

ARTICLE

The Logic of Measurement: A Defense of Foundationalist Empiricism

Mariam Thalos 

University of Tennessee, Knoxville, Knoxville, TN, USA
Email: m.thalos@gmail.com

(Received 5 October 2022; revised 6 March 2023; accepted 14 April 2023)

Abstract

Practitioners of science treat *evidence* as a separate and objective body of materials that is independent of, and possibly also prior to, all of theorizing. Philosophers of science, by contrast, are increasingly wary of the role of theory in testing and measurement contexts, and hence have problematized the notion of evidence as prior or independent, even in the context of measurement. This paper argues that there is an important sense in which empirical certification of a quantity, via measurement, is indeed prior to theorizing, albeit not necessarily in order of time. The case for this priority distinguishes between the certification of the measurability of a given quantity, as a quantity appropriately measured on a specified scale, and the epistemic warrant due to an assignment of a specific magnitude to that quantity on a given occasion. The result is an account of the certification of a measurable quantity, independent of any theory in which that quantity features. The effect is to render certification of quantities theory-neutral. The aim of the essay is thus to bolster and re-establish a more nuanced empiricist view, via building a case for quantity certification as the epistemic basis (i.e., foundation) of the scientific enterprise.

Keywords: Theory of measurement; RTM; foundationalism; measurement realism; theory of scales; Patrick Suppes; R. Duncan Luce; Louis Narens; Eran Tal; Hasok Chang

1. Introduction

Foundationalist empiricism (FE) is the conceptual muscle behind the language of “evidence-based X” in the larger scientific community. FE rests on the idea that the epistemic corpus can – and indeed must – be divided between that which is “evidence,” on one side, and everything else on the other. The evidence is made up of a wide variety of “data points”: bits of bone, fingerprints, survey data, materials and measurements generated in a laboratory setting, and compilations, tables, and sometimes more elaborated graphical representations of all these, and other things besides. The other side – the side that may be said to “make something” of evidence – might in addition to the fancy theories and hypotheses that loom large on that side of things, also include *rules* and *norms* of evidence, evidence collection practices and assessment, and a dynamic body of broader practices that have evolved over time. FE, in its rawest form, is a doctrine

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

that the former corpus of materials (the evidence) is to be treated as the ground and guarantor – indeed, the foundation – of all that lies on the other side of the dividing line, on the basis that it is an independent repository of materials, and consequently, for reasons I will take up in this essay, above a certain sort of suspicion.

Philosophers have prosecuted the case against foundationalisms for hundreds of years, the most recent skirmishes directed specifically against the positivist movement of the early twentieth century, with its distinctive form of foundationalism. I shall not be defending any form of positivism. Still, any foundationalist posture has to confront the wedge that has proved most successful in opening up and then prosecuting the case against positivism: the highly sticky idea of the theory-ladenness of observation. A certain loose cluster of philosophers of science has recently joined that prosecution by advancing the idea of theory-ladenness *of measurement itself*, which has long been viewed as the cornerstone of the scientific enterprise. And so we have arrived recently at the idea that the quantities theories speak of as the objects of measurement (direct or otherwise) belong to the category of inventions rather than that of discoveries. This idea is the target of my argument in this essay.

I will be arguing that, even if there is merit to the idea of theory-ladenness of observation, as it concerns occasions of observation (thought of as specific sequences of events involving psycho-social activity that results in assignment of specific magnitudes to observables/quantities), this theory-ladenness does not impugn the construal of measurement as *the assignment of appropriate scales to the quantities in question*, in the process *certifying* the quantity as occupying a rightful place on the “evidence” side of the dividing line we have been speaking of. This construal of measurement is the cornerstone of a body of theory around what is traditionally known as the “mathematical theory of measurement.”

The argumentative core of the present essay will be to separate the case for the *certification* of quantities – the (logical/mathematical) case-building procedure that assesses appropriateness of a scale for measurement of a given quantity, usually within a mathematical discipline – from the case for *accurate measurement* of those self-same quantities usually by individual researchers, with or without improvised instrumentation, on a wide range of specific occasions (e.g., as an air temperature of 28 degrees C, at the Dallas airport on Friday, September 15, 1928). To do that I will describe a body of theory that is completely separate and independent of substantive scientific theories within the scientific subdisciplines. This independent body of theory – the *mathematical theory of measurement* – can provide the basis for an independent explanation of *quantity certification*, without reference at all to such *scientific* theories of substance (theories of heat and temperature, theories of economic growth, theories of the contingent objects of the world as we know them) as the quantities in question themselves might enter into.

The body of theory on which I shall draw for making the case for quantity certification explains such confidence in the status of measurement as science might legitimately place. And, when measurement does deserve that confidence, the explanation will be completely separate and independent of such scientific theories (in physics, or economics or psychology or what-have-you) as the quantity in question might figure. I will say that this body of explanatory theory *explains measurement success* of the relevant quantities (when that success has been earned) – that measurement theory explains *certification* of the associated quantities, as a matter entirely separate from all scientific theory of empirical phenomena associated with the relevant quantities.

If the case against FE is thought to rest most securely upon the theory-ladenness of measurement, then *certification* – as an independent explanation of measurement success – should be deemed to undermine that case, and thereby to rehabilitate the viability of FE. The aim of the essay was thus to provide a basis for a more nuanced empiricist view of measurement.

It is important to emphasize that the certification I shall speak of is on a case-by-case basis; it is not a global case for measurement of any dreamed-of (or, if you prefer, made-up) quantity. In other words, a case for certification has to be made for each proposed measurable quantity, as a type rather than as a token. For example: length (as a type, rather than as the length of a specific object), mass (as a type, rather than as the mass of a particular body, or even a type of body), and so on. The case for certification is contingent upon identifying suitable empirical operations, as I will explain. It is also important to emphasize that certification is largely a matter of the *metaphysics* of measurement, even though it rests partly on observations. To be clear, the case that shall be made here is *not* that observation, as such, can be or has been conducted in such a way as to secure *epistemic accuracy* (so as to warrant a status of knowledge or anything else) on any given occasion. Certification of a quantity parallels closely the logic of the question of scientific explanation construed very widely, according to which the explanans draws on a body of *scientific* theory. In certification, the explanans draws on a body of *measurement* theory, the *mathematical theory of measurement*. Thus certification *explains* success in measurement of the relevant quantity or quantities, **in the most abstract or general case of that quantity**; it decidedly does *not* account for knowledge or accuracy of the magnitude of the quantity as measured on any given occasion. Certification of a quantity is an **explanation** whose target is the success of measuring the aforementioned quantity; it is *not* justification of any assertion as to the magnitude of the quantity measured on any particular occasion or any particular context. Because certification is not a matter of assigning specific magnitudes to quantities on specific occasions, it *can* be construed as independent of scientific theorizing around the phenomena in which the certified quantity might be implicated. The remainder of this essay explains why it also *must* be so construed.

Before we can get to the account of certification, we will need to review (in the next three sections) some foundations of measurement theory, the theory of scales, and the classical representational theory of measurement (RTM), which I will endorse with some innovations. The innovations are important to understanding the role of certification.

I am here painting foundationalism as something to preserve, but it is well known that many philosophers and philosophies of science draw heavily on the idea that foundationalism is neither necessary nor desirable. One competitor to foundationalism – namely, coherentism – has been widely embraced (Tal (2017b) and (2019) perhaps chief among them, supported by a number of arguments that one finds in Chang (2004), both of which I will engage at further length below). And at least one philosopher of science (Boyd 2010; 2019) has made hay out of theory-ladenness, arguing that realism and naturalism are equally well served by them because only the truth of theories and the reliability of the scientific methods employed in their development can explain how well they perform in practice. But Boyd's argument leaves it open just how one can, neutrally, judge that a theory performs well in practice. The standard of "projectability" he invokes begs to be interpreted as "via confirmed predictions," and this in turn to be understood in terms of reliable (and indeed confirmed) measurement outcomes. So Boyd's argument depends, indeed quite heavily,

upon the independence of *measurement*, despite being welcoming of theory-ladenness more broadly. Boyd's position cannot pair comfortably with the non-independence of measurement. And so my account can give comfort to Boyd's brand of realism, while rebutting allegations meant to undermine the independence of measurement.

2. On the logic of measurement

As Tal (2017a) observes, we have reached a new point in the discipline of epistemology of measurement. Although the threat to empiricism used to be theory-ladenness (the doctrine that theoretical assumptions are unavoidable in designing measurement apparatus and interpreting their readings, and just as unavoidable in observations made with the naked eye), "contemporary scholarship has come to view measurement as theory dependent *by default*" (Tal 2013, emphasis added). That is, the burden of proof is increasingly being shifted onto the shoulders of the defenders of the traditional conception of measurement.

Chang (2004), among many others, employs detailed historical case studies of measuring temperature to highlight a highly iterated process by which measurement practices are refined even as the theoretical understanding of temperature and its connections with other observables is advanced. These case studies, he maintains, support the notion of "epistemic loops" around measurement,¹ which in turn give way to a highly potent form of epistemic coherentism, whereby validation of measurement procedures is irreducibly infected by the contingent history of developing theories, to say nothing of the practical contexts in which purported success in measurement serves specific human interests. Building on these and related ideas, Tal (2017b; 2019) advances a "model-based" theory of measurement, according to which measurements can be viewed as successful only in relation to an (idealized) model of the measurement process being employed.

