


SYMPOSIA PAPER

Escape from Zanzibar: The Epistemic Value of Precision in Measurement

Alistair M. C. Isaac 

School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, 40 George Square, Edinburgh, UK

Email: a.m.c.isaac@ed.ac.uk

(Received 14 October 2021; revised 14 March 2022; accepted 14 May 2022; first published online 09 June 2022)

Abstract

A “Zanzibar” is an island of measurement values that internally cohere, but are detached from independent contact with reality. One manifestation of Zanzibars is through “bandwagon effects,” the tendency of contemporaneous measurements to agree. Bandwagons illustrate how the otherwise virtuous drive towards coherence can have negative epistemic consequences. I argue that precision is an epistemic virtue that mitigates against bandwagon effects and illustrate this claim with a case study from the history of measurements of c . This precision-first reasoning motivates the practice of blind data analysis in bleeding edge precision measurement, where outcomes can point the way to new physics.

1. Introduction

A “Zanzibar” is an island of measurement values that internally cohere, but are detached from independent contact with reality. Contemporary philosophy of measurement has embraced *coherence* as the highest epistemic virtue to which successful measurement can realistically aspire. The criterion of global coherence undermines *local* Zanzibars, demanding that local measurements cohere with background theory, model-based simulations, and alternative measurement techniques. Nevertheless, global coherence without a constitutive contribution from observer-independent reality seems no more than a giant Zanzibar: intersubjectivity unmoored from objectivity. Indeed, biting this anti-realist bullet, contemporary coherentists assert that measurement is “the continuation of theory construction” (van Fraassen 2008, 112), that quantity values are not discovered but “come into being” through measurement (Chang 2004, 213), and consequently that simulations and measurements have the same epistemic standing (Morrison 2009).

I do not think this coherentist picture accurately reflects the epistemic achievements of contemporary physical measurement. The methods of measurement and data analysis employed to determine values for the fundamental physical constants

include a suite of techniques explicitly designed to allow the world to “push back” against the tendency toward coherence, and thereby disrupt not only local, but global Zanzibars. Here, I aim to establish just one such technique, the use of precision-first reasoning to identify and reduce systematic error. This result should be surprising. First, precision per se is rarely discussed as an epistemic virtue, yet I demonstrate it is treated as such by the measurement community. Second, precision is the inverse of random error, and does not *prima facie* concern systematic divergence from “true” value at all. Yet important kinds of systematic error may be transformed into noise, and thereby treated as random and reduced by improving precision. This result is significant, as neither the assessment nor the improvement of precision constitutively involves substantive physical theory; consequently, high precision serves as a theory-neutral mark of measurement success.

2. Whither Zanzibar?

The term “Zanzibar” derives from a cautionary tale: a tourist in Zanzibar asks the retired captain of a ship moored at harbor how he knows his noonday gun is accurate. The captain replies that he sets his watch by the clock in the window of the town watchmaker. Later, as the tourist passes the watchmaker’s shop, she asks him how he knows his clocks are accurate. The watchmaker replies that he sets them by the noonday gun fired each day from the harbor. This parable has been oft repeated by metrologists, but its exact message is unclear. We can identify at least three distinct conceptual problems in the neighborhood.

First is the problem of local versus global units. The watchmaker and captain can arrange to meet at 2 p.m., and their respective timekeeping devices won’t steer them wrong. If, however, the captain wishes not to miss his favorite radio show, he’d better ensure his watch synchronizes with that of the BBC. Petley (1988, 9) calls globally divergent local standards “Zanzibar units,” giving the drift of physical standards, such as the shrinking of the British Imperial yard stick, as instances. It is easy to see how pressure toward global coherency resolves this problem, demanding a *fortiori* globally consistent units.

If we consider measurement as an attempt to determine the values of quantities in the world, however, then a second problem emerges, inverse of the first. For any measurement in the first instance produces a result in local units, which must then be translated, through a complex modeling and calibration procedure, to the final outcome value represented on an intersubjectively available scale (Giordani and Mari 2019; Tal 2017). During this translation procedure, the pressures toward coherence may have subtle effects, resulting in the “tendency for experiments in a given epoch to agree with one another” (Petley 1988, 294). This phenomenon has been called the “bandwagon effect” (Franklin 1986) or “intellectual phase locking” (Cohen and DuMond 1965). Section 4 analyzes bandwagon measurements of the velocity of light between 1935 and 1941, which converged on a value 17 km s^{-1} below the one we accept today.

This brings us to the third challenge posed by Zanzibars: given discrepant measurements, how can we tell whether they are due to systematic error, or require new physics? Do discrepant values of c indicate unidentified sources of physical interference in an experiment, or that the velocity of light is not constant, but changes over

time, decreasing or modulating sinusoidally?¹ In the limiting case, divergent measurements may not even target the same quantity (Tal 2019). If the watchmaker hears the Captain's noonday gun when his clocks read 12:08, he has several options available: (i) take the gun as a standard and rewind his clocks to 12:00; (ii) assume both devices exhibit some error and correct to a weighted mean, e.g. 12:04; or (iii) conclude that the gun is tracking a different quantity. For instance, the ship's captain may have decided to fire his gun whenever his stomach grumbles for lunch.

If there is a principled way to choose between these options, then captain and watchmaker can escape their Zanzibar loop. Any principled strategy must be grounded in virtues, and *coherence* with other aspects of knowledge is surely one such virtue. Another virtue, which historically has played a crucial role in assessing the quality of physical measurements, is *precision*.

3. In praise of precision

"Precision" is used ambiguously in the reporting of experimental results.² Technically, precision is the degree to which repeated measurements by the same procedure agree. Even in technical reports, however, "precision" is sometimes used casually to refer to inverse degree of uncertainty. These uses are not equivalent, because reported total uncertainty includes both contributions from "random" error, that is, statistical effects with an expected value of zero, and those from "systematic" error, that is, uncertainty in corrections made to the reported result on the basis of modeling or theoretical considerations. Only the first of these is a measure of agreement across repeated applications of the same procedure.

Historically, random and systematic errors were treated as radically different in character; the inverse of the first was precision, the inverse of the second, accuracy (Beers 1957). The reason for this is that precision can be assessed in a straightforward way with statistical techniques conceptually independent of physical theory (Isaac 2019). In contrast, analysis of systematic error requires both theoretically-informed imagination to guess at possible physical effects on the apparatus, as well as creativity in quantifying the degree to which these affect the outcome. Thus, the assessment of systematic error is complicated by both "known unknowns"—systematic effects one knows to be present but is uncertain how to quantify—and "unknown unknowns"—systematic effects that one may not have imagined or considered. The relative difficulties here are clearly illustrated in the history of recommended values for the fundamental constants: while the uncertainties around these values monotonically shrink, these values often leap to new centroids, outside the range of uncertainty assigned to earlier recommendations (Henrion and Fischhoff 1986, Figure 1). Ultimately the reason for these leaps is the unknowns of systematic error.

The upshot of this is that random and systematic errors have radically different epistemic status: the first may be evaluated, and indeed reduced, without any substantive appeal to physical theory, while the latter may only be evaluated through the informed application of theoretical knowledge, calculation, and modeling. High precision indicates stability in the physical system of the measurement procedure

¹ These analyses of inconstant c were suggested by de Bray and Edmondson respectively (Hüttel 1940; Henrion and Fischhoff 1986).

² This ambiguity is unrelated to that analyzed by Teller (2013).

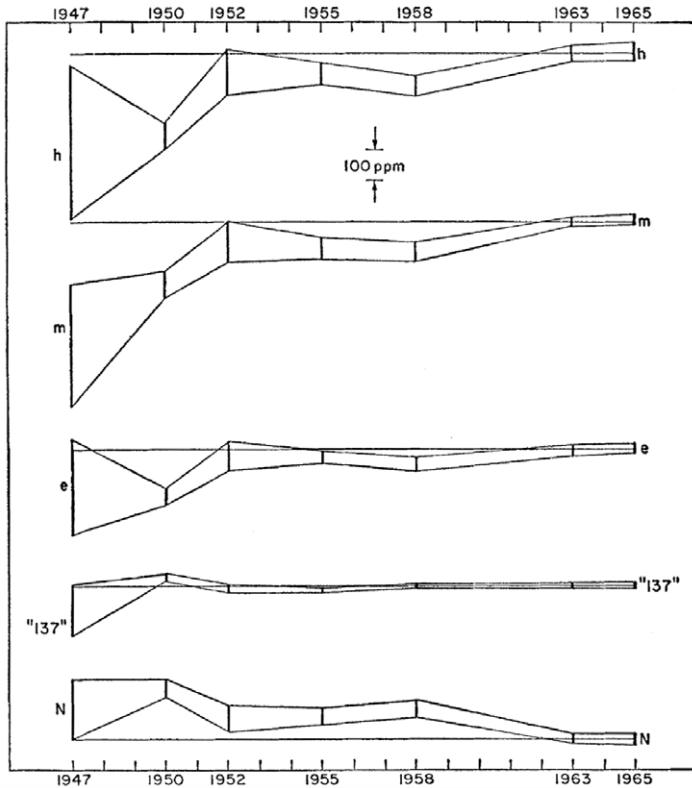


Figure 1. Change in values assigned to physical constants 1947–1965. Reprinted from Cohen and DuMond (1965, 592) by permission of the American Physical Society.

that obtains independent of the theoretical analysis of that system, and knowledge of this stability *itself* constitutes a theory-independent kernel of the overall knowledge we obtain through measurement. From an evidential standpoint, the precision of a measurement provides a lower bound on its overall uncertainty, and therefore the experimentalist’s task is to maximize precision above all else, crystalizing this theory-independent kernel. Systematic errors may always be recalculated or reduced in hindsight, thereby further refining the knowledge from a particular measurement (Birge 1942; Taylor 1971).

Metrologists know this, and they respect and elevate their peers who are able to achieve high-precision results. This is why apprenticeship in a laboratory is crucial for acquiring expertise in high-precision measurement: only through hands-on experience can that pragmatic, un-textbook-able skill of reducing random error be achieved (Ferrario and Schiaffonati 2012, 5.1.1), and without this skill, the creative assessment of systematic effects is irrelevant. In a forum on the least squares method for determining recommended values for fundamental constants, Peter Franken makes the point forcefully: “If in fact you are really limited by the random errors of your measurement you should measure some more. You want to get down to

the point where you *do* have to apply subjective judgments on what you are doing [i.e. assess systematic effects]” (1971, 507). Failure to achieve sufficiently high precision renders moot the analysis of systematic error.

So, increased precision, or the reduction of random error, has a more robust epistemic status than corrections for or estimations of systematic error (Petley 1988, 293). This insight motivates one of the most powerful techniques of precision measurement: the physical transformation of a source of systematic error into statistically “random” error. This technique turns a “known unknown”—namely a known source of error with unknown magnitude—into noise, which may then be reduced through repeated measurements and statistical analysis (292). For instance, in order to measure the gravitational attraction between two 1mm gold spheres, Westphal et al. (2021) needed to factor out the complex gravitational effect on these spheres due to larger moving massed bodies in the vicinity. This effect was systematic, in the sense that its expected value was not zero; yet it was too complex and variable to reasonably calculate, or even estimate, as it involved uncontrollable, and in general unknowable, movements of objects outside the laboratory. The solution was to sinusoidally modulate the position of one sphere, rendering external gravitational effects into background noise, against which a sinusoidal signal, the attraction between the spheres, could be detected.³ This is just a special case of a general technique in gravitational research, namely rotating a torsion balance to induce a sinusoidal gravitational “signal” that can be detected against the complex background “noise” of attractive effects from other massed bodies (Roll et al. 1964). Changing the physical setup of the experiment, by introducing rotation or harmonic modulation, thereby transforms “known” but incalculable external effects into noise.

Contemporary metrology obscures the epistemic difference between random and systematic errors through its practice of treating all errors “identically” in the calculation of total uncertainty (JCGM 2008). To be clear, I agree with this decision and the reasoning behind it wholeheartedly. The critical point for the JCGM is that all uncertainties be justified and commensurate; in particular, they reject the line of response to the phenomenon illustrated in Figure 1 that advocates adding large *arbitrary* uncertainties to account for unknown unknowns. Rather, distinct sources of uncertainty should be collated into an “uncertainty budget” (JCGM 2008; Tal 2017) that permits the separate evaluation of each source of error, whether random or systematic, as is absolutely essential at the bleeding edge of high precision measurement.

4. An island at c escaped

The history of physical measurement contains ample evidence of the evidential power of precision—both, in the first instance, in lending credence to spurious values, and later in the very detection and overturning of these errors. Here we see a cycle, of a seemingly high quality result establishing a bandwagon, which then itself is finally

³ Examples like this illustrate why contemporary metrology recommends uncertainty estimations be classified into Type A (using statistical analysis alone) and Type B (using more elaborate methods), rather than “random” and “systematic”—in these examples, the same cause has been, through design, transformed from producing a systematic effect to producing a random effect.

broken by the achievement of a higher degree of precision. Let's examine one example of this general phenomenon.⁴

One method for calculating the speed of light c , first successfully applied in the mid-19th century, reflects a point light source from a rotating mirror to a fixed mirror and back. In the intervening time, the rotating mirror has shifted enough to displace the image of the light source; the degree of displacement can then be used to calculate c . Albert A. Michelson, of Michelson and Morley fame, made significant advances in the precision with which c could be measured by this method throughout his career, and in 1929 began an ambitious new measurement near Laguna Beach, California. Although Michelson himself died early in the experiment, his collaborators Pease and Pearson carried on, collecting almost 3000 data points over a two year period, and publishing in 1935 a value for c that claimed higher precision than any preceding:

$$299,774 \pm 6 \text{ kms}^{-1} \text{ (Michelson, Pease, and Pearson 1935)}^5$$

Over the next decade, subsequent measurements of c using different methods converged on this value. For instance, Anderson (1937) and Hüttel (1940) both employed a photoelectric Kerr cell to convert sinusoidally modulated light that had been diverted down two paths of different distances into an electrical signal. This signal could then be amplified to determine the phase difference between the two beams of light, and c determined from the different distances they had traveled. These two independent measurements fell close within the expected error of Pease and Pearson (MPP):

$$299,771 \pm 15 \text{ (Anderson 1937)}$$

$$299,771 \pm 10 \text{ (Hüttel 1940)}^6$$

A further refined measurement by Anderson agrees with MPP even more closely:

$$299,776 \pm 14 \text{ (Anderson 1941)}$$

Birge (1942) performed a detailed analysis of prior measurements of light to derive a recommended standard value. He was impressed by the high precision of the MPP and Anderson 1941 measurements, especially the stability they displayed over a very large number of observations, approximately 3000 in both cases. Birge, taking the

⁴ Other examples of bandwagon effects broken by increased precision are the victory of x-ray diffraction over oil drop determinations of e (Birge 1942) and the 8σ discrepancy in the value of η_{+-} measured before and after 1973 (Franklin 2018, 7–8).

⁵ The original publication reported not the standard deviation, but the *average* deviation of ± 11 ; Birge (1942) assigned the uncertainty of ± 6 on grounds discussed below. Throughout, I have suppressed subtleties around methods for calculating uncertainty; these details do not affect the overall trajectory of change in precision illustrated by these examples.

⁶ These values have been corrected from those originally published to account for the group velocity index of light in air, as suggested by Birge and reported by Anderson (1941); the published values were $299,764 \text{ kms}^{-1}$ (Anderson) and $299,768 \text{ kms}^{-1}$ (Hüttel).

convergence of these (and other) different methods into account, ultimately recommended

$$299,776 \pm 4 \text{ (Birge 1942).}$$

Surveys over the next few years, for instance Dorsey's (1944) book-length study, and Warner's (1947) assessment of the velocity of electromagnetic waves in general, seemed to vindicate Birge's analysis.

And yet, this whole pattern of results was overturned. In 1947, Essen measured the velocity of electromagnetic waves in a vacuum using a cylindrical resonator, and in 1948, Bergstrand performed a Kerr cell measurement over several kilometers, both delivering values significantly higher than Birge's recommendation. The turning point came with Carl Aslakson's (1949) publication of radar triangulation measurements which claimed a new significant figure, followed rapidly by further measurements by Bergstrand achieving precision an order of magnitude greater than Birge's estimation.

$$299,793 \pm 9 \text{ (Essen 1947)}$$

$$299,796 \pm 2 \text{ (Bergstrand 1949)}$$

$$299,792.4 \pm 2.4 \text{ (Aslakson 1949)}$$

$$299,792.7 \pm 0.25 \text{ (Bergstrand 1950)}$$

A survey by Bearden and Watts, appearing in January 1951, split the difference between the earlier and later groupings in recommending a value for c . Assessing the data available in 1952, DuMond and Cohen (1953) excluded all values prior to Aslakson's when performing their least squares adjustment to find consistent values for fundamental constants. By 1965, measurements of c had grown sufficiently precise that Cohen and DuMond treat it as an auxiliary constant in their new least squares adjustment. They use the value for c that had been officially adopted in 1957 by the International Scientific Radio Union (Weber 1958):

$$299,792.5 \pm 0.4$$

There are several remarkable points to note about this incident. The first is that the post-1947 values clustered 4 standard deviations away from Birge's estimate, and 3 away from the accepted analysis of the uncertainty in MPP's result. This divergence is remarkable in part because of the seeming convergence in value across several different measurement procedures, including some not discussed here. Reflecting on the affair in 1957, Birge emphasized this point, and its surprising implications: "[T]hese eight results, obtained by six different investigators, using *four* completely different experimental methods, agreed with one another, on the average, *twice* as well as was to be expected on the basis of the probable errors that had been assigned to the individual results [While] all the different methods are subject to systematic errors . . . one would scarcely anticipate that the several final systematic errors should all be in the same direction and of roughly the same magnitude" (50). This led

Birge to a startling conclusion about the origin of bandwagon effects: “In any highly precise experimental arrangement there are initially many instrumental difficulties that lead to numerical results far from the accepted value of the quantity being measured . . . the investigator searches for the source or sources of such errors, and continues to search until he gets a result close to the accepted value. *Then he stops!*” (51, emphasis in original).

In hindsight, it’s hard not to see Birge himself as a contributor to this particular bandwagon. Already in 1934, in a note arguing for the stability of c against Edmondson’s sinusoidal variation hypothesis, and drawing in part on preliminary MPP data, Birge (1934) suggested the value of $299,776 \pm 4$. The accepted standard deviation in MPP’s result, ± 6 , is also due to Birge. The distribution of MPP’s original data, while highly peaked near 299,774, is non-gaussian with long tails. Birge (1942) decomposed this data into a sharply peaked distribution of “good” observations and a relatively flat distribution of “bad” observations, presumably infected by unknown systematic errors. If, instead of taking uncertainty to coincide with the standard deviation in only the “good” data, one takes it to coincide with the standard deviation of the data as a whole, namely ± 13.3 , the current value of c falls the much less “shocking” distance of 1.3 standard deviations away (Cohen and DuMond 1965).

A second aspect of this incident worth noting is the role of engineering in pushing back against theory. In World War II, both US and UK employed radar triangulation to guide aircraft. By sending radar signals between an aircraft and two ground positions, the location of the craft could reliably be predicted from the time of signal return. The British system Oboe was used for blind bombing, allowing, for instance, bombing of German industrial sites in conditions of low visibility. The triangulation calculation includes the speed of light in a vacuum, which is then corrected for the refractive effects of atmosphere and altitude to determine distance the radar signal has traveled. Already during the war, Oboe engineers discovered that the official (i.e. Birge’s) suggested value for c introduced systematic discrepancies, which were largely eliminated by replacing it with the value 299,787.6 km/sec (Hart 1948). Aslakson, working to improve the accuracy of the US system Shoran for geodetic surveying discovered a similar discrepancy.

Aslakson’s (1949) study used Shoran to generate a set of triangulated distances, with a small core of these anchored to ground surveys. Discovering a seeming systematic discrepancy with the ground distances, he rechecked a variety of possible sources of systematic error in the devices and calculations. As a last resort, he reversed the calculations for speed of radar waves to find the best value of c for each length, then took the average, arriving at $299,792.4 \text{ kms}^{-1}$. Remarkably, after this calculation Aslakson discovered that the Army Map Service had undertaken their own independent analysis of this discrepancy. Mary Jane Camplair performed a least squares analysis of the discrepancy between *all* Shoran and map distances to find a constant multiplier for correcting the Shoran results, implying a value for c of 299,792.3. Camplair’s analysis confirmed that the systematic effect varied with distance, lending credence to the identification of c as the source of the discrepancy.

Aslakson explicitly takes the high precision of Shoran measurements in his experiment as indicative of their evidential quality. He stresses the higher “internal consistency” of his repeated measurements than those of Anderson as evidence

his value stands as a legitimate competitor to Birge's recommendation. In the context of discovery, the size of the discrepancy between the Shoran measurements and ground surveys was "shocking" precisely because of Shoran's exceptional precision (480).

More generally, the reported uncertainties in both Aslaksen (1949) and Bergstrand (1950) are entirely uncertainties due to random error (i.e. measures of "precision" in the narrow sense). Indeed, it was this high precision that drove the acceptance of the new value, independent of uncertainty assignments to systematic errors. Bearden and Watts (1951, 74) arbitrarily increased the uncertainty in Bergstrand's result to ± 2 in order to cover uncalculated systematic errors, a practice not adopted in later least squares adjustments, as subsequent, higher-precision measurements converged on the later value, and the pre-Aslaksen measurements were dropped.

A final point about Aslaksen's result. Yes, there is a story to be told here about the value of coherence, for indeed it was the lack of agreement between Shoran measurements and ground geodetic surveys that drove the search for some correcting factor. Nevertheless, this search was anchored firmly in the world, in actual distances on the globe that mattered for actions like effectively guiding aircraft and bombing strategic targets. Vividly, it is closeness of fit with conditions in the world, not with other pieces of theory, that drove this result. And in this case, the pushback of the world against our human actions, the role it played in determining their success conditions, demanded a level of precision beyond that previously attained in the laboratory.

5. The blind against the bandwagon

But how can we ensure the world has power to "push back" in measurements with no immediate implications for action? This problem seems to arise at the bleeding edge of contemporary high-precision measurement, which depends on elaborate laboratory-created phenomena. The systematic effects at stake are so numerous and arcane in their theoretical grounding that the pressures to correct until an expected result is reached "and then stop" may seem unavoidable.

And yet, there are techniques to block bandwagons. One particularly important method "blinds" the data from a high-precision experiment before analysis, e.g. by shifting it uniformly to a new centroid. If the experimenters themselves are blind to the exact shift, they can analyze the data, correcting for systematic errors, in a position of ignorance about the degree to which their value agrees with prior results or not. Blind analysis cannot possibly proceed until an expected value is reached "and then stop." This technique is especially important in cases where high-precision measurement indicates new physics, as it removes the possibility of theoretical bias toward one theory or another.

A recent extreme form of blinding was carried out by the Fermilab group in their measurement of the anomalous magnetic moment of the muon (Muon $g-2$ Collaboration 2021). This measurement is of great theoretical significance, as it shows a 4.2σ discrepancy between measurement and the predictions of the Standard Model, pointing the way to new physics (Arcadi et al. 2022). Fermilab measured the precession frequency of muons trapped in a cyclotron by means of the positrons emitted during weak decay; this direct measurement of precession may be used in conjunction

with knowledge of the stable magnetic field in which the muons are bound to calculate the anomalous magnetic moment. Initial data are digitized by a high-precision digital clock; an outsider offset this clock rate by a small amount to blind the data collection at the hardware level (Muon $g-2$ Collaboration 2021, 5). Then, six further blindings of data at the software level were introduced, for each of six different groups separately analyzing the data (18). These groups applied different statistical techniques to different parts of the data, subject to different systematic errors (6). Only after each group had finalized its results were the software blindings removed, and only after the comparison of results and decision to publish was the hardware blinding finally removed (19). This elaborate procedure protects against both local bandwagon effects across analyses within the group, and any general bandwagon effect with other measurements.

Blinding of this form is conceptually similar to the transformation we discussed above in the case of gravity research. Rotating a torsion balance turns external gravitational effects into noise. Likewise, blinding data turns researcher bias into noise. Such unconscious bias is a “known unknown” in the sense that it is known to exist (from the history of bandwagon effects), but it is impossible to meaningfully quantify. Blinding turns this potential systematic effect into mere noise, the effects of which are reflected in the overall precision of the outcome.

6. Conclusion

Coherence is an important value in contemporary measurement, but the drive toward coherence also poses an epistemic danger. This danger manifests in the form of bandwagon effects. Yet these effects have been escaped, and can be avoided in future, by elevating the value of precision, that is, the reduction of random error. Because the assessment of precision involves purely statistical techniques, this epistemic value is not infected by the contingent features of any particular physical theory. The precision of a measurement indicates the stability of the physical system it comprises, and this stability itself is a sign of fixity in the world, independent of our interests or theory. Moreover, some systematic effects may be transformed into noise, and thereby their reduction subsumed into the reduction of random error. This transformation may even be applied to the drive toward coherence itself, by blinding data, and thereby turning experimenter bias into noise. This technique is especially important for measurements that indicate new physics, as it lends those measurements a kernel of theory-neutral value.

Acknowledgements. This paper has benefitted from the comments of Nora Boyd, Allan Franklin, Adam Koberinski, Teru Miyake, George E. Smith, and Eran Tal. This research was supported by the Alexander von Humboldt Foundation.

References

- Anderson, Wilmer C. 1937. “A Measurement of the Velocity of Light.” *Review of Scientific Instruments* 8: 239–47.
- Anderson, Wilmer C. 1941. “Final Measurements of the Velocity of Light.” *Journal of the Optical Society of America* 31(3):187–97.
- Arcadi, Giorgio, Álvaro S. de Jesus, Tércio B. de Melo, Farinaldo S. Queiroz, and Yoxara S. Villamizar. 2022. “A 2HDM for the $g-2$ and Dark Matter.” Preprint. arXiv:2104.04456v2

- Aslakson, Carl I. 1949. "Can the Velocity of Propagation of Radio Waves Be Measured by Shoran?" *Transactions American Geophysical Union* 30(4):475–87.
- Bearden, J. A., and H. M. Watts. 1951. "A Re-Evaluation of the Fundamental Atomic Constants." *Physical Review* 81(1):73–81.
- Beers, Yardley. 1957. *Introduction to the Theory of Error*. Reading, MA: Addison-Welsey.
- Bergstrand, Erik. 1949. "Velocity of Light and Measurement of Distances by High-Frequency Light Signalling." *Nature* 163(4139):338.
- Bergstrand, Erik. 1950. "Velocity of Light." *Nature* 165(4193):405.
- Birge, Raymond T. 1934. "The Velocity of Light." *Nature* 134(3394):771–2.
- Birge, Raymond T. 1942. "The General Physical Constants as of August 1941 with Details on the Velocity of Light Only." *Reports on Progress in Physics* 8:90–134.
- Birge, Raymond T. 1957. "A Survey of the Systematic Evaluation of the Fundamental Physical Constants." *Nuovo Cimento* 6:39–67.
- Chang, Hasok. 2004. *Inventing Temperature*. New York: Oxford UP.
- Cohen, E. Richard, and Jesse W. M. DuMond. 1965. "Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965." *Reviews of Modern Physics* 37(4):537–594.
- Dorsey, N. Ernest. 1944. "The Velocity of Light." *Transactions of the American Philosophical Society* 34(1): 1–110.
- Dumond, Jesse W. M. and E. Richard Cohen. 1953. "Least-Squares Adjustment of the Atomic Constants, 1952." *Review of Modern Physics* 25(3):691–708.
- Essen, L. 1947. "Velocity of Electromagnetic Waves." *Nature* 159(4044):611–2.
- Ferrario, Roberta, and Viola Schiaffonati. 2012. *Formal Methods and Empirical Practices: Conversations with Patrick Suppes*. Stanford, CA: CSLI Publications.
- Franken, Peter. 1971. "Comments on the Assignments of Experimental Uncertainties." In *Precision Measurement and Fundamental Constants*, edited by D. N. Langenberg and B. N. Taylor, 507–9. Washington, D.C.: National Bureau of Standards.
- Franklin, Allan. 1986. *The Neglect of Experiment*. Cambridge: Cambridge UP.
- Franklin, Allan. 2018. *Is it the 'Same' Result: Replication in Physics*. San Rafael, CA: Morgan & Claypool.
- Giordani, Alessandro, and Luca Mari. 2019. "A Structural Model of Direct Measurement." *Measurement* 145:535–550.
- Hart, C. A. 1948. "Some Aspects of the Influence on Geodesy of Accurate Range Measurement by Radio Methods, with Special Reference to Radar Techniques." *Bulletin Géodésique* 10(1):307–352.
- Henrion, Max, and Baruch Fischhoff. 1986. "Assessing Uncertainty in Physical Constants." *American Journal of Physics* 54(9):791–798.
- Hüttel, A. 1940. "Eine Methode zur Bestimmung der Lichtgeschwindigkeit unter Anwendung des Kerreffektes und einer Photozelle als phasenabhängigen Gleichrichter." *Annalen der Physik* 5(37):365–402.
- Isaac, Alistair M. C. 2019. "Epistemic Loops and Measurement Realism." *Philosophy of Science* 86(5): 930–941.
- JCGM (Joint Committee for Guides in Metrology). 2008. "Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement." https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/
- Michelson, A. A., F. G. Pease, and F. Pearson. 1935. "Measurement of the Velocity of Light in a Partial Vacuum." *Astrophysical Journal* 82:26–61.
- Morrison, Margaret. 2009. "Models, Measurement and Computer Simulation: The Changing Face of Experimentation." *Philosophical Studies* 143(1):33–57.
- Muon $g-2$ Collaboration. 2021. "Measurement of the Anomalous Precession Frequency of the Muon in the Fermilab Muon $g-2$ Experiment." *Physical Review D* 103:072002.
- Petley, Brian W. 1988. *The Fundamental Physical Constants and the Frontier of Measurement*. Bristol: Adam Hilger.
- Roll, P. G., Krotkov, R., and R. H. Dicke. 1964. "The Equivalence of Inertial and Passive Gravitational Mass." *Annals of Physics* 26(3):442–517.
- Tal, Eran. 2017. "Calibration: Modelling the Measurement Process." *Studies in History and Philosophy of Science* 65–66:33–45.
- Tal, Eran. 2019. "Individuating Quantities." *Philosophical Studies* 176(4):853–78.

- Taylor, Barry N. 1971. "Comments on the Least-Squares Adjustment of the Constants." In *Precision Measurement and Fundamental Constants*, edited by D. N. Langenberg and B. N. Taylor, 495–99. Washington, D.C.: National Bureau of Standards.
- Teller, Paul. 2013. "The Concept of Measurement-Precision." *Synthese* 190(2):189–202.
- van Fraassen, Bas. 2008. *Scientific Representation*. Oxford: Clarendon Press.
- Warner, J. 1947. "The Velocity of Electromagnetic Waves." *Australian Journal of Science* 10(3):73–6.
- Weber, Ernst. 1958. "Report on URSI Commission I—Radio Measurement Methods and Standards." *Proceedings of the IRE* 46(7):1354–57.
- Westphal, Tobias, Hans Hepach, Jeremias Pfaff, and Markus Aspelmeyer. 2021. "Measurement of Gravitational Coupling Between Millimetre-sized Masses." *Nature* 591(7849):225–28.