Epistemic Loops and Measurement Realism

Alistair M. C. Isaac

Abstract

Recent philosophy of measurement has emphasized the existence of both diachronic and synchronic "loops," or feedback processes, in the epistemic achievements of measurement. A widespread response has been to conclude that measurement outcomes do not convey interest-independent facts about the world, and that only a coherentist epistemology of measurement is viable. In contrast, I argue that a form of measurement realism is consistent with these results. The insight is that antecedent structure in measuring spaces constrains our empirical procedures such that successful measurement conveys a limited, but veridical knowledge of "fixed points," or stable, interest-independent features of the world.

§1 Introduction

Recent philosophy of measurement has employed detailed case studies to highlight the complex, iterative process by which measurement practices are refined. Typically, these examples are taken to support some form of epistemic coherentism, on which the validation of measurement procedures, and thus their epistemic import, is irreducibly infected by the contingent history of their development in aid of human interests. This coherentism in turn undermines *measurement realism*, the view that outcomes of successful measurement practices veridically represent objective (i.e. interest-independent) features of the world. For instance, van Fraassen (2008) takes the historical contingency of measurement practice to support empiricism, and Chang (2012) argues that only a pragmatic, interest-relative "realism" about measurement outcomes is plausible, not one which interprets them as corresponding to objective features in the world. More generally, Tal (2013) identifies coherentism as a major trend within contemporary philosophy of measurement.

I argue that the iterative and coherentist features of measurement practice these authors rightly emphasize are nevertheless consistent with realism about measurement outcomes. Nevertheless, my position contrasts significantly with that of other measurement realists, such as Byerly and Lazara (1973) or Michell (2005), who take measurement realism to be continuous with global scientific realism. On their view, measurement realism is a *stronger* position than traditional realism, imputing reality not only to theoretical objects and laws, but also to their quantitative character. The view defended here reverses this priority, articulating a realism about measurement outcomes *weaker* than traditional realism. In particular, I argue that the convergent assignment of increasingly precise values that constitutes successful measurement serves as incontrovertible evidence for *fixed points* in the world — features or events standing in stable quantitative relationships — even though the evidence it provides for any non-numerical theoretical description of these points is defeasible. The insight here is that measurement is more evidentially demanding than traditional confirmation, i.e. it requires a greater contribution from the interest-independent world to succeed than mere qualitative experiments. I argue that this greater evidential demand is a consequence of the antecedent numerical structure in which measurement outcomes are represented. This antecedent structure blocks the possibility of gerrymandered categories that crosscut the joints of nature. Consequently, successful measurement constitutes a substantive enough epistemic achievement that we may legitimately "factor out" the contribution to success made by human interests, and accept the outcome as representing an objective feature of the world.

After surveying the motivations for measurement coherentism, I elaborate on the notion of "successful" measurement, and why it poses a challenge to coherentism. The paper concludes with a more careful articulation of the distinctive features of fixed point realism.

§2 Epistemic Loops in Measurement Practice

Contemporary measurement coherentism is motivated by two types of case study, each identifying a different kind of epistemic "loop," or feedback process driving knowledge formation. Chang and van Fraassen emphasize diachronic examples of epistemic iteration, where the feedback process extends over several stages of mutual influence between theory change and refinement of measurement practice. A different kind of epistemic loop has been discussed by Tal and metrologist Mari, who highlight the role of models in the calibration of measurement instruments and the assignment of quantity values, illustrating a synchronic epistemic interdependence between theory and measurement.

§2.1 Epistemic Iteration

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). He takes this process to support a "progressive coherentism": on the one hand, the criteria for measurement success are internal to a practice, so scientific knowledge does not rest on an independent foundation; on the other hand, these internal criteria may be used to evaluate new practices as improvements or refinements on their predecessors, thereby allowing for scientific progress (in contrast to traditional coherentism, Chang 2007). In the context of measurement, this means that later measurement practices may be understood as in some sense "better" than earlier ones, yet these "epistemic achievements" should not be cashed out as greater degree of correspondence to quantities in the world.