Both Chang and Tal invoke the contingency in scientific developments, in both measurement procedures and in scientific discovery more broadly – which is a historical fact, arguably metaphysical in character – to cast aspersions on the idea that there can be anything like a history-independent conceptualization of either measurement, as such, or the object-in-the-world measured, and via that route to throw doubt on every sort of foundationalism (mine among them) which would set measurement as independent of theory. The correct response to historical contingency, according to their brand of skepticism, is a coherentism whose main tenet is the notion that we retain today what coheres today with present standards, but modulated by the idea that the standards of today can constitute advances on those of yesterday. Chang refers to this as "progressive coherentism": this doctrine provides at once that the criteria for measurement success are internal to the practice of science – so scientific knowledge does not rest on an independent foundation – but at the same time that these internal criteria may be used to evaluate new practices as improvements, thereby allowing for scientific progress (in contrast to traditional coherentism – cf. Chang 2017a, 2017b). "In the context of measurement," writes Isaac (2019: 932), this should be interpreted to mean "that later measurement practices may be understood as in some sense 'better' than earlier ones, yet the 'epistemic achievement' involved should not be understood in

¹The term was not invented by Chang, but it was prominently utilized by him. Chang (2004) defines *epistemic iteration* as "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" (45).

terms of greater degree of correspondence to quantities in the world.” Therefore it is important to be mindful of how one’s notions today may not be consonant with the notions had by predecessors who might have used the very same terminology, and can constitute genuine progress over them.

Of course everyone should endorse caution in connection with attributing contemporary ideas or concepts to our predecessors, especially those who inhabit the distant past. But of course coherentism is *not simply* the indisputable doctrine that scientific ideas and scientific practices change over time; the view that history is contingent is compatible with the foundationalism that I am defending. Observing caution with respect to the history of ideas and/or measurement practices does not require forswearing the notion that scientific theorizing must be founded on an independent body of evidence, including measured evidence. Still, as Tal himself observes, while any view on an ontological or conceptual issue (for instance the status of a given quantity as independent in some way) might be compatible with any view on the epistemics associated with it (such as the reliability of a measurement process associated with the named quantity), the issues are routinely interconnected by philosophical arguments, and connected also with the history of science – so that, for example, coherentists tend to take a dim view of certain conceptualizations of measurement (as we have observed in connection with Chang and Tal), and argue for their ideas by reference to the historical record. My aim here is therefore also partly to sever some of the lines of argument from epistemics (e.g., reliability of measurement processes) to ontological status of the associated quantities (for instance as objects of measurement on certain scales). It is to show that one can argue about appropriateness of scales in a way that is in fact independent of the epistemics – and indeed that this is the most philosophically sophisticated construal of the classical RTM.

I begin with a most anodyne observation: we must observe the substantive distinction between two things:

- (1) the confirmation of a *specific magnitude assignment* to a measurable quantity on a specific occasion, which we can refer to as the *accuracy* of a specific measurement (say of the temperature of a room on a given occasion as 21 degrees C); and
- (2) the validation of that quantity (*air temperature*, or more broadly, *temperature*), as such – as a thing broadly susceptible to measurement on a specified scale – which I will refer to as *quantity certification*.

This distinction will be the wedge that will allow me to take up and carry the (temporarily shifted) burden of proof in connection with whether evidence is independent of theory, and thereby to redeem the traditional conception of measurement as (contrary to the coherentist view) establishing a relationship, not mediated by scientific theories, between empirical structures, perhaps secured with the help of measuring devices, and a mathematical structure such as a scale. In this tradition, measurement is the central and most important means of assigning a scale to a quantity.

The distinction I insist upon, between certification of quantities and confirmation of magnitudes, begins by anchoring it in the distinction between (on the one hand) a *quantity*, understood as a characteristic of the universe that may vary in magnitude with time, taking on no more than a single magnitude at a single moment in time, and (on the other) the *event* of a quantity taking on a definite magnitude. There is the length of object O, on the one hand, and on the other, that length taking on the

length of (say) 2.5 m; there is the temperature of object O, on the one hand, and on the other, that temperature taking on the value of (say) 98 degrees Fahrenheit.

Consider now a situation in which, by hypothesis, a certain class of things can be said to enjoy a variety of different *flavors* (which together form a *field F*, if you will, of flavor), but that no two of the identified flavors bear any intelligible relation to one another. For instance, it might be the case that the flavor “chicken” has nothing whatever in common, in their conception, with the flavor “beef” or “potato” or “onion” – that they cannot be compared to one another in any meaningful way once apprehended. All one can say about the elements of the field *F* is that they are different.

The best one could do, in such a situation, would simply be to categorize something bearing one of the relevant flavor attributes into its appropriate flavor bucket, and nothing more; let’s refer to this categorization scheme as a *bucket taxonomy*. When the best we can do, for this field *F*, is to devise a bucket taxonomy for it, it becomes clear that the field *F* has no structure. And while of some interest to those who would toil daily in this field (to achieve culinary goals, for instance), the bucket taxonomy of the field *F* would appear not to lend itself to significant theorizing. One could enumerate those buckets, if the inclination to do so urges: “1,” “2,” “3,” and so on. But that way of designating the buckets would have to come with a caution *not* to use the number labels to signify anything other than mere identity and difference – specifically, use of the number labels should not be taken to designate relations among the buckets outside of identity and difference. It should be forbidden to draw inference as to ordering or priority, for instance. And arithmetic operations on the number labels, for instance to attain a “sum,” “distance,” or “average,” has to be forbidden, because it is fundamentally meaningless. The significance of number labels has to be suitably circumscribed. Herman von Helmholtz cast the key question of assigning values to quantities as: “[W]hat is the objective meaning of expressing through denominate numbers the relations of real objects as magnitudes, and under what conditions can we do this?” (1887: 4)

By contrast, consider the familiar case of simple length measurement. Length, as such, is a quantity type; the length of a particular object is a token. Assignment of a magnitude of length to a selected object, on the basis of a measurement, requires a measurement *procedure*. Now, one may produce a meter stick and determine via a suitable application of it that a certain object O enjoys a length of 2.5 m. When, pursuant to this act, one says that “the length of this body is 2.5 m,” such a statement is meaningful only if the reality spoken of in it bears multiple logical relations to *other* statements of the form “O bears length *x*” – the “2.5” statement is meaningful because the number itself belongs to a structured field. Moreover, we would find a “2.5 m” assignment of length completely uninteresting if nothing else could be said to enjoy that same length, or some other length, longer or shorter. It’s in virtue of the fact that length is something many things enjoy, which renders it valuable to assign a length magnitude to some object. Length assignment becomes not merely a property, unique or not, but also a dimension along which other things too can be compared (cf. Swoyer 1987 and Mundy 1987).

The true difference between unconnected flavors, on the one hand, and lengths on the other, does not lie so much in the limited possibilities for theorizing upon unconnected flavors – for it might well be that the possibilities are sufficiently interesting for theorizing with an unstructured taxonomy. The difference lies rather with the manner in which the field is applied to the objects it organizes. For, with length there arises also a specifiable *operation* of measurement that comes into view – associated with the measuring device to which it is so cinematically linked: the measuring rod. And it is

this very rod, together with its numerous emulators, that carry the possibilities of scope coiled within. And thus is born the notion of “measuring operation” that is so important in the theory of measurement.²

It’s important to be clear here that the concept of a measurement *operation* – the concept that becomes one of the two foci of the mathematical theory of measurement (with the second being the mathematical objects, normally, scales) that are associated with the operation – is in no way a concept of measurement *procedure* or measurement *device*; the theory of measurement devices belongs elsewhere. A measurement operation, according to the conception we will be using, like everything in the mathematical theory of measurement, is an abstract, logical/mathematical object or set of objects, about which axioms are postulated, and concerning the “outcomes” of which the question of how best to conceptualize their representation becomes the central question of the theory.

3. Metaphysics: the theory of measurement

How is measurement of a quantity (e.g., length) different from measurement of the unstructured field F of flavors we discussed above? The answer might seem straightforward: it is because the field being subjected to measurement has the relevant mathematical structure – for instance, it is a structure where elements in it enjoy a mathematical ordering relation of less-than/greater-than, together with the *additivity* of that relation. Mathematical relations can therefore *represent* the shorter-than/longer-than relation. It is because of these mathematical relations that the field of lengths has a more substantial structure than the unstructured field of flavors. And it is this structure that measurement invokes.

The **fundamental problem of measurement**, in connection with a given quantity, is therefore the determination of the appropriate kind and amount of (mathematical) structure that is required to model measurement of that quantity. For one should expect that an arbitrary pairing of a quantity with a mathematical object will result in a mismatch of amount and type of structure in the representing tool – with the representational power of the instrument frequently outstripping that in the target quantity. A simple example to demonstrate the point: there is a great deal of structure, say, in a three-dimensional mathematical space – too much for the measurement of rods. Ascertaining the correct amount (as well as quality) of structure is the fundamental problem of measurement theory.³

²There is some controversy over whether a procedure that “operationalizes” our bucket taxonomy of flavors does not also deserve calling a true operation and its associated scale a true measurement scale. A permissive or liberal stance on the question (like that of Stevens 1935) might not take us too far off course, but since “operationalization” will be an important feature of the thesis I am developing, I am afraid I have to come down closer to the opposite side of the permissiveness continuum on which Campbell (1920) sits, thereby denying a bucket taxonomy the status of measurement. My position is therefore somewhat closer to the strict operationalism of Campbell (1920), who insisted on measurement being closely associated with the assignment of numbers, than to that of Stevens (1935); it is strictly somewhere in between the two positions, endorsing the idea that measurement is to be thought of as associating a mathematical structure of some kind (not necessarily numbers) to an operation. Buckets, according to this conception, are insufficiently operationalizable. This allows for a distinction between measurements, as such, and a more free-wheeling thing we might refer to as “conceptualization.”

