Studies in History and Philosophy of Science 96 (2022) 51–67

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

Pluralizing measurement: Physical geodesy's measurement problem and its resolution

Miguel Ohnesorge

Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, United Kingdom

ARTICLE INFO

Keywords:
Measurement
Coordination
Geodesy
Geophysics
Astronomy
Measurement error

ABSTRACT

Derived measurements involve problems of coordination. Conducting them often requires detailed theoretical assumptions about their target, while such assumptions can lack sources of evidence that are independent from these very measurements. In this paper, I defend two claims about problems of coordination. I motivate both by a novel case study on a central measurement problem in the history of physical geodesy: the determination of the earth's ellipticity. First, I argue that the severity of problems of coordination varies according to scientists' predictive and experimental control over perturbations of the measurement process. Second, I identify a methodology by which scientists can solve hard problems of coordination and gradually increase their predictive control over perturbations. I dub this methodology ‘operational pluralism’ since it is driven by the introduction of alternative measurement operations that involve different physical indicators.

1. Introduction

When conducting derived measurements,[1] scientists infer the magnitude of a theoretical parameter from a set of quantitative indicators. It has been noted widely that such inferences can be affected by an epistemic circularity (Chang, 2004; van Fraassen, 2008; Mach, 1986; Tal, 2017a). Establishing measurements often requires detailed theoretical knowledge about their target parameter, while our theoretical models of that parameter can lack sources of evidence that are independent from these very measurements. As a consequence, many philosophers have argued that justification in measurement takes the form of bi-directional *problems of coordination*.[2] Both our measurement procedures and theoretical models are modified *iteratively* to account for prediction-measurement discrepancies, making them cohere with each other as needly as possible. If they are successfully coordinated, measurements converge within the space of possible outcomes permitted by our best theoretical model of their target and their former disagreement can be theoretically explained.

Coordination is significantly harder to achieve when measuring the parameters of large and partially inaccessible physical systems – the earth being a prime example. Despite growing philosophical interest in the geosciences (Bokulich, 2018, 2020; Bokulich & Oreskes, 2017; Miyake, 2015, 2017a, 2017b; Parker, 2014; Smith, 2007; Watkins, 2021), it remains insufficiently understood how such complicated epistemic conditions affect the dynamics of measurement coordination. In this paper, I analyse a foundational geoscientific measurement problem – the measurement of the earth's polar flattening – to develop two interrelated philosophical arguments. The first argument is conceptual, sharpening the existing epistemological vocabulary in light of my case study. I introduce the notion of *hard problem of coordination* to refer to situations in which scientists can neither predict nor experimentally control the relevant perturbations[3] of the measurement process. My second argument is methodological. I argue that hard problems of coordination can be resolved through a diachronic and iterative methodology that I dub *operational pluralism*. Operational pluralism is distinct from other kinds of pluralism existing in philosophy of science. While existing views focus on the proliferation of theories, models, or taxonomies, operational pluralism denotes a particular methodology in *measurement*, which aims at isolating and anticipating sources of measurement error. Given that hard problems of coordination are found all across physical science, the

E-mail address: mo459@cam.ac.uk.

[1] In what follows, I take for granted that I am talking about derived measurements and drop the “derived” label. There are good reasons to reject the fundamental distinction between direct and derived measures, but this is not the place to argue for them.

[2] While this notion was popularised by Hans Reichenbach (1920, 1932), Mach (1986) first used “coordination” to describe the dynamic relationship between quantity concepts and measurement procedures.

[3] I use “perturbation” in the sense usually ascribed to it in applied mathematics. Roughly, a perturbation is a disturbance of an initial, approximate model of a physical system. Perturbations of the measurement process are physical effects that are not included in the initial model of the measurement process.

https://doi.org/10.1016/j.shpsa.2022.08.011
Received 7 August 2022
Available online 22 September 2022
0039-3681/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

strategy with which geodesists effectively solved their problem should offer epistemic lessons of general methodological value.