For instance, thermometry as a practice begins with subjective assignments of relative heat on the basis of our bodily experiences. Noticing that fluids appear to change volume in rough correspondence with these subjective sensations, one may construct a thermoscope, or device allowing comparison of relative fluid volumes in different circumstances. Already a theoretical leap is required to identify the cause of these changes in relative volume with the cause of our differing subjective sensations, especially given the discrepancies between these sensations and our thermoscopic readings (e.g. contrary to experience, caves are warmer in summer than they are in winter). Nevertheless, the move to the thermoscope constitutes an epistemic achievement, in the sense that it allows for greater regularity in the assignment of relative temperatures, both across contexts and across observers. A similar pattern is seen in the move from thermoscope to thermometer, which enables assignment of numbers to temperatures. Numerical representation constitutes a yet greater epistemic achievement, insofar as it allows comparison of temperature assignments across devices. Nevertheless, this practice does not itself guarantee greater veracity of temperature assignments, since it rests on the assumption that temperature varies linearly with changes in the height of thermometric fluid. But this assumption cannot itself be verified, as that would require access to temperature in the world by some means independent of thermometry. Similar achievements, (seemingly) inextricably entangled with theory, may be seen at each further stage in the development of thermometric practice.

The moral of this case study is the historical contingency of thermometry, and thus of its results. At each stage in the development of thermometry, an advance in theory was required to extend measurement practice. Internal criteria of consistency and increased precision in the assignment of numerical values establish the new practice as an advance over the previous one. Yet, the application of these criteria is not empirically constrained. When one assumes that "temperature" (whatever it may be) varies linearly with changes in the height of the indicator column in an air thermometer, one is making an assumption both necessary for measurement progress and in principle non-empirical, since no independent access to "temperature," outside the behavior of the very devices and procedures under investigation, is possible: "*Prior to the construction of a thermometer, there is no thermometer to settle that question!*" (van Fraassen 2008, 126, emphasis in original). Chang (2004) argues that, in order to make sense of the "progress" exemplified by cases like these, we have to "look away from truth," and appeal only to historically contingent criteria for success (227)—"scientific progress … cannot mean closer approach to the truth" (228); "Truth, in the sense of correspondence to reality, is beyond our reach" (Chang 2007, 20). The delusion that one may evaluate the correspondence between our assignment of temperatures and the objective state of the world rests on the mistaken and "impossible god-like view in which nature and theory and measurement practice are all accessed independently of each other" (van Fraassen 2008, 139). Rather, the only relevant notion of "truth" for assessing the success of thermometry "rests first and foremost on coherence with the rest of the system" (Chang 2012, 242).

§2.2 Models and Calibration

Another kind of epistemic loop is found in synchronic measurement practice, where *models* play a constitutive role in determining measurement outcomes. The crucial concept here is *calibration*, the process of correcting a measurement device for inferred discrepancies between its readout and the target value. Calibration is a necessary feature of all sophisticated measurement, yet the process of calibration illustrates the ineliminable role of theoretical posits in the very assignment of quantity values in an act of measurement. When measuring, scientists do not (as one might naively suppose) read values directly from nature, rather they employ models of the interaction between measurement device and target system in order to "correct" the readout value to a final assigned value (Mari and Giordani 2014).

Tal (2014) illustrates this point through the example of the measurement of time, in particular coordinated universal time (UTC). The second is presently defined as 9,192,631,770 periods of the hyperfine transition between the two ground states of a caesium-133 atom at zero degrees Kelvin.[1] Models feature at every step of the process leading from devices that interact directly with caesium atoms to the UTC. First, it is impossible to probe caesium atoms at absolute zero, so the enumeration of hyperfine transitions output by a caesium clock must be corrected for this discrepancy. This, as well as other corrections, rely on models of the physical interaction between the device and the atom in order to infer the discrepancy between the actual state of the system and the idealized state referred to in the definition. Caesium clocks are too complex to run continuously, so their output is used to calibrate more mundane atomic clocks (301). Furthermore, the UTC itself is not identified with the output of any one clock; rather, it is calculated retrospectively by a weighted average over all participating atomic clocks, with weights determined by the degree of past fit between each clock and previous calculations of UTC (302–3).

The lessons of this example are analogous to those of epistemic iteration: measurement improvement appears to rest on internal standards of coherence rather than on correspondence with external quantities. The weighting procedure that leads to UTC, for instance, "promotes clocks that are stable relative to each other" (304). Success at achieving this stability indeed demonstrates "genuine empirical knowledge," but not knowledge in the first instance about a regularity in the objective world, but rather a regularity "in the behaviour of instruments" (327). Consequently, it is a "conceptual mistake" to think that "the stability of measurement standards can be analysed into distinct contributions by humans and nature" (328). On an extreme interpretation of this view, even computer simulation constitutes a form of measurement (Morrison 2009). The basic idea is that, once we grant the ineliminable role of models in measurement, it is a small conceptual step to accept that the aspect of measurement involving empirical contact with the world may be arbitrarily distant from that involving modeling (Parker 2017).