³Batitsky (1998) takes the fundamental problem to which RTM seeks a solution to be one of reduction: how to think about the assignment of quantitative values to something like counting, and launches

One may, as for example Michell (2005: 292), largely sidestep the logic of asking this question formally, and instead pose it empirically:

Measurement is always based upon the hypothesis that some attribute is quantitative. This hypothesis, like any in science, says that certain empirical conditions obtain and it, thereby, rules out other possibilities. The scientific method of critical inquiry, according to which hypotheses are only accepted following serious attempts to put them to the test and given evidence in their favour, applies to this hypothesis as much as to any.

I agree with the spirit of the idea that the modeling of a quantity by a specific mathematical structure is, logically speaking, a hypothesis. But one has to insist that testing such a hypothesis must proceed by way of an appreciation of the nature of excess mathematical structure, and thus by way of appreciation of the body of theory devoted to this task. Yes, each such hypothesis, vis-à-vis a target quantity, requires testing at the bar of observation. But one requires an entire theory of mathematical structure in order to appreciate how that testing must proceed. The mathematical theory of measurement *is* that theory. It is a theory that explains, and in that process justifies, the choice of a specified amount of mathematical structure. And execution of that explanation proceeds orthogonally to the execution of the testing of any scientific theory in which the target quantity might figure.

The most influential philosophical account of measurement today – known as the representational theory of measurement (RTM), originating largely with Patrick Suppes and collaborators in the 1950s (Krantz *et al.* 1971; Luce *et al.* 1990; Suppes *et al.* 1989), defines measurement as the construction of mappings from empirical relations (longer-/shorter-than) into mathematical/numerical relations or structures. I propose, and accordingly shall argue at further length below, that we must add to this definition – we must add that the mapping for any given quantity (from empirical relations to mathematical structures) has to come with instructions detailing, in mathematical terms, the *amount and type of structure* in the mathematical objects being pressed into service of representing the quantity itself. This, it must be understood, is in service of a *faithful rendering of the amount of structure in the world (in this case the measurable quantity)*. My distinctive contribution here, to the philosophy of measurement, is to insist that appreciation of the precise amount of mathematical structure needed to represent a quantity is fundamentally what the mathematical theory of measurement is about: the fundamental aim of the theory of measurement is to provide an account of exactly how much is being asserted about the logical structure of the space of possible outcomes of measurement of the feature being measured. The theory of measurement is concerned with the structure that a collection of measurement outcomes directs attention to, but does not make direct claims of realism.⁴

polemics against RTM from that controversial basis. Still, his counterproposal (Domotor and Batitsky 2010) is a worthy account of the realism of quantities – treating some but not all the questions one might wish answered by such a theory. For example, it is incapable of raising and answering the question that I take (see text) to be central to the theory of measurement: how much structure is in the quantity we seek to represent, and how much excess structure might a given scale mis-attribute to it?

⁴The structure spoken of resides in the reality of the quantity, according to some theorists (Byerly and Lazara 1973; Domotor and Batitsky 2010; Swoyer 1987); but it is nowise my intention to be advancing an argument for that sort of strong (structural) realism here; my view is that realism about quantities is

Recall the hypothetical field of flavors F we discussed before. We said then that we could enumerate the resulting buckets if we fancied doing so: 1, 2, 3, and so on. But we also noted that doing so must always come with a caveat: a caution *not* to use the numerals to signify anything other than mere identity and difference, since arbitrary (conceptually unconnected) symbols would work equally well. The significance of those particular labels has to be suitably circumscribed. And this is what the theory of scales does, as we shall soon discuss.

The point I shall be continually stressing is that there can exist features of a mathematical object, useful for representation purposes, that go above and beyond the features associated just with the labels themselves (numbers, or vectors, or what-have-you). For example, the labels “boy” and “girl” might be used to indicate not merely contraries, but also (if the user of the label is so inclined) opposites. There is a great deal of mathematics and mathematical structures one may press into service these days (as more mathematics is devised/discovered), and a great deal of it is eligible for use in a representational capacity. The task for any representation theorist, however, is to fit the complexity of the representation to match precisely and in all the ways that matter for the purposes at hand, the complexity of the object being represented. For example, if one is measuring a length, one should employ a mathematical entity that has sufficient complexity to handle appropriately all the possible ways that the measurement process of length might result in – and no more. We require a correct fit between the entity/feature/reality (an exact fit with the totality of possibilities/complexities/structures) and the entity representing it, in order to declare measurement victory. Indeed, much theory in mathematics is itself devoted to describing how to handle excess mathematical structure when it’s unavoidable – discussions of subjects treated under the labels of *symmetry*, *supersélection*, and *gauge* – when utilizing standard mathematical structures to represent features of the world in spacetime. It is important, as these concepts make clear, to keep track of how many differences in a description are “real” differences. For representation purposes, we require an entity that is capable of depicting this reality correctly. It is no small thing to get that right. And when a perfect fit is impossible, one has to do what one can to avoid misuse of excess structure.

It has become fashionable to strike, with Eugene Wigner (1960), a posture of awe and fascination at “The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics [--it] is a wonderful gift which we neither understand nor deserve.” But there is no “gift” about it: it is *work* to fit the right mathematics to quantities in the world, including measurement processes that arrive at magnitudes for quantities (cf. Wilson 2008). One might equally appropriately wonder about the miracle of the appropriateness of clay or stone for the shaping of statues, as wonderful gifts that we neither understand nor deserve. (Perhaps one also believes that sausages are miracles we do not understand or deserve if one has never seen one made.) But the right way of viewing the matter is to appreciate that it takes hard work to find the appropriate structure; every choice stands in need of justification, and no justification comes for free. The world of mathematics is full of structure, the vast preponderance of it inappropriate for the representation of quantities, and perhaps much of it inappropriate for the study and analysis of any phenomenon in the world.

Proofs of the appropriateness of a structure or structures to representing quantities came along altogether late in the game of science, but they are (some of them, anyway)

unnecessarily strong – realism should be, more modestly, about the *possibilities for measurement outcomes* in connection with measurements of that quantity (the field, as we discussed above).

here now. It is well worth the philosopher's while to let them do such work as they can in connection with some of the most notorious philosophical doctrines in the philosophy of science.

4. The theory of measurement, especially scales

Before proceeding further, we must take note that RTM itself has recently come under fire, in spite of its standing in mathematics, and Wigner's adulations of mathematics as gift of which we are undeserving. But before we can address the critics, we will do well to hold firmly in mind the reality that RTM is proclaiming as being represented. What problem is it purporting to solve? Or, to put it another way, if RTM is the answer, what is the question?

An initial attempt might be this: Measurement theory answers the question of how mathematical objects can represent structures antecedently present in quantities. This blunt attempt simply assumes that quantities are antecedently structured in the way that the ultimately selected mathematical objects also are. This is not the case, for instance, when we allow numbers to be utilized in representing the outcomes of flavor measurements, as in the hypothetical example we have been discussing. The simple point is that there might be much more structure available in the mathematical objects doing the representing than there is in the features being represented. So the true objective of measurement theory is to address the question of how ensure an *appropriate fit* between represented quantity and representing (mathematical) object. A more measured response to the question of the previous paragraph is therefore that, in order to ensure the appropriate such fit, RTM seeks to associate measurement with an empirical process whose outcomes can be structured in a way that is suitably modeled by the mathematical objects selected. This readily makes sense of some of the language measurement theory's proponents employ, frequently using the terms "empirical structures" and "operations." But as critics point out, nothing in the mathematical theory of measurement focuses on the realities of putting the empirical world under a microscope or any processes of measuring anything or constructing measuring devices for measurement (Boumans 2007; Decoene *et al.* 1995; Michell 1990; 1995; Reiss 2008). In the critics' view, the analysis of measurement in measurement theory is simply unrealistic, giving no attention whatever to real-time processes, historical contingencies and prospects for error.

The critics are correct in pointing out that in some sense there is a lack of realism. But this is because the "empirical structures" and "operations" that measurement theorists have in view are *not* the messy things of the world. They are, instead, features of the logic of an array of measurement *prospects* for which a mathematical correlate is sought (perhaps this is what Heilman (2015) means by "concept"). For example, the logic associated with measurement of length is what is referred to as a "concatenation operation." This is the manifestly nonphysical, nonrealistic operation of laying down identical rigid rods end to end in absolutely straight lines, without leaving a gap between. No considerations are given to our fortunes and fates in the labors of measurement: of leaving gaps or falling prey to overlaps, rod misalignments or mishaps in rod counting. All those details are irrelevant to the logic of measuring length, which is the true target of the analysis. More relevant is, for example, the question of the *dimensionality* of the array of measurement prospects; do we need to represent the possible measurement results in a two-dimensional array, or is one sufficient? (If the answer is one-dimensional array, then an operation involving rods laid end to end will do the trick.)

The idea behind “operation” and “empirical structure” is to capture the most general, logical features of the phenomenon, so as to answer the question of an appropriate fit between quantity (in the world) and mathematical object doing the representing (modeling the world); it is *not* to acknowledge specifics and epistemological vicissitudes in the real-time labor of assigning a magnitude. It is to distinguish, for instance, whether the operation belongs to the category of so-called *extensive* measurement, distinguished by the principle that addition of matter or substance or what-have-you in the measurement process results in an “additive” difference. It does in the case of concatenation: adding more of the subject to be measured results in more length in a specific (“additive”) way, whereas in the case of so-called *intensive* measurement (of the likes of temperature or density) it may or may not. Length is *additive*, and therefore can be represented in a particular way, because of the properties of *concatenation* – this is the sort of thing that measurement theory can prove. It can prove that a measurement operation that is like putting rods end to end deserves representation by a particular sort of scale.

If we accept all this, we really are giving a *representational* interpretation of the RTM. We are speaking of quantities in the world being subjected to modeling as a result of their logical properties – properties that are features also of mathematical objects. Like Krantz *et al.* (1971: 9) write, we are regarding measurement “as the construction of homomorphisms from empirical relational structures of interest into numerical relational structures that are useful” – on the basis of enjoying certain properties in common. This is the received view of measurement, to which I take myself to be giving voice. Unlike the notion that RTM is to be interpreted as directing comments only at concepts (as Heilman (2015) avers), I take the view that these comments make sense only in the context of an attempt to model quantities in the world.

More questions arise about representation once we delve into further complexities regarding representation. We know, for instance, that there is “more” and “less” when it comes to both temperature and density. Do they therefore warrant being put on the same sort of scale as length, or must they remain purely a measurement of more or less – a measurement of ordering purely? What do we lose by making a choice? We could of course use numbers, as practice might dictate. But without further guidelines for deciding, it will be a matter of controversy whether doing so imputes too much structure to the phenomenon. The hope is that, once we have a formal proposal for how to model the measurement operation, as such, we should be able to prove that a certain structure is appropriate.

That model, available for a larger range of quantities, is a more recent mathematical innovation: the theory of *additive conjoint measurement* (Krantz *et al.* 1971: 17–21 and chs. 6–7; Luce and Tukey 1964; and the lucid Narens and Luce 1986). Without going into unnecessary detail, the strategy for demonstrating the appropriateness of a number-line representation in such cases is to prove that it is appropriate to attribute to the quantity Q in question an effectively unrestricted ability to order or compare states of it through measurement. This proof is possible when it is available to triangulate comparisons of levels of Q, as indexed by levels of two other quantities (say, R and S), with which Q is known to vary. Axioms regarding their comparison with levels of Q are written down (instances of which are subject to confirmation through observation). And from the axioms Luce and Tukey (1964) first proved the condition called “additivity.” As Narens and Luce (1986) write: “The two [other quantities, R and S] are, in some sense, measures of the two components of the attribute, and Q is the rule that describes how these measures trade off in measuring the attribute [Q].” We might say that R and S are “levers” for shifting Q, as for example pressure and volume are levers on temperature.

Later work showed that the conjoint strategy works because it reduces mathematically (if not empirically) to extensive measurement.⁵ Since then, many variants of extensive and additive conjoint measurement have been used by scientists in a number of fields, from physics to psychology and beyond, to certify the measurement of a variety of quantities. (We will take up discussion of psychological quantities presently.) The theory of measurement has made numerous theoretical advances since then (Narens and Luce (1986) provide a very lucid history up its point of publication), and culminating especially in the theory of scales.

Before moving to that topic, I want to emphasize the philosophical significance of what we have reviewed here so far. The example of additive conjoint measurement makes it crystal clear that measurement theory is not directing attention to measurement processes in the messy world, so much as to *models* for establishing the extent of the possibility of ordering states of some quantity Q whose status as a quantity is in question. This is less about the objective reality of the quantity as such, and more about a field of measurement outcomes of the associated quantity and the ability to organize/structure them in an appropriate way. Facts about this field, as we noted before, might turn out to be very widely repeatable, for instance with other objects; when that's the case, as turns out to be so in the case of length, a length L has the same meaning whether applied to rope, train tracks, or wing span.

Neither concatenation structures nor conjoint measurement structures should be construed as referring to measuring specific instances of the quantities in question, as such, or about various real-time processes associated with ascertaining such features as spatial extent. Operations represent the logic of our means of such ascertainment – its logic-metaphysics – and not its epistemology. Concatenation operations and conjoint measurement models of extensive quantities are *models*, in the philosophical sense, of what success in measuring length or temperature or anything else looks like. The theory of measurement is about nothing so much as about the very success in measurement, and the mathematical structures that are appropriate to its representation. Thus the theory of measurement is more like a *normative* theory than it is a descriptive one. It's no wonder, therefore, that no attention is given over in it to possibilities of measurement gone wrong.

One might say that the culmination of the theory of measurement is a theory of scales. In 1946 the psychologist Stevens published in the journal *Science* a typology of scales:

The classification of **scale** (Stevens 1946)

- (1) **Nominal** (color, sex, particle spin; the bucket taxonomy where labels are essentially meaningless).
- (2) **Ordinal** (orderings, rankings; where the labels are only meaningful for the order they inherently provide: I, II, III, ...; 7, 8, 9...; 1st, 2nd, 3rd...).

⁵One can do little better than to quote Narens and Luce (1986: 170–71): “So, in sum, conditions that are sufficient to construct an additive representation of a conjoint structure are: weak ordering, independence, the Thomsen condition, unrestricted solvability, and the Archimedean property. What Krantz (1964) and Holman (1971) did was to show that, despite the fact that there is no empirical operation visible in an additive conjoint structure, the trade-off formulated in that structure can be recast as an equivalent associative mathematical operation. This allowed the earlier representation theorem for extensive structures to be used to prove the existence of an additive conjoint representation. This construction is such that it can actually be mimicked in practice by constructing standard sequences and using these to approximate, within a specified error, the desired measure.”

- (3) **Interval** (single-dimensional scale of numerical assignments where the most meaningful unit is the interval; the zero of the scale is fixed only by a convention).
- (4) **Ratio** (length, mass, volume; special case of interval, where the zero is independently meaningful).

Stevens listed scale types more or less intuitively, drawing attention to the fact that numbers could be used equally well in all the cases, but their meaning in each case is entirely different. The intuitive list leaves open the question whether there aren't more categories of scale. Maybe this four-item list is an exhaustive list. But if it is, how can we know? What is the proof? This is a question of metaphysics, since it is concerned less with the labels (in every case numbers, real and/or natural) and more with mathematical features that are separately and independently manifested in the numbers/number line.

It turns out that mathematics can itself address the question – mathematics can answer matters of metaphysics, when the relevant concepts are sufficiently formally articulated.

In the 1970s and 1980s Duncan Luce and Louis Narens worked out a representational theory of the question that needs answering, in the context of larger questions about measurement and the modeling of empirical operations representing measurement that we have been discussing. Narens and Luce write (1986: 169):

Although [Stevens] recognized more than anyone else at the time the significance of scale type in contrast to the particular structures exhibiting it, he seemed not to appreciate that, in fact, the concept of scale type is a theoretical one that can only be formulated precisely in terms of an explicit axiomatic model of an empirical process. He failed to acknowledge that it takes more than one's intuitions to establish that a measurement process, such as magnitude estimation, leads to a ratio scale.

Put slightly differently, Stevens's list, without a theory, amounts merely to intuitions that, in their turn, require explaining. Although Narens and Luce do not put it quite this way, their point is that a scale type is not only a way of representing facts across a wide variety of settings and objects – so that the metric has semantic meaning. But it also carries syntax or metaphysical structure that is representable and susceptible of being theorized about mathematically.

Finally, Luce and Narens's answer to the question is that the choice of scale does indeed have to be commensurate (pardon the pun) with a measurement *process* that has a certain acknowledged set of possible outcomes, but that the process does not really exhaust the meaning of the model; the model is as theoretical an object as is any model, and has to be studied and employed accordingly. In still other words, scale type is a theoretical conception, and as such is due to a certain amount of theorizing. A theory of scale type (appropriately also mathematical/metaphysical) gives that conception its due. As the present study is not focused upon scale type specifically, I will not tarry on this point longer, as interesting and important as it is.

5. Some caveats

The view I am setting down here has many affinities with an old-fashioned *operationalism* regarding meaning of quantities, at least in its emphasis on empirical operations. I wish to emphasize however that mine is not a view about the *meaning* of quantities; it is

instead a view about the appropriateness of their measurement-on-a-particular-scale. I am referring to this as *certification*. And its role is to provide a philosophical anchor for an FE and the special status it accords to observation and measurement.

Another point to emphasize is that the account here accords a special status to quantities that are measurable by an empirical operation or via a process of conjoint measurement. By contrast, it does not accord that same status to quantities to which the only access is via a computation through a model, or that must be estimated from a model or a statistical distribution. Thus this account respects the traditional distinction between direct and indirect measurement, and accords special status only to direct measurement.

The measurements we therefore can speak of as guarantors of a quantity's measurability stand in stark contrast to the estimations of quantities or parameters made using a scientific model, in whole or in part. We do not deny the status of quantities or parameters that require estimation; we merely cannot undertake to guarantee that status through the quasi-empirical methods of RTM.

6. A philosophical theory of measurement

My position here regards what we ought to mean by certification of a quantity Q. Success in measurement of a quantity, as such, should never be issued as a blanket statement, covering all assignments of magnitude made in the name of science. It should, instead, follow this schema:

Schema M

- (1) There are a variety of scale types available (nominal, ordinal, interval, ratio).
- (2) The overwhelming preponderance of experience/observation of measurement of Q are a best fit with the mathematical modeling assumptions of scale S.⁶

Therefore, Q is fittingly measured/measurable on a scale of type S.

When the status of Q, in our measurement enterprise, can be summarized this way, I propose we say that it is *certifiable*.

This (philosophical) account of certifiable measurement is not free of all theory – but it *is* free of what might be properly termed *scientific theorizing of the facts*. In other words, it is free of what we may call *scientific substance* that features Q.

Certainly (1) is very much *not* an observational statement. Equally, it contains nothing of what may rightly be called scientific substance. (2), by contrast, can be purely observational. But now one can insist: if observation, as such, is theory-laden, why isn't certification? *The answer lies in the logic of premise (2) above. It is not a statement about the accuracy of any measurement of Q; it is a statement about the logic of the array of outcomes. Thus it concerns the logic/metaphysics of the quantity for which a case for certification is being mounted, rather than any concerns around assignment of specific magnitudes of it on specific occasions.*

Still, one must not oversell the case. It is quite possible that our confidence in (2), vis-à-vis some specific Q, might be premised largely on a theory. For example, taking the case of the less familiar sort of certification of a conjoint measurement operation, it might well be the case that confidence in our ability to order states indexed by

⁶If we are using a conjoint theory of measurement, we must add: taken together with observations of associated quantities that function as “levers.”

pairs of states of quantities R and S, viewed as levers, might be premised largely on confidence in, or commitment to, a theory or some other of the way the levers are operating, rather than on a purely observational basis. It might be, in other words, that we have no theory-independent way of indexing Q itself. This is not out of the question. In these cases, direct certification will fail, and we might therefore propose an associated notion of *indirect certifiability*, which should be construed as related to *direct certifiability* exactly as direct measurement is related to indirect measurement. But these failures are manifestly absent in the case where we use direct means such as measuring rods or thermometers. Hence the case *against* certifiability of quantities should not be exaggerated, the historical contingencies of measurement processes notwithstanding. Temperature, contrary to some opinions, is decidedly accessible and therefore certifiable. This is the question to which we turn next.

7. Temptations of contingency

The effect of my theory of quantity certification is to provide a sound philosophical basis for assertions that certain scales and scale types are appropriate, on a case-by-case basis, to measurement of specified quantities, and thereby certified for that purpose. Such assertions are (at least in the central and high-profile quantities favored by coherentists) premised on concrete bodies of observation. I claim that this philosophical account is one that can be a fitting pedestal or foundation upon which the balance of the edifice of associated scientific theorizing may be said to rest. It is, as I've argued, a thing independent of the associated scientific theory, but *not* of mathematical–philosophical theory, or of observation simpliciter. It can thus serve as a defense of FE. Let's turn now to how this position addresses specific coherentist challenges.

Temperature has served as a lightning rod for arguments around measurement, and foundationalism generally, for some time. The quantity is palpably real to scientists and lay persons alike, but also, by contrast with something like length, has enjoyed a checkered history in its measurement. Moreover, it is now measured on numerous scales and by multiple device types that operate on a variety of quite diverse principles. Finally, it has been embroiled in philosophical controversies around the metaphysics (including the reducibility) of quantities in classical and statistical thermodynamics, including the truth of the so-called Second Law. It is thus no wonder that temperature lends itself so easily to skeptical argumentation.

Chang (2004) recounts how the history of measurement of temperature is tangled up with the history of theorizing about temperature, with the effect that we have devised the very concept of temperature in the very process of measuring it over a certain historical interval. Tal (2017b), building on some of the historical material Chang (2004) discusses, maintains that the very process of working out a theory of the nature of temperature (the theory of thermodynamics) is implicated in constructing the many means of measuring temperature that scientists came ultimately to employ. This is apparently evidenced, for instance, by the need at certain points in history to make assumptions – best to call them decisions, they say, if we're honest. There is for instance the prominent example of the relationship between melting point and freezing point of any given substance, namely whether these should be construed as always and everywhere the same and equal to one another (the question of their so-called *fixity*). The fixity of melting point/freezing point was, according to Chang, simply assumed in the effort to coordinate measurement at certain points in history. The most forceful statement of this critique might be stated as follows: since we measure as we do because

our predecessors assumed what they assumed (or rather, decided), all of temperature measurement is equally contaminated by theory, and to that extent compromised – i.e., literally, un-founded, if a foundation of pure observation is the only one allowable as a true foundation. The term “loopiness” has sometimes been applied in connection with a quantity like temperature.

Chang is comfortable with the “loopiness” as an important feature of the process of stabilizing our activities of measurement, arguing that the iterations in the process in which theoretical advancement is intertwined with standardization of measurement of temperature, is in fact virtuous, so long as the “epistemic iterations” in the history of standardization respects foregoing traditions, while at the same time aiming at correcting perceived shortfalls (Chang 2004: ch. 5). While pre-theoretical construals of temperature aimed in haphazard ways at ordinal rankings of things from cold to increasingly hot, this effort was honored when more quantitative attempts came along – from thermoscopes to early thermometers to better thermometers utilizing different principles in service of better triangulation. Coordinations and convergences of these efforts progressed toward every more stable standards of measurement, and with them more stable theoretical constructs, virtuously moving the process toward ever more stability. Stability, he suggests, signals measurement success on a coherentist basis. But does it?

Isaac (2019), a critic of this coherentist position who advances what he calls “fixed point realism,” argues that the variety of increases in *precision* of measurement (for instance through technological innovation) – while confirmatory of success generally – also points to stouter and more durable support for measurement as more firmly rooted in the quantitative realities of the world.

My own response to Chang’s observations regarding the development of the growth in *theory* of temperature (i.e., thermodynamics) as developing with and alongside a growing understanding of *how to measure* temperature is that they may well all be true. But I would nonetheless insist that, even so, nothing in the parallel stories has bearing on fundamental questions about the case for certification of temperature *as measurable* (and even *directly* measurable) on a given scale. First of all, the history of a theory is routinely mixed up with developments in the certifiability of many of its quantities all at once; and so one will always encounter elements of both in a history of the surrounding science. (This is another confirmation of my account: historical developments in the certification or certifiability of a quantity, especially the emergence in scientists’ consciousness of facts that render a quantity certifiable, can come on the historical stage either much earlier or much later than the quantity whose credentials are in question.) Certification is a *logical* relation holding among the quantity, the mathematical scale proposed for it, and a body of observations in connection with that certification: the body of observations is the evidential basis for certifying a relationship of fit between the quantity in question and the proposed scale. It is rare in the history of the employment of a quantity, that the scientific community focuses explicitly on the question of its certification; much more prominent is the question of when, in the history of the subject, a scale is simply adopted and presumed appropriate. The logic of certification is apparent more in the rear view mirror than in the moment of adoption.

More urgently, Chang (2004: 59) speaks of “the problem of nomic measurement” – the problem of choosing a standard for measuring (for example) temperature, via a procedure-cum-appropriate-device, when doing so seems to require invoking a theory of the device – or worse, the phenomenon of which it takes advantage. Normally there

are choices to be made among alternative measuring devices, which likely (or so the presumption goes) would result in different judgments as to whether the same objects/events are deemed equal-temperature. But – he implores us – isn’t appealing to a theory (or worse, nothing at all) to make a choice tantamount to a vicious circularity? It must be circular, since there is no way to judge that it is correct *before* employing a measuring device. For instance, should we pick the sealed column of mercury, or alcohol instead? And can we pick before we know just how much the substance expands upon being heated? Either way, how do we know where to put markings such “10” and “50” – assuming “0” and “100” are selected on the basis of the boiling point and freezing points of (again) an arbitrarily chosen fluid?

The most forceful statement of the coherentist position can be quite seductive⁷: it is tempting to believe that, because a decision, or even a series of them, had to be made (and no way around it), different decisions could have been made in an alternate scenario, with the consequence that our theories-cum-standards of temperature would have turned out totally differently from what they currently are.

But this fantasy of contingency is an illusion. My contention is that the theoretical outcome would have been the same no matter which, of the available choices vis-à-vis temperature measurement discussed by Chang, were made. So for example, imagine, contrary to fact, that someone had the brilliant idea to inscribe a logarithmic scale between freezing point and boiling point of some arbitrary thermometric fluid encased in a closed tube – orange juice, let’s suppose. Would that have resulted in, for example, a different ideal gas law? Certainly not. Assuming nothing else is different in connection with volume and pressure measurements across the actual world and this hypothetical scenario, the only difference would be that the quantity referred to as T by this hypothetical scientific community, we could refer to it as $\log T$, and accordingly translate their pronouncements into the true pronouncements in our own language. Nature remains in the driver seat, nature calls the shots. Our power is only over labels.

By the same token, suppose we did NOT equate melting point of orange juice with its freezing point – suppose we did not take that “decision.” Would our thermodynamic laws be different? Certainly not. We would soon have *discovered* that the identity of freezing point and melting point (in those cases where in fact they are). We would have learned that it is true in water (for example), and that would have suggested the principle as true for orange juice as well. And for precisely the reasons that Chang himself acknowledges: because there is constant “looping” checks that are performed: decisions are made, and then further measurements are performed to substantiate that nothing has been “destabilized” by them. The community is always looking over each others’ shoulders, and in the rear view mirror, to double-check, once again, that nature is in the driver seat.

Someone who maintains otherwise only affirms that we are somehow creating the laws of nature themselves – not merely our concepts and tools – as we adopt measurement standards. Chang does not believe we are creating the laws – indeed he confesses: “it is always by favour of Nature that one knows something” (Chang 2004: 50). The question is naturally how to characterize that knowledge. While FE proclaims that knowledge rests on an independent basis of observation and measurement, the coherent view, however temporized, removes all proposed bases as privileged. But removing basis removes the basis for the coherentist position itself; it is to insist – without a basis – that our decisions on any matter, including on matters of measurement, will invariably,

⁷This might not be Chang’s own view, though it might be that of Tal (2017b).

irretrievably, and substantively be formative of theory, and to foreclose (again without basis) that we can ever learn something to gainsay assumption and/or decision. To affirm this is to throw into doubt, beyond hope of redemption, that inquiry can aim at discerning facts on the ground.

There is perhaps a more profound worry. There is the worry that, if you selected orange juice as your fluid in a sealed tube, and orange juice behaves sufficiently differently from (for example) air or mercury, your measurements with an OJ thermometer will not really be of temperature – or anyway, not of temperature as we measure it now. It should instead be referred to as measurement of “twemperature.” If, when one subsequently constructs a mercury thermometer, and finds systematic differences between the OJ thermometer readings and those with a mercury thermometer, it will be an unanswerable question at that point whether the two thermometers are measuring the same thing, albeit one with systematic errors. Thus our choices around measurement cannot settle what they are measurements of. This is what Tal (2019) refers to as the problem of “individuating” quantities. While Chang might be optimistic that a sufficiency of “loops” will result in a resolution of genuinely different quantities, Tal believes that the problem is ultimately not subject to empirical resolution, that ultimately when two measuring processes disagree there can be no evidential grounds for answering the question of whether they are measuring the same quantity, and so that we have to settle for a “relativist” solution – a solution of choices relative to theories or models of the measurement processes we ultimately employ. This is the extreme coherentist position.

My response to this worry is much the same as before. You look at the two tables of measurements, one with an OJ thermometer and a second with a mercury thermometer. Are they measurements of temperature, with one thermometer systematically in error? Or are they instead measurements of different quantities? I have not conducted this experiment, but there is something one can say without having done so. Suppose that there are three objects to be measured for temperature. And suppose that the three objects are ordered differently by the two thermometric processes – for example, that process OJ orders them O_1 , O_2 , and O_3 , in ascending order, but that process mercury orders them O_2 , O_1 , O_3 , in ascending order. If this is a systematic finding (multiply repeated, the results stay the same), these data are sufficient to conclude that (at least) one of them is NOT measuring temperature. This result follows directly from the principles of conjoint measurement (to be discussed at further length below). If, by contrast, the two thermometric processes (levered by the same secondary quantities) result in the same ordering, even if nonlinearly spaced, we are back to the situation analogous to the former example, where the two processes are referring to the same quantity (temperature) by different labels (e.g., “T” and “log T”).

Tal (2019) discusses this type of quandary in the context of the history of temperature measurement. It was well known by the late 1700s that sealed tubes filled with different fluids expand at different rates when heated. It became paramount to determine which fluid expands most regularly (i.e., linearly) with “temperature.” This research eventuated in work of Henry Regnault in the 1840s, which prompted the adoption of the air thermometer as the standard. Tal (2019) writes that Regnault had to assume, without any evidence, that all the thermometers he studied were measuring temperature. Indeed, he says that there was no choice but to make this assumption – because no evidence is ever available to establish the fact. “That is not to say that Regnault could not test whether his instruments were thermometers,” writes Tal. “*At a crude level* [my emphasis], he could show that all the instruments responded in expected

ways to being heated, e.g. by exhibiting an increase in volume or pressure. However, showing that an instrument measures a quantity at a given level of accuracy does not entail that the instrument also measures that quantity at a higher level of accuracy.”

I confess to being at a loss as to how to understand Tal’s argument. The question is whether there can ever be evidence that settles whether there is reference to temperature rather than something else – evidence determinative of whether two processes are measuring the same quantity. Here Tal seems to be agreeing that there can be such evidence (“... not to say that Regnault could not test whether his instruments were thermometers”), albeit not always “at a given level of accuracy.” Still, he concludes that there is never sufficient evidence to say that two thermometers are measuring the same thing – temperature. He does say that there is evidence “at a crude level.” But this is not the same thing as saying there is not evidence enough to settle the question. Furthermore, what is so crude about this evidence – what makes this (or any) evidence crude? To my mind, there is nothing crude about evidence that settles whether the processes are in fact responding to self-same temperature. The theory of conjoint measurement gives insight into why that evidence is actually not crude.

8. How certifying temperature works

The earlier point about the historical appearance of a certification for a quantity is less true in the history of the discipline of psychology, where so-called *construct validity* is very much front and center when a quantity is introduced. This is because psychology and allied social sciences have been very much under pressure to defend the legitimacy (certifiability) of their quantities as measurables. In these sciences, certification of the measurement of their quantities was (and indeed continues to be) very much a matter for disputation at the time of theory development. It is not surprising, therefore, that these sciences distinguish between *validity* for their so-called *constructs* (which corresponds to what I call certifiability) and *reliability*. Reliability is, at least in part, what philosophers study under the label of “epistemology,” concerned as it is with the correctness, on numerous occasions of measurement, of the magnitude assignments made on those occasions. Validity, in social sciences, is another matter entirely – as scientists in the field well recognize. It is this distinction that my account of certifiability seeks to make philosophically more visible.

So let’s follow the outlines of a case for certification of temperature, to confirm that it does not, in fact, rely on any implicit appeals to theory. Let’s, for purposes here, assume that measurement of temperature is best certified through an account of additive conjoint measurement. To be clear, this is a point I draw from measurement theory in connection with temperature, taken simply as read for purposes here. Nothing depends on this being true. (Still, this is a reasonable assertion, although it is also reasonable to view it as properly extensive; either way, a certification of the conjoint measurement process will do the trick for certification purposes, in light of the mathematical reduction result discussed by Narens and Luce 1986: 170–71.) The aim of conjoint measurement is to describe formal conditions (which we will name below) under which a quantity (for instance temperature) is appropriately quantified on a cardinal (ratio or ordinal) scale, on the basis of purely ordinal data.

Thus the certification of temperature through an additive conjoint methodology will proceed, per the theory of measurement, via identifying two other quantities – for purposes of argument let’s appoint pressure (P) and volume (V) as appropriate levers on temperature, at least in cases where the substance in question is a gas. The observations

that will certify measurement of temperature (of gases in this case) will, again according to the theory of measurement, involve ordering, in a prescribed fashion, paired observations (p_i, v_j) over a range $i = 1, \dots, n; j = 1, \dots, m$ (the details of which are not important for the argument here). For this, we require actual measurements, and so the conducting of experiments. These experiments are meant to confirm that the relevant pairs obey certain axioms (according to one certification scheme, *single and double cancellation*, according to another case *solvability and Archimedean axioms*, to establish temperature as continuous⁸). The details of these conditions are not important for our purposes (Sherry (2011) offers more detail); we need to only note that they are thoroughly independent of any substantive theory of the phenomenon – of any thermodynamic principles, and of any assumptions about the functional relations between the quantities at play. Indeed, one must recognize that this must be the case if indeed they are schemes that apply over a wide range of possible phenomena, from thermodynamics to psychology to sociology and beyond – and they are indeed invoked over this broad spectrum of phenomena.

Now, assuming results are confirmatory when the prescribed measurements are conducted, the conclusion can subsequently be drawn that temperature is a quantity (i.e., appropriately measured on a cardinal scale), and accordingly deserving of representation on such a scale. Nowhere in the case to be made for the certification of measurement do we need to develop, or adjust, a theory of temperature, pressure, and volume. We need only take for our purposes that the latter two serve as levers, capable of moving temperature up and down – something that itself is directly confirmed in the process of certifying measurement.

Temperature is a most wickedly difficult case for my account, but also an opportunity to delve into complexities of nature that require flexibility in the account I am putting forward. The difficulties have been in view for centuries, despite RTM and conjoint measurement's more recent vintage. Lord Kelvin himself points out the obstacles posed for a concept of absolute temperature arising from the apparently indispensable reference substance, and the confounds introduced by specific heat:

Although we have thus a strict principle for constructing a definite system for the *estimation* of temperature, yet as reference is essentially made to a specific body as the standard thermometric substance, we cannot consider that we have arrived at an absolute scale, and we can only regard, in strictness, the scale actually adopted as an arbitrary series of numbered points of reference sufficiently close for the requirements of practical thermometry...

It is ... now recognized (from the variations in the specific heats of bodies) as an experimentally demonstrated fact that thermometry under this condition is impossible, and we are left without any principle on which to found an absolute thermometric scale. (Kelvin 1851)

The trouble is this⁹: the abstractly described procedure in the theory of conjoint measurement could be thought to achieve its ambitions if pressure and volume *always* allowed for demonstrating continuity of temperature in *any* substance and at *every*

⁸The Wikipedia entry on the topic of conjoint measurement (https://en.wikipedia.org/wiki/Theory_of_conjoint_measurement) is an excellent account, if more details of these concepts are wanted.

⁹I am grateful to a kind referee for this journal for clearly articulating this concern.

temperature range. And this is where the trouble with heat capacity comes in. The theory of conjoint measurement construes temperature independently of theory, and more specifically as the property of a substance to vary in a fixed way with (i.e., levered by) certain other macrovariables. The trouble is that there is a confound – that temperature variations are modulated also by a confounder: variations and discontinuities in heat capacity, defined as the amount of heat required to raise a fixed amount of a substance one temperature unit. Even worse: specific heat varies by substance. So how is it possible to arrive at an absolute temperature scale when levering temperature, when different substances behave so differently? True, heat capacity is clearly defined independently of theory (it is, to repeat, the amount of heat required to raise a fixed amount of a substance one temperature unit), but this definition is in terms of the quantity (a unit of that temperature) that we seek to certify.

The answer to confronting complexities in temperature is to meet it honestly within our theories, both scientific and philosophical. Therefore we must always be open to the possibilities of yet unknown confounders – which signal that our theory is still insufficiently sensitive to the factors that play a role in shifting still other factors in a system. This is the simple recognition that science is defeasible, demanding revisability also in claims that substantially rely upon the relevant scientific facts. This defeasibility must be matched by revisability in our certification efforts. And so we must answer this challenge to certification, via newly identified confounds, by a story of certification that is itself open to parallel revision – all efforts to certify a quantity must be amendable as follows: that they must be open to the possibility that perhaps there are more than the number of levers thus far identified. Thus a conjoint theory for temperature must be amended whenever we learn of new factors that show themselves in specified circumstances, in some ranges of value, under some conditions, and so on.

The trick to working with such complexity is as follows: once we identify a new confound, then we devise a new procedure that ensures that, when we hold the newly identified confound fixed throughout the specific range of its impact (i.e., when we keep its impact constant), together with those of all other known confounds, the familiar procedure of certification (the levering by, e.g., via pressure and volume) can be applied and will prove successful. The work of certification must itself be revised when new confounds are discovered, or we are not responding to the potential complexities of nature.

The simple point is this: discovery of a new confound is not an occasion to throw up one's hands in defeat. If we were to do so in the present case, we would be denying that (for example) 72 degrees is the same fact no matter what the substance we're attributing it to; furthermore we would have to give up the highly dear notion that it does not matter which sort of thermometer we use, or which substance was used to calibrate the thermometer. These ideas are dear to the scientific enterprise. And giving them up is positively *not* what scientists have opted to do, even in the wake of Lord Kelvin's comments.

Discovery of complexity in one phenomenon has long been treated as an opportunity for complexifying a range of related theories that have managed to confront the world so far successfully, but would benefit from the new twist. And this is precisely what thermodynamicists have done in the time since Kelvin. They have consistently spoken of temperature as a feature of bulk states of matter that can be described by an increasingly larger number of constants or variables (different constants for heat capacity under isobaric conditions, under isochoric conditions, under phase changes, and much else), that may vary across substances, even before there is theory in

which these confounders came to feature. True, variations in heat capacity throws a monkey wrench into the works for the theory of conjoint measurement of temperature, but we have tools for ensuring that we can work around it in practice. So theory of measurement should not despair.

To recapitulate, the crucial question is this: did we *rely* upon any theory – for instance the ideal gas law ($PV = nRT$) – in the case for certification, explicitly or otherwise? And to repeat: the answer is a resounding NO. True or false, the ideal (or any other) gas law plays no role whatsoever in the certification process for temperature. Neither does any other theory, old or new. The same goes for all the assumptions that Chang identifies, including the assumption that melting point and freezing point (or boiling point/condensation point) are the always the same. The certification proceeds by demonstrating that the measured magnitudes among the three quantities obey the certain axioms mentioned before (e.g., *single* and *double cancellation*, or *solvability* and the *Archimedean axioms*, to establish temperature as continuous), within an appointed range, moving through as many ranges as required so as not to fall afoul of confounds, without calling on theoretical principles: all the heavy lifting is done by the empirical measurements themselves. If it is ever necessary to bring in theoretical principles for the heavy lifting, I would consider the effort at (direct) certification to have failed.

There is, finally, the question of whether taking P and V as levers of temperature T is a theoretical assumption that should itself count as an infecting one. I answer, once again, resoundingly NO. Taking P and V as levers of T amounts to taking these as causally effective ways of shifting T. But this assumption is not a scientific theory proper. It is no more scientific theory than the assertion that fire causes smoke is scientific theory, as contrasted with mere common sense grounded in observation. Philosophers who are attracted by the notion that identification of levers is ultimately a matter of theory must give up the distinction between science proper, on the one hand, and common sense on the other – something that Aristotle insisted we must never do (cf. Thalos 2013). Finally, if one were recalcitrantly partial to the idea that there is no line between common sense and science, then the notion that measurement is theory-laden should have no bite whatsoever. Erasure of that line is what theory-ladenness amounts to, so insisting upon erasure of the line is simply begging the question against the independence of theory and measurement that I am building here.

As a matter of history, the questions that required settling before a scale type could be pronounced for temperature were: (1) Are intermediates always possible between any two meaningful temperature points?; and (2) Is zero, as such, meaningful? The answer to (1) was simply taken for granted (yes) in the history of the subject – and without regard to substantive theory. The answer to (2) did not need to be settled at the same time. Indeed there are many possible argumentational pathways of insisting upon a meaningful zero (as there actually were in the case of temperature), as well as many possible theoretical motivations for actually determining its value relative to conventionally appointed ones. In the case of temperature the question took some time to settle (in the affirmative, and as to its precise value in relation to interval scales already in use; from the surprisingly good extrapolations of Guillaume Amontons in 1702 of $-240C$ to Lord Kelvin's calculations of $-273C$ based at least in part on Carnot's theory of heat in 1848). Still, settling that question does not – and historically did not – depend on any substantive theory of the thermodynamic quantities that were initially measurable phenomenologically (temperature very prominently among them). Efforts to reach the “limit to the degree of cold” had perhaps more to do with reaching agreement on

where that point lies than any theoretical considerations – a fact that (if indeed a fact) bolsters the case for theory neutrality of scale choice that I have been pressing.

What's more, substantial scientific questions still remain as to the correct theoretical account of temperature – many disputes are as yet live. Still, the scale type for temperature is now settled: it is a ratio scale. In 1851 Lord Kelvin (1851) wrote to the Royal Society of Edinburgh:

The characteristic property of the scale which I now propose is, that all degrees have the same value; that is, that a unit of heat descending from a body A at the temperature T° of this scale, to a body B at the temperature $(T - 1)^\circ$, would give out the same mechanical effect, whatever be the number T. This may justly be termed an absolute scale since its characteristic is quite independent of the physical properties of any specific substance.

And no amount of theorizing since that time has unsettled this point of agreement. It has remained settled from 1848 onward, and throughout the controversies swirling around the conceptualizations of temperature in statistical–mechanical terms, the status of the second law of thermodynamics and the theoretical upheavals and innovations wrought by ergodic theory within it, beginning in the 1860s and running to this day. This point by itself tells against the notion that certification of the measurement of any given quantity should (by default) be construed as theory-laden.

Temperature is thus certifiable without any appeal to scientific theory. This certification allows us to say that measurement of temperature (and to reiterate: this does *not* apply to any specific instance of an assignment of magnitude to a temperature measurement) has an independent observational and theoretical basis – that is, a basis independent of and orthogonal to any scientific theorizing involving the quantity temperature. This allows us to confirm an FE.

While the dictum that the justification of science is a separate matter from its historical development is oft-repeated (that we must distinguish the context of justification from the context of discovery), operationalizing it (to coin a phrase) has been less clear. The account of measurement certification, methinks, makes a strong beginning: the focus on certification of quantities as separate from the justification of theory and the assignment of specific magnitudes (which latter has to involve, in turn, testing of predictions and repeatability).

9. Further engagements with the critics

Sherry (2011) might be viewed as containing an implicit criticism of the present position. Like Chang (2004), he adopts what he refers to as a “pragmatic” approach. By that, he means that if one makes an assumption (baselessly if necessary, as perhaps can be said of chemist Joseph Black’s attitude vis-à-vis the measurement of temperature¹⁰) that something is a quantity (i.e., appropriately measured on a cardinal scale), and this proves a fruitful path to the elaboration of one’s theory (e.g., to the formulation

¹⁰He writes: “Black treated Fahrenheit’s instrument as a *bona-fide* measurement device, even though one could object reasonably to his doing so. Our three objections would have been decisive if Black hadn’t been able to *use* the instrument in company with the medieval quantification of quantity, to turn familiar qualitative observations into a pair of quantitative concepts fundamental to the science of heat – capacity for heat (nowadays, specific heat) and latent heat”. (2011: 515)

of concepts like specific heat and latent heat) as well as a body of confirmatory measurements, then this is grounds enough for answering objections (perhaps he means to the assumption that we have a quantity) – but not grounds for quantity realism. Still, unlike Chang (2004), Sherry (2011) does not advance a “pragmatic” or coherentist approach to answering the problem of nomic measurement; nor does he view it as insoluble in the style of Tal (2017b). Instead, he – like myself – looks to the precepts of the theory of measurement, particularly the theory of conjoint measurement. Sherry (2011: 520) writes: “Conjoint measurement gives, then, conditions under which a thermoscope can be used as a thermometer, and in this way it provides a rigorous mathematical solution the problem of nomic measurement.” While I, by contrast, do not agree that there can be a purely mathematical solution to this philosophical problem, I do agree that conjoint measurement is the basis of a philosophical answer to it. (This paper, give or take, is that philosophical answer.)

Sherry’s position is therefore in agreement with my own in connection with how to approach the problem of nomic measurement, although he denies that conjoint measurement status is sufficient to guarantee quantitative status (his argument focuses on temperature). Sherry insists that quantitative status is tied to what he refers to as “predictive and explanatory success” – concepts for which he gives no analysis. This is not the place to discuss explanatory and predictive notions of success, but suffice it to say that, unlike Sherry, I see no extra role for such success concepts to play in quantitative certification – i.e., the suitability of a certain mathematical object (a scale) for representing the outcomes of measurements of that quantity. Thus on this point, my view and Sherry’s part ways.

Finally, one might be suspicious of my position that it is ignorant of modern metrology: the meter is now defined in reference to the speed of light in vacuum and the kilogram is defined (since 2019) in reference to Planck constant. So how is it possible that measurement of these things, in whatever manner, can be free of theories that assure us the speed of light is constant or bring Planck’s constant to our attention in the first place? The answer here is that there has to be a distinction between, on the one hand, the operationalization of measuring a particular quantity (say, via sticks laid end to end or a balance scale), and on the other, a unit of measure for the quantity (the meter or kilogram). These are different things, conceptually speaking. Defining the one (the meter or kilogram by reference to something found in a theory) does not impugn its operationalization via a previously devised measurement operation.¹¹ Surely it hands too much power to metrologists if a new definition of the second (for whatever theoretical purpose) could render our current timepieces obsolete – if, in other words, metrologists’ work were in a position to require the invention of new timekeeping devices.

10. Conclusion

The present paper has come to the philosophical defense of the doctrine of FE via an enhanced account of the RTM, in the process of offering an account of the certification

¹¹As Sherry too observes, physicists may also be prepared to deal with various foundational problems in their discipline by abandoning the continuity of space and time – and, presumably also in that same spirit to nominate measuring (small) distances and even (small) masses by reference to constants like the speed of light and Planck’s constant – but presumably would not “claim that the correctness of their speculations would reveal that time-tested methods for measuring length and distance were a sham”. (Sherry 2011: 523)

of quantities. I have shown that the question of quantity certification is settled independently of the question of the reliability of observation of any particular quantity on any particular occasion. This position in the metaphysics and epistemology of measurement is not a position on realism as such. Still, the effect is to give confidence to the measurability of quantities, on a case-by-case basis – as certain fixed points in the body of evidence, independent of any particular theory but not independent of experimental evidence around the quantity in question. Thus certification of quantities is, in a very precise sense, scientific-theory-free. Certification thus stands apart from scientific theory. It is legitimately regarded as an independent pedestal in the knowledge enterprise.

Science has no more faithful ally than the doctrine of FE. They are a dying breed today, those who come to its defense. But defense of FE is more important than ever. Still, one might wonder whether the case I have made is sufficient for foundationalism – whether the certification of quantities is enough to ensure that there is indeed a sufficiently independent basis of evidence, for the ultimate testing of scientific theories. Certification might still be thought to be insufficient to ensure foundationalism. Sufficient or not, it must be granted that quantity certification removes at least one roadblock to foundationalism: the theory-ladenness of measurement.¹²

References

- Batitsky V.** (1998). 'Empiricism and the Myth of Fundamental Measurement.' *Synthese* **136**, 51–73.
- Boumans M.** (ed.) (2007). *Measurement in Economics: A Handbook*. London: Academic Press, Emerald Group.
- Boyd R.** (2010). 'Realism, Natural Kinds and Philosophical Methods.' In H. Beebe and N. Sabbarton-Leary (eds), *The Semantics and Metaphysics of Natural Kinds*. London: Routledge, 212–234.
- Boyd R.** (2019). 'Rethinking Natural Kinds, Reference and Truth: Towards More Correspondence with Reality, Not Less.' *Synthese*. **198**, 2863–2903. <https://doi.org/10.1007/s11229-019-02138-4>.
- Byerly H. and Lazara V.** (1973). 'Realist Foundations of Measurement.' *Philosophy of Science* **40**, 10–28.
- Campbell N.R.** (1920). *Physics: The Elements*. London: Cambridge University Press.
- Chang H.** (2004). *Inventing Temperature*. New York: Oxford University Press.
- Chang H.** (2017a). 'VI – Operational Coherence as the Source of Truth.' *Proceedings of the Aristotelian Society* **117**(2), 103–22.
- Chang H.** (2017b). 'Scientific Progress: Beyond Foundationalism and Coherentism.' *Royal Institute of Philosophy Supplements* **92**, 1–20. <https://doi.org/10.1017/S1358246122000303>.
- Decoene S., Ongheena P. and Janssen R.** (1995). 'Representationalism Under Attack: Review of an *Introduction to the Logic of Psychological Measurement*, by J. Michell and *Philosophical and Foundational Issues in Measurement Theory*, by C. Wade Savage and P. Ehrlich.' *Journal of Mathematical Psychology* **39**, 234–42.
- Domotor Z. and Batitsky V.** (2010). 'An Algebraic-Analytic Framework for Measurement Theory.' *Measurement* **43**(9), 1142–64.
- Heilman C.** (2015). 'A New Interpretation of the Representational Theory of Measurement.' *Philosophy of Science* **82**, 787–97.

¹²Earlier versions of this paper were presented at the Workshop on Scientific Explanations: Competing and Conjunctive (2019) at the University of Utah, and as a symposium presentation at the 2021 Eastern Division of the American Philosophical Association. I am grateful for comments from Lu Chen and Peter Tan at the APA session, and for helpful suggestions and comments from audience members at both meetings. I acknowledge the debt I owe to friends and colleagues who read and provided feedback, suggestions, and encouragements: Thi Nguyen and Robert Richardson. It is not always possible to remember the sources of help and inspiration, but I am uncommonly grateful for the many directions from which all these come, even if I cannot always pay them proper credit.

- Holman E.W.** (1971). 'A Note on Conjoint Measurement with Restricted Solvability.' *Journal of Mathematical Psychology* 8, 489–94.
- Isaac A.M.C.** (2019). 'Epistemic Loops and Measurement Realism.' *Philosophy of Science* 86(5), 930–41.
- Kelvin W.T.** (1851). 'Dynamical Theory of Heat.' *Transactions of the Royal Society of Edinburgh. In Mathematical and Physical Papers.* Cambridge: Cambridge University Press, 174–332. <https://www.originalsources.com/Document.aspx?DocID=B8DVYIJQDSW5XR9>.
- Krantz D.H.** (1964). 'Conjoint Measurement: The Luce–Tukey Axiomatization and Some Extensions.' *Journal of Mathematical Psychology* 1, 248–77.
- Krantz D.H., Luce R.D., Suppes P. and Tversky A.** (1971). *Foundations of Measurement, Vol. I: Additive and Polynomial Representations.* New York: Academic Press.
- Luce R.D., Krantz D.H., Suppes P. and Tversky A.** (1990). *Foundations of Measurement Vol 3: Representation, Axiomatization, and Invariance.* San Diego and London: Academic Press.
- Luce R.D. and Tukey J.W.** (1964). 'Simultaneous Conjoint Measurement: A New Scale Type of Fundamental Measurement.' *Journal of Mathematical Psychology* 1, 1–27.
- Mitchell J.** (1990). *An Introduction to the Logic of Psychological Measurement.* Hillsdale, NJ: Erlbaum.
- Mitchell J.** (1995). 'Further Thoughts on Realism, Representationalism, and the Foundations of Measurement Theory: Author's Response to Review by Decoene *et al.* of "An Introduction to the Logic of Psychological Measurement."' *Journal of Mathematical Psychology* 39, 243–47.
- Mitchell J.** (2005). 'The Logic of Measurement: A Realist Overview.' *Measurement* 38, 285–94.
- Mundy B.** (1987). 'The Metaphysics of Quantity.' *Philosophical Studies* 51(1), 29–54.
- Narens L. and Luce R.D.** (1986). 'Measurement: The Theory of Numerical Assignments.' *Psychological Bulletin* 99, 166–81.
- Reiss J.** (2008). *Error in Economics: Towards a More Evidence-Based Methodology.* London: Routledge.
- Sherry D.** (2011). 'Thermoscopes, Thermometers, and the Foundations of Measurement.' *Studies in History and Philosophy of Science Part A* 42(4), 509–24.
- Stevens S.S.** (1935). 'The Operational Definition of Psychological Concepts.' *Psychological Review* 42(6), 517–27.
- Stevens S.S.** (1946). 'On the Theory of Scales of Measurement.' *Science* 103, 677–80.
- Suppes P., Krantz D.H., Luce R.D. and Tversky A.** (1989). *Foundations of Measurement, Vol 2: Geometrical, Threshold and Probabilistic Representations.* San Diego and London: Academic Press.
- Swoyer C.** (1987). 'The Metaphysics of Measurement.' In J. Forge (ed.), *Measurement, Realism and Objectivity*, pp. 235–90. Dordrecht: Reidel.
- Tal E.** (2013). 'Old and New Problems in Philosophy of Measurement.' *Philosophy Compass* 8(12), 1159–73.
- Tal E.** (2017a). 'Measurement in Science.' *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.). <https://plato.stanford.edu/archives/fall2017/entries/measurement-science/>.
- Tal E.** (2017b). 'A Model-Based Epistemology of Measurement.' In N. Mossner and A. Nordmann (eds), *Reasoning in Measurement*, pp 233–253. NY: Routledge.
- Tal E.** (2019). 'Individuating Quantities.' *Philosophical Studies* 176, 853–78.
- Thalos M.** (2013). *Without Hierarchy: The Scale Freedom of the Universe.* New York: Oxford University Press.
- von Helmholtz, H.** (1887). *Counting and Measuring*, C.L. Bryan (trans.), New Jersey: D. Van Nostrand, 1930.
- Wigner E.** (1960). 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences.' *Communications in Pure and Applied Mathematics* 13(1), 1–14.
- Wilson M.** (2008). *Wandering Significance.* New York: Oxford University Press.

Mariam Thalos (PhD, University of Illinois/Chicago, 1993) is a distinguished professor of philosophy at the University of Tennessee, Knoxville. She is a philosopher of science, first and foremost, focusing on physical, decisional, and human sciences, including formal decision and game theories. She is also active on philosophical topics around practical reasoning, including action theory, phenomenology, and theories of freedom. The author of two monographs and more than 100 articles and essays, she is currently working on a comprehensive philosophical study of human reasoning in the wild.