/psa.2024.23.

Boge, F. J. and Zeitnitz, C. (2020). Polycratic hierarchies and networks: What simulation-modeling at the LHC can teach us about the epistemology of simulation. *Synthese*. https://doi.org/10.1007/s11229-020-02667-3.

Boumans, M. (2005). Measurement outside the laboratory. *Philosophy of Science*, 72(5).

Cable, M. (2005). *Calibration: A Technician's Guide*. ISA.

Carlson, S. (1985). A double-blind test of astrology. *Nature*, 318:419–425.

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

Corcella, G., Franceschini, R., and Kim, D. (2018). Fragmentation uncertainties in hadronic observables for top-quark mass measurements. *Nuclear Physics B*, 929:485–526.

Currie, A. (2020). Bottled understanding: The role of lab work in ecology. *The British Journal for the Philosophy of Science*, 71(3):905–932.

Dardashti, R., Thébault, K. P. Y., and Winsberg, E. (2017). Confirmation via analogue simulation: What dumb holes could tell us about gravity. *The British Journal for the Philosophy of Science*, 68(1):55–89.

de Regt, H. W. (2017). *Understanding Scientific Understanding*. Oxford, New York: Oxford University Press.

De Regt, H. W. and Gijsbers, V. (2017). How false theories can yield genuine understanding. In Grimm, S. R., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: New Essays in Epistemology and the Philosophy of Science*, pages 50–75. London: Routledge.

Dellsén, F. (2016). Scientific progress: Knowledge versus understanding. *Studies in History and Philosophy of Science Part A*, 56:72–83.

Durán, J. M. (2020). What is a simulation model? *Minds and Machines*, 30(3):301–323.

Durán, J. M. (2024). Computer simulations. In Knuuttila, T., Carrillo, N., and Koskinen, R., editors, *The Routledge Handbook of Philosophy of Scientific Modeling*, pages 149–163. Routledge.

Elgin, C. (2017). *True Enough*. MIT Press.

Frost-Arnold, G. (2019). How to be a historically motivated antirealist: The problem of misleading evidence. *Philosophy of Science*, 86(5):906–917.

Giere, R. N. (2009). Is computer simulation changing the face of experimentation? *Philosophical Studies*, 143:59–62.

Grimm, S. (2012). The value of understanding. *Philosophy Compass*, 7(2):103–117.

Gross, C., Strobel, H., Nicklas, E., Zibold, T., Bar-Gill, N., Kurizki, G., and Oberthaler, M. K. (2011). Atomic homodyne detection of continuous-variable entangled twin-atom states. *Nature*, 480(7376):219–223.

Halvorson, H. and Clifton, R. (2002). No place for particles in relativistic quantum theories? *Philosophy of science*, 69(1):1–28.

Hasse, H. and Lenhard, J. (2017). Boon and bane: On the role of adjustable parameters in simulation models. In Lenhard, J. and Carrier, M., editors, *Mathematics as a Tool*, pages 93–115. Springer.

Heidelberger, M. (2003). Theory-ladenness and scientific instruments in experiementation. In Radder, H., editor, *The Philosophy of Scientific Experimentation*, pages 138–151. Pittshurgh, Pa.: The University of Pittshurgh Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* University of Chicago Press.

Ladyman, J. (2000). What's really wrong with constructive empiricism? van fraassen and the metaphysics of modality. *The British Journal for the Philosophy of Science*, 51(4):837–856.

Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth.* Berkeley, CA: University of California Press.

Lenhard, J. (2007). Computer simulation: The cooperation between experimenting and modeling. *Philosophy of science*, 74(2):176–194.

Lenhard, J. and Küppers, G. (2006). From hierachical to network-like integration: A revolution of modeling style in computer-simulation. In Lenhard, J., Küppers, G., and Shinn, T., editors, *Simulation: Pragmatic Construction of Reality*, pages 89–106. Springer.

Lenhard, J. and Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Science Part B: Studies in History and*

Philosophy of Modern Physics, 41(3):253–262.

Lenhard, J. and Winsberg, E. (2011). Holism and entrenchment in climate model validation. In Carrier, M. and Nordmann, A., editors, *Science in the Context of Application*, pages 115–130. Springer.

Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature*, 444(7122):1022–1023.

MacMahon, D., Pearce, A., and Harris, P. (2004). Convergence of techniques for the evaluation of discrepant data. *Applied Radiation and Isotopes*, 60(2):275–281. Proceedings of the 14th International Conference on Radionuclide Metrology and its Applications, ICRM 2003.

Massimi, M. and Bhimji, W. (2015). Computer simulations and experiments: The case of the higgs boson. *Studies in History and Philosophy of Modern Physics*, 51:71–81.

Michelson, A. A. and Morley, E. W. (1887). On the relative motion of the earth and the luminiferous ether. *American Journal of Science*, 34:333–45. https://doi.org/10.2475/ajs.s3-34.203.333.

Morgan, M. S. (1999). Learning from models. In Morgan, M. S. and Morrison, M., editors, *Models as Mediators*, pages 347–88. Cambridge University Press.

Morrison, M. (1999). Models as autonomous agents. In Morgan, M. S. and Morrison, M., editors, *Models as Mediators*, pages 38–65. Cambridge University Press.

Morrison, M. (2009). Models, measurement and computer simulation: the changing face of experimentation. *Philosophical Studies*, 143(1):33–57.

Morrison, M. (2015). *Reconstructing Reality: Models, Mathematics, and Simulations.* Oxford University Press.

Morrison, M. and Morgan, M. S. (1999). Models as mediating instruments. In Morgan, M. S. and Morrison, M., editors, *Models as Mediators*, pages 10–37. Cambridge University Press.

Moses, J., Vijumon, P., and Muthu, M. (2024). *Engineering Thermodynamics.* RK Publication.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation.*

Harcourt, Brace & World.

Parker, W. S. (2009). Does matter really matter? computer simulations, experiments, and materiality. *Synthese*, 169(3):483–496.

Parker, W. S. (2017). Computer simulation, measurement, and data assimilation. *The British Journal for the Philosophy of Science*, 68(1):273–304.

Psillos, S. and Zorzato, L. (2020). Against cognitive instrumentalism. *International Studies in the Philosophy of Science*, 33(4):247–257.

Ritson, S. and Staley, K. (2020). How uncertainty can save measurement from circularity and holism. *Studies in History and Philosophy of Science*. https://doi.org/10.1016/j.shpsa.2020.10.0040.

Rowbottom, D. (2019a). *The Instrument of Science: Scientific Anti-Realism Revitalised*. Taylor & Francis.

Rowbottom, D. P. (2019b). Extending the argument from unconceived alternatives: observations, models, predictions, explanations, methods, instruments, experiments, and values. *Synthese*, 196:3947–3959.

Staley, K. W. (2020). Securing the empirical value of measurement results. *The British Journal for the Philosophy of Science*, 71(1):87–113.

Stanford, K. P. (2006). *Exceeding Our Grasp. Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Stanford, P. K. (2016). Instrumentalism: Global, local, and scientific. In Humphreys, P., editor, *The Oxford Handbook of Philosophy of Science*, pages 318–36. Oxford, New York: Oxford University Press.

Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation. *Philosophy of Science*, 64(S4):S65–S74.

Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A*, 44(3):510–515.

Tal, E. (2012). *PhD Thesis: The epistemology of measurement: A model-based account.* University of Toronto (Canada).

Van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective.* Oxford University Press.

Wright, A. J. and Jackson, J. J. (2022). Childhood temperament and adulthood personality differentially predict life outcomes. *Scientific Reports*, 12(1):10286.

Zee, A. (2010). *Quantum Field Theory in a Nutshell.* In a Nutshell. Princeton University Press.