§3 Achieving Successful Measurement

For the remainder of this paper, I wish to grant the basic descriptive features of this account: both diachronically and synchronically, successful measurement involves epistemic loops. Nevertheless, I will argue, there is a form of measurement realism consistent with these loops; one on which the contingent, interest-relative, and theory-laden aspects of measurement may indeed be factored out, leaving the bare, objective facts about the world conveyed by successful measurement.

---

[1] Arguably, the process of establishing UTC is not measurement at all — since the length of the second is *defined* by caesium-133 transitions, it is not subject to empirical determination. The purpose of the project Tal examines is not to establish a value, as in paradigmatic cases of measurement, but rather to coordinate time-relevant activities across the globe with maximal precision. I set this concern aside for the discussion here, since Tal's analysis has been so influential in philosophy of measurement, and his conclusions concerning the model-mediation of measurement incontrovertibly reflect the practices of metrologists.

But what is "successful measurement"? For the purposes of discussion here, I take *measurement* to be any empirical procedure for assigning points (or regions) in a metric space to states of the world, where a *metric space* is any set of elements with a distance metric defined over it. This means, on the one hand, that I rule out degenerative forms of "measurement" that simply assign objects to categories, or place them in an order (the *nominal* and *ordinal* scales of Stevens 1946). On the other hand, I include measurement procedures that map states of the world into any geometrical space, not just the real line, so long as they have an assigned distance metric (siding with Suppes, et al. 1989, against Díez 1997); nevertheless, in the interests of simplicity, I will refer to these outcomes as "numerical" assignments, since they may be represented by vectors of real numbers. In line with Krantz et al. (1971), I take it that one can determine whether or not an empirical procedure constitutively requires the metric features of a geometrical space by analyzing whether these remain invariant across permissible transformations over the mapping into that space.[2]

I take *successful* measurement to exhibit two key features: *convergence* and *precision*. These features pose a significant challenge to the thoroughgoing coherentist.

§3.1 Convergence

Coherentists have emphasized the theory-ladenness of both diachronic and synchronic aspects of measurement refinement. However, a hallmark of sophisticated scientific measurement is its attempt to factor out the role of theory in measurement by employing different theoretical commitments to measure the same quantity. A measurement practice *converges* when procedures employing different theoretical commitments arrive at the same outcome.

For instance, in the early 20th century, a wide variety of phenomena were investigated, employing distinct methods and theoretical commitments, in the attempt to measure Avogadro's constant N_A, the number of particles in a mole of substance. Perhaps most well-known are Perrin's experiments on Brownian motion, which, in combination with Einstein's theoretical analysis, allowed an assignment of value to N_A. However, similar values were achieved by radically different means. For instance, Millikan was able to determine N_A by measuring charge of the electron through his oil drop experiments and dividing the Faraday constant (charge of a mole of electrons) by his result. Millikan's measurement relied on Stokes' theoretical analysis of the movement of spheres through a viscous fluid — insofar as Brownian motion was a factor, it was as a source of noise, not (as for Perrin) a source of evidence. Black body radiation and the blue

---

[2] For instance, consider two procedures for assigning real numbers to my students. On the first, I assign a number to each letter-type with which a student's name begins (e.g. A=1, B=3,…); on the second, I hold a meter stick up to each student and note their height. The former procedure is indifferent to the algebraic structure of the real line (letters do not add or subtract from each other systematically), and thus metric features of the real line are not invariant across alternative, equally permissible assignments of numbers (e.g. A=7, B=15,…). The second does make use of algebraic structure (as heights do "add" through concatenation), and thus metric features remain invariant across alternative assignments (Jamal is twice the height of Leslie, whether their heights are represented in inches or centimeters). So, on the present definition, the latter procedure is measurement, but the former is not.

of the sky are examples of other phenomena that, when combined with theoretical models of photon emission and diffraction respectively, allow alternate means of measuring N_A. Insofar as these procedures assign the same value to N_A, they converge.

I want to stress that the point being made here is *not* the traditional realist one, that these practices provide converging evidence for the particulate nature of matter, whether as "common cause" (Salmon 1984) or most likely hypothesis (Psillos 2011). Those arguments are instances of *abduction*, while I am interested in whether a stronger, non-abductive conclusion may be drawn from convergence. A better analogy is with the discussion of robustness in the modeling literature: a result is *robust* if it is obtained by a plurality of models that each make different simplifying assumptions (Weisberg 2006). The particulate nature of matter is not robust in this sense across different measurement practices, since it is assumed by all of them. However, the value of N_A is robust, since that value is not itself assumed, and is obtained with a great degree of agreement despite differences in the assumptions made by each measurement practice (and its supporting models). I claim that convergence toward this value provides robust, non-abductive evidence for an objective feature of the world.

This example is in no way exceptional: convergent measurement practices are rife across the sciences. Smith and Miyake, for instance, have investigated a number of examples. Thomson's convergent measurements of the charge of the electron employed a variety of different methods and assumptions (Smith 2001). Early attempts to measure the density of the interior of the earth likewise assumed a variety of different theoretical models (Miyake 2018). In more recent research, measurements of the constants that govern molecular vibration converge across spectroscopy, chemistry, thermodynamics, and femtochemistry (Smith and Miyake, *manuscript*). To pick an example from an entirely different area of science, measurements of the spectral sensitivity of mammalian retinal receptors employing psychophysical methods (extracting sensitivity curves from behavioral color matching experiments, as performed by Helmholtz in the late 19th century) converge closely with 20th century physiological methods (detecting rate of nerve firing in (e.g.) cow retinal tissue in response to single wavelength lights, Wandell 1995). In all of these cases, "What is being shown through the convergence of these measurements is that the discrepancies between the different measurements … are due to the particularities of the models being used" (Miyake, 2018, 336). In other words, convergence factors out model-sensitive features of measurement; in order for it to occur, "the empirical world has to cooperate" (Smith 2001, 26).

§3.2 Precision

Traditionally, measurement success was evaluated with respect to two features: accuracy and precision. *Accuracy* was degree of approach to true value, while *precision* was degree of specificity in the value provided. The considerations in §2 undermine the criterion of accuracy, since they show we have no independent access to "true values" and thus cannot use them as standards for evaluating measurement (Mari 2003). Nevertheless, we can still assess measurements for precision, since it may be defined operationally: a measurement is *precise* to the extent that it returns the same result when performed repeatedly. The number of *significant figures* in a numerical assignment indicates the degree of measurement precision, since these characterize the size of the region within which repeated measurements fall.

Coherentists stress the fact that increased precision is a purely internal criterion for improving measurement. Here, however, I want to stress the way in which increased precision constitutes a qualitatively different, and more impressive, epistemic achievement than other forms of empirical success, such as qualitative prediction or improved coherence of classification. These qualitative achievements are subject to worries about semantic and theoretical holism: one may always succeed in classification, or correct qualitative prediction, by suitably redrawing the boundaries of one's theoretical concepts. As LaPorte (2004) argues, when faced with anomalies in the relationship between guinea pigs and prototypical rodents, or birds and dinosaurs, scientists face a *choice* whether to expand or contract their previous categories to include or exclude perceived outliers (a similar case is made by Slater 2017 for Pluto and planethood). Nothing about the prior conceptual framework itself forces this choice one way or another, nor do demands for internal consistency.

Measurement is different from mere categorization precisely because it maps states into a metric space. The crucial point to note here is that a metric space has *antecedent structure*: the distances between points on the real line, and the algebraic relationships between them, are fixed *before* we employ it to represent height or temperature or electric charge. This antecedent structure constrains the relationship between measurement outcomes, independently restricting our assessment of them as same or different, or converging or not, in a manner impervious to ad hoc revision. Increase in precision occurs when successive measurement practices are able to shrink distances (between repeated measurements within each practice) determined by the metric of the representing space. Thus, the metric of this space serves two functions: (i) it represents the distances between different measured quantities, but (ii) it also provides a directed metric for improving measurement of a single quantity, since it determines the distances between repeated measurements that characterizes their precision. Consequently, pace van Fraassen, attempts to increase precision are empirically constrained, since this directed metric for improvement can only be satisfied through the cooperation of nature: if nature is not sufficiently stable where we probe it, no choice, convention, or increased coherence can reduce the distances between our repeated attempts to measure it. Some examples will illustrate this point.

Consider, for instance, determinations of the boiling point of water. Chang (2004, Ch. 1) surveys the sequence of choice points in the early practice of thermometry leading to relative stability in the measurement of this temperature: what are the visual indicators of boiling, where should the thermometer be positioned, what should be the shape of the vessel holding the water, its material, etc.[3] Decisions on each of these points affect the relative stability in the thermometric reading, illustrating the naivety of a view on which

---

[3] The issue here is the phenomenon of "superheating," whereby water with relatively little dissolved gas, or in a flask with very small surface area, may be heated to a higher temperature without bubbling.

boiling point is a simple phenomena merely waiting to be observed. Nevertheless, in committing to represent the boiling point numerically, investigators subjected themselves to a criterion for success distinct from coherence. If the numbers assigned by thermometers within this-shaped vessels and that-shaped ones differ during phenomenologically similar bubblings, then the distance between those numbers provides a criterion of difference that must be respected if thermometric practice is to count as measurement. Restricting attention to those vessels that minimize distances between numerical outcomes is thus not a mere choice, or gerrymandering of the category "boiling," since it is forced upon the investigator by an antecedent metric for success.

Likewise, consider again the determination of UTC through the retrospective weighting of the comparison set of atomic clocks. For Tal, the success of this procedure is evidence for stability in our clocks, but not for any human-independent feature of the world. Nevertheless, UTC is constrained by the world in two distinct ways. First, through empirical contact with caesium atoms. While this contact is mediated by models, these models themselves are the result of convergent measurements of atomic phenomena through a wide variety of means, employing distinct theoretical assumptions. Second, the distance metric of the real line constrains the assessment of fit between clocks in the set. While the algorithm that weights them takes degree of internal agreement as the standard for higher weighting, the metrical structure of the space in which relative rates of the clocks are assessed ensures relative agreement cannot be stipulated, fudged, or gerrymandered. The clocks need to cooperate by performing stably enough that they may be compared with a high degree of precision, and this stable point remains tethered to a robust regularity in the world through checks with the convergent behavior of caesium.

While UTC is in some respects atypical (see footnote 1), these three features — internal coordination of outcomes, empirical checks, and directed improvement constrained by the real line — are features of scientific measurement in general. What Tal's discussion of the UTC obscures is the sheer number of empirical checks typically involved, and the strictness of the demands placed by conformity to the metric of improvement the measuring space provides. In official determinations of fundamental physical constants, convergence is demanded across *all* measurement procedures, as assessed by the law-governed interrelationship between physical quantities, and the degree of precision achieved illustrates the strictness of this demand. For instance, in late 19th century measurements of N_A by Perrin and e (charge of electron) by Thomson, only 2 to 3 significant figures were typically obtained within method, and convergence across methods often only agreed as to order of magnitude. By 1911, Millikan was measuring both e and N_A to 4 significant figures, and demonstrating that the models employed to calibrate the oil drop method converged closely with other aspects of physical theory (1911). As of 2014, N_A was being measured at upwards of 9 significant figures, and e upwards of 11 (Mohr et al. 2016).[4] In each case, the increase in precision has been constrained by the antecedent structure of the real line, and thus is not itself a matter of mere convention or coherence. Rather, the world must cooperate by remaining

---

[4] It is expected that after the 2018 26th General Conference on Weights and Measures, N_A and e will be fixed as constants to which other quantities may be referred during measurement.

sufficiently stable if such precision is to be possible; consequently, precise values constitute robust evidence for points of objective fixity in the world revealed through measurement.

§4 Conclusion: Fixed-Point Realism

Traditional scientific realism rests on an abductive inference from observed empirical success to presumed underlying causes. Successful measurement may certainly be used in such an inference, but I claim here that it non-abductively supports a more modest realism:

> **Fixed Point Realism** – values obtained through successful measurement veridically represent objective fixed points in the world, which may be exhaustively characterized by the pattern of distances that obtain between them in a metric space.

FPR is a form of *epistemic structural realism*. It differs from traditional realism insofar as it claims a veridical characterization of the world is possible independent of any particular theoretical description. Our theory of the nature of temperature or of state changes may change radically, yet the points of relative stability characterizing, e.g., boiling point of water, "absolute zero," freezing point of oxygen, etc., will stay robust across any such change, and that robustness may be represented by their relative positions within a numerical scale.

FPR differs from other flavors of structural realism in the type of structure to which it is committed. Structural realists typically focus on the rich mathematical structure of physical theory, and derivation or limit relations that hold between successive theories, e.g. Newton's laws are a limit case of relativistic mechanics (Worrall 1989). FPR commits itself only to *geometric* structure, i.e. the pattern of relative distances that obtain between points of stability as represented in a metric space. Just as our theoretical description of these stable points may change, so may our mathematical account of their relationship — if new mathematical physics fails to derive old equations as limit cases, this in no way jeopardizes the veridicality of this geometric structure.

Finally, FPR disagrees with coherentism, insofar as it asserts that the geometrical structure uncovered through acts of successive measurement obtains in the world independent of our practices. It does not deny the importance of epistemic loops for understanding the process of measurement. Nevertheless, it takes convergence in measured values to indicate that the points of stability they represent obtain independent of the theoretical commitments encapsulated in the models used for calibration. Likewise, it takes increased precision to constitute a criterion for measurement success over and above that of coherence, one that is only realized when the interest-independent world cooperates with us by remaining stable when we probe it.

Bibliography

Byerly, H., and V. Lazara (1973) “Realist Foundations of Measurement,” *Philosophy of Science* 40:10–28.

Chang, H. (2004) *Inventing Temperature*, Oxford UP.

Chang, H. (2007) “Scientific Progress: Beyond Foundationalism and Coherentism,” O’Hear (ed.) *Royal Institute of Philosophy Supplement* 61:1–20.

Chang, H. (2012) *Is Water* H_2O*?* Springer.

Díez, J. (1997) “A Hundred Years of Numbers: An Historical Introduction to Measurement Theory 1887–1990, part ii,” *Studies in History and Philosophy of Science* 28:237–265.

Krantz, D., R. Luce, P. Suppes, and A. Tversky (1971) *Foundations of Measurement*, vol. 1, Dover.

LaPorte, J. (2004) *Natural Kinds and Conceptual Change*, Cambridge UP.

Mari, L. (2003) “Epistemology of Measurement,” *Measurement* 34:17–30.

Mari, L., and A. Giordani (2014) “Modeling Measurement: Error and Uncertainty,” in Boumans, Hon, and Petersen (eds.) *Error and Uncertainty in Scientific Practice*, Pickering & Chatto: 79–96.

Michell, J. (2005) “The Logic of Measurement: A Realist Overview,” *Measurement* 38:285–294.

Millikan, R. (1911) “On the Elementary Electrical Charge and the Avogadro Constant,” *Physical Review* 2:349–397.

Miyake, T. (2018) “Scientific Realism and the Earth Sciences,” in Saatsi (ed.) *The Routledge Handbook of Scientific Realism*, Routledge: 333–344.

Mohr, P., D. Newell, and B. Taylor (2016) “CODATA Recommended Values of the Fundamental Physical Constants: 2014,” *Review of Modern Physics* 88:035009.

Morrison, M. (2009) “Models, Measurement and Computer Simulation: The Changing Face of Experimentation,” *Philosophical Studies* 143:33–57.

Parker, W. (2017) “Computer Simulation, Measurement, and Data Assimilation,” *British Journal for Philosophy of Science* 68:273–304.

Psillos, S. (2011) “Moving Molecules above the Scientific Horizon: On Perrin’s Case for Realism,” *Journal for General Philosophy of Science* 42:339–363.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton UP.

Slater, M. (2017) “Plato and the Platypus: An Odd Ball and an Odd Duck – On Classificatory Norms,” *Studies in History and Philosophy of Science* 61:1–10.

Smith, G. (2001) “J.J. Thomson and the Electron, 1897–1899,” in Buchwald and Warwick (eds.) *Histories of the Electron*, MIT Press.

Smith, G., and T. Miyake (*manuscript*) “Realism, Physical Meaningfulness, and Molecular Spectroscopy”

Stevens, S. (1946) “On the Theory of Scales of Measurement,” *Science* 103(2684):677–680.

Suppes, P., D. Krantz, R. Luce, and A. Tversky (1989) *Foundations of Measurement*, vol. 2, Dover.

Tal, E. (2013) “Old and New Problems in Philosophy of Measurement,” *Philosophy Compass* 8/12:1159–1173.

Tal, E. (2014) “Making Time: A Study in the Epistemology of Measurement,” *British Journal for Philosophy of Science* 67:297–335.

Wandell, B. (1995) *Foundations of Vision*, Sinauer.

Weisberg, M. (2006) “Robustness Analysis,” *Philosophy of Science* 73:730–742.

Worrall, J. (1989) “Structural Realism: The Best of Both Worlds,” *Dialectica* 43:99–124.

van Fraassen, B. (2008) *Scientific Representation*, Oxford UP.