In the extensive case study underlying my proposal, I present novel historiographical research reconstructing how geodesists, astronomers, and geophysicists[4] first came to measure convergent outcomes for earth's polar flattening between 1880 and 1924. This marked an immense achievement, solving a prestigious measurement problem that had persisted since the seventeenth century (Ohnesorge, 2021). In 1924, the *International Association of Geodesy* accepted a uniquely parametrised ellipsoid model of the earth, motivated by a convergence between all available measurement procedures and significant advances in controlling them for systematic errors (Torge, 2017, p. 50). As I will show, this convergence was not merely a result of accumulating more data or deriving an accurate model from theory. Rather, it required the use of additional measurement indicators that are subject to different perturbational effects, vindicating the value of operational pluralism.

The plan is as follows. Section 2 provides a historical and epistemological introduction to geodesy's long-standing measurement problem, together with a brief sketch of the models and measures of the earth's figure. Section 3 contains the bulk of my case study. The key takeaway is that operational pluralism was instrumental for measuring convergent values of the earth's polar flattening. In Section 4, I systematise operational pluralism and defend it against possible objections.

2. Physical geodesy and its measurement problem

2.1. A brief history

While geodesy did not become a cohesive discipline until the second half of the nineteenth century, “geodetic” problems were known and studied as such since the seventeenth century. The principal aim of geodesists was to determine the figure, gravity field, and interior constitution of the earth. These tasks all involve deriving a mathematical model of the rotating earth's shape and density distribution and measuring its theoretical parameters. Since problems of coordination involve an epistemic interdependence between theory and measurement, I have to say some words about the mathematical modelling before moving on to the physical measurement practices.

The mathematical “problem of the earth's shape” (Greenberg, 1995) was to derive the earth's general geometric figure and some quantitative limits for its parameters from assumptions about gravitational attraction, the planet's interior density distribution, and its rotational motion. Approaches to the problem assumed that (i) the earth is in a state of hydrostatic equilibrium,[5] and that (ii) its formation can be modelled by treating it as a homogenous fluid. Thus, mathematical geodesy aimed to determine under which conditions homogenous, uniformly rotating fluid bodies whose constituent particles attract according to the inverse-square law of gravity can be in a state of hydrostatic equilibrium.

Newton turned this into a feasible mathematical problem by introducing an empirical parameter representing the approximately 1/288 ratio between centrifugal ‘force’ and gravitational acceleration at the equator. His argued that an approximately ellipsoidal spheroid of revolution with an ellipticity of 1/230 is the only possible equilibrium figure, noting that inwardly increasing densities inside of the fluid could result in larger ellipticities (Greenberg, 1996; Todhunter, 1873, chap. 1) (Fig. 1).[6] Christian Huygens and various French philosophers proposed

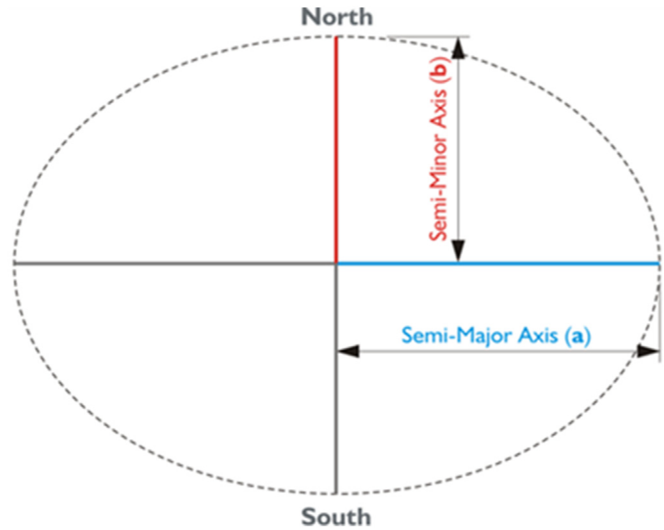

Fig. 1. Meridian ellipse of an ellipsoid of revolution, where a and b are parameters in terms of which polar flattening $\left(f = \frac{a-b}{a}\right)$ is defined. Wikimedia commons.

alternative but ultimately unsuccessful ellipsoidal equilibrium models, which can be derived by replacing Newton's law with their respective theories of gravity, according to which gravity is not a universal force acting on all particles of matter but is directed at the center of planets. Alexis Clairaut articulated the most sophisticated equilibrium theory in 1743, showing that an ellipticity between 1/230 and 1/597 is a necessary condition for uniformly rotating fluid bodies that are composed of different ellipsoidal shells of homogenous density to be in a state of dynamical equilibrium, given that Newton's law applies. Clairaut notably showed that an increasing density towards the center will result in a *smaller* ellipticity, contrary to what Newton had assumed (Chapin, 1995; Greenberg, 1988).[7]

To produce evidence for the universal inverse-square law and the global accuracy of their ellipsoid model, physical geodesists had to measure convergent values for the model's parameters. The central parameter that geodesists tried to determine was the earth's polar flattening, which is equivalent to the model's ellipticity. For this task, two separate measurement indicators had been proposed by Newton, Huygens, and the famous French astronomers Cassini I and II: latitudinal variations in the strength of surface gravity (I_P) and latitudinal variations in the lengths of triangulated meridional arcs (I_T).[8] In both cases, the magnitude of the indicators is assessed at different, astronomically determined coordinates on the earth's surface. After multiple such local measurements, the polar flattening is inferred from the ellipticity of the model which accounts for the latitudinal variations with as little residual error as possible.[9] As Fig. 2 below illustrates, such measurements were “*theory-mediated*” (Harper, 2012; Smith & Seth, 2020). The very definition of polar flattening relied on the theoretically derived ellipsoid model and the amalgamation of different local measurement results assumed their model-conform, elliptic variation with latitude.

Edward Sabine's, 1825 survey of latitudinal surface gravity variations offered overwhelming evidence of a systematic disagreement between

[4] In what follows, I will sometimes refer to the *group* of historical actors in question as “geodesists”, because the measurement of the earth's polar flattening was generally understood to be a “geodetic” problem.

[5] Attempts at defining hydrostatic equilibrium started with Newton and culminated in Clairaut's idea that the net force acting on any fluid channel between two surface points must be 0, which he articulated through Fontaine's novel partial differential calculus (Greenberg, 1995).

[6] In light of several revisions between the different editions of the Principia, the 1/230 value from the third edition is taken as representative here.

[7] The competition between the alternative predictions and the controversy about whether Newton's law is compatible with 18th century measurement results are discussed in Chapin (1995) and Ohnesorge (2021).

[8] Newton also proposed measurements based on the precession of the equinoxes. As I show in detail below, it took until the late nineteenth century for such measurements to be empirically feasible.

[9] This is a simplified description of these two measures For a more detailed discussion see Ohnesorge (2021).

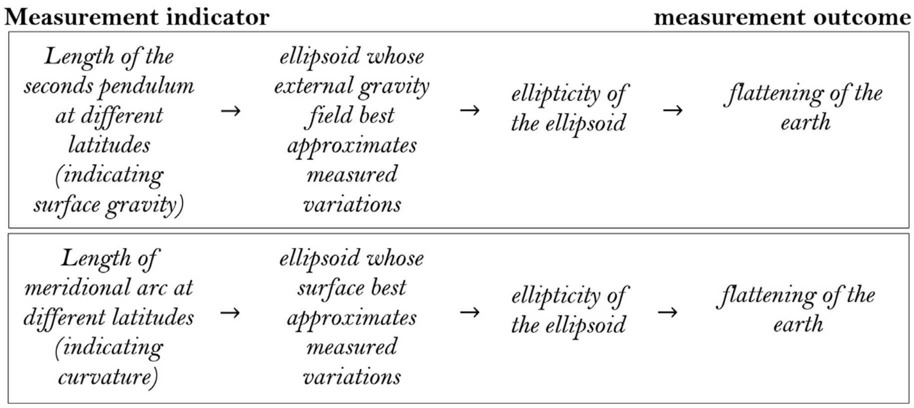

Fig. 2. Measurement inferences from latitudinal variations in pendulum (I_P) and arc lengths (I_T) to polar flattening.

the outcomes inferred from I_P and I_T (Sabine, 1825, p. 341). At the same time, local conflicts between the measured and predicted parameters of surface gravities and curvatures gave evidence for the *local* shortcomings of the ellipsoid model underlying both measurement procedures (Airy 1826; Bessel, 1838; Gauss, 1828; Laplace, 1796, p. 12). Since the theoretical ellipsoid model is presumed by any inference to polar flattening, the causes of the inconsistent outcomes could neither be uniquely attributed to the model nor to any specific measurement procedure. Consequently, the discordance between I_P and I_T confronted geodesists with a problem of coordination.

> **Geodetic Problem of Coordination:** To determine whether the theoretical ellipsoid model could accurately represent the earth's *global* figure despite its local shortcoming, geodesists needed accurate measures. To improve the consistency between their measures, they needed more accurate empirical laws connecting their indicators to a theoretical definition. The available laws, however, relating ellipticity to variations in surface gravity and curvature, were defined relative to the theoretical ellipsoid model (Fig. 3).

Nineteenth-century geodesists were very aware of this problem. Following the abovementioned results, Carl Friedrich Gauss and George Biddell Airy discussed this measurement problem in two seminal papers during the late 1820s. Both of them, moreover, proposed to solve it through the iterative analysis of prediction-measurement discrepancies, hoping to understand the physical sources of errors in the earth's heterogenous subterranean density and, if necessary, replace the ellipsoid model with a more sophisticated successor in a piece-meal manner.[10] Gauss already suggested the outline for a potential successor model, whose operationalisation required geodesists to determine the extent to which the equipotential surface roughly coinciding with mean sea-level deviates from an ellipsoidal reference surface (Ohnesorge, 2021).

Notwithstanding arc and pendulum measurements on a monumental scale in Great Britain, Continental Europe, India, Peru, Scandinavia, and Russia, geodesists did not manage to solve their measurement problem throughout the nineteenth century. Neither was the available data about the distribution and causes of curvature and surface gravity discrepancies sufficient to estimate whether flattening measurements based on the ellipsoid model could, in principle, be made consistent. In the 1880s, I_P-values were still concentrated in the range between 1/298–1/310 while I_T-values were spread between 1/284 and 1/292, showing no significant convergence since the beginning of the century (Strasser, 1957, Appendix, 91–93). As things stood then, (i) the value of geodesists' target parameter was underdetermined (providing inconclusive evidence for the ellipsoid model and Newtonian gravitation) and (ii) the perturbations acting on I_P and I_T were not sufficiently understood. Roughly 200 years after Newton had first attempted to derive a model of the earth's figure from his theory of gravity and scattered empirical data, convergent measurements of its defining parameter were still lagging.

2.2. Epistemological assessment

I take this case to offer *epistemological* insights because persistent discordances in theory-mediated measurements are not unique to polar flattening. Gregory Good has illustrated the difficulty of determining the earth's interior composition based on magnetic measurements (Good, 2011). As Gordon Belot and Teru Miyake note, an analogous problem was faced by geophysicists trying to determine the earth's interior density distribution based on gravimetric and seismological measurements at its surface (Belot, 2015; Miyake, 2017b).[11] All these problems shared two pertinent features, which, loosely building on Miyake's work, can be characterised as follows: (i) scientists do not have empirical access to a parameter without relying on a idealised theoretical model and (ii) their measurements are subject to multiple overlapping perturbations that they can neither predict theoretically nor shield their measurements against (Miyake, 2011). In what follows, I will refer to such situations as *hard problems of coordination*. In the case of geodesy, the unaccounted perturbations resulted primarily from ignorance about the earth's irregular topographic and subterranean density distribution. As we will see later, these perturbations can lead to various systematic errors in different measurement procedures. Examples include large, non-elliptic undulations of the terrestrial gravity field that result in mismatches between data from different regions, strong gravity anomalies that severely affect specific measurements, and asymmetries in the earth's general density distribution that affects inferences from astronomical quantities to the earth's figure.

To be sure, I do not intend to draw any sharp demarcation between standard and hard problems of coordination, nor can I offer necessary or sufficient conditions. My talk of hard problems of coordination aims to pick out cases that require a different methodological treatment, given the respective *degree* to which the perturbations of the measurement process resist predictive and experimental control. This can be

[10] This corresponds to methodologies discussed by Chang (2004, ch. 4), Harper (2012), and Smith (2014), according to which the discrepancies between a theory (or its associated models) and physical measurements are either repeatedly explained by that theory or lead to its iterative revision.

[11] This problem was harder because measurement data was scarcer. Surface gravity measurements are in principle insufficient to uniquely determine the earth's interior structure and seismological measurements can only be conducted after (effectively unpredictable) seismic events.

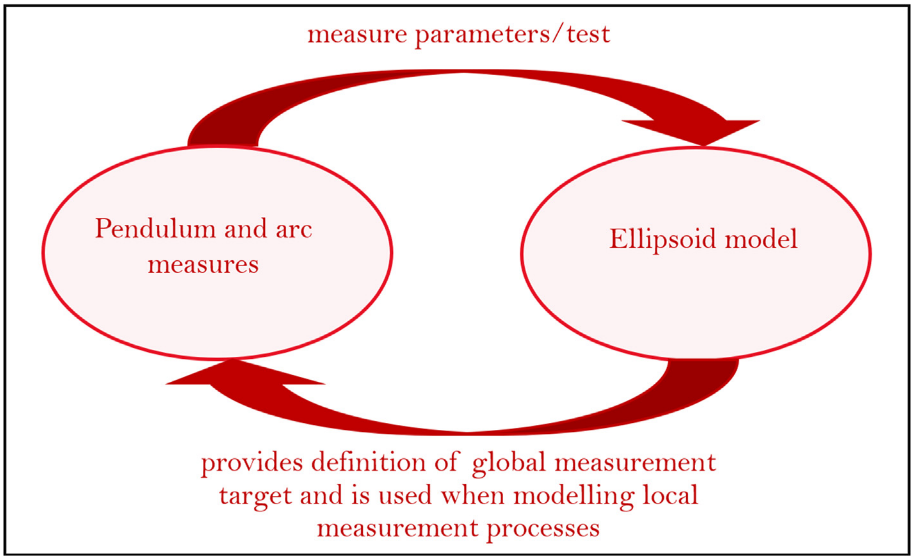

Fig. 3. Epistemic dependence between theoretical ellipsoid model and pendulum and arc measures of the earth's polar flattening.

illustrated by a small juxtaposition to other problems of coordination. Eran Tal gives a nice account of how metrologists coordinate the measurement of the standard second, which acts as the basic unit of the *Coordinated Universal Time* (Tal, 2016, p. 302, p. 326). In the previous *International System of Units* that Tal discusses, the second was defined as "the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom" (BIPM, 2006, p. 113). Coordinating this theoretical definition with measurements is complicated by the fact that any de facto measured state transitions will be subject to background perturbations introduced by (i) gravitational and (ii) magnetic forces, as well as non-absolute-zero temperatures. The crucial differences to our case, and hard problems of coordination in general, is that metrologists are able to *experimentally shield* atomic clocks from several perturbations and *predict* the specific uncertainties arising from the effects of (i) and (ii) in different physical realisations of the ideal caesium 133 atom. This means that their models of the measurement process can reliably anticipate sources of error uniquely affecting specific measurement indicators. Coordination is restored without major difficulties since the operative perturbations were indicated by types of clocks that deviate and the extent to which they deviate. Geodesists, in contrast, did not have sufficient theoretical knowledge about the perturbations resulting from the earth's heterogenous interior and topographic density distribution to identify the unique source of conflicting outcomes. Neither could they dispose of these perturbations by placing the earth in a highly controlled experimental set-up corresponding to the sophisticated machinery of modern atomic clocks. As a consequence, any one of their indicators were unavoidably exposed to multiple perturbation arising from the difference between the physical earth and the idealised model used in the measurement inferences.

My distinction between standard and hard problems of coordination is related to but not identical to a distinction that some philosophers draw between two kinds of epistemic "coordination" in scientific practice. *Measurement coordination* in a narrow sense – and as discussed in this paper – refers to the coordination between physical indicators and models of the target system. This process is also often referred to as "correlation" or "calibration" (Boumans, 2007; Heidelberger, 1994; Tal, 2017b). In a broader sense, models are also coordinated with abstract theories that describe and predict their parameters, sometimes involving stipulative principles (Reichenbach, 1932; Stump, 2015). Put this way, the ellipsoid is a model of the earth that is coordinated (in the broad sense) with Newtonian gravitation based on the principles of planetary equilibrium figures and several idealising assumptions about its rotational motion and internal density. This model, in turn, is coordinated (in the narrow sense) with the measured variations in the length of seconds-pendulums and meridional arcs. As Flavia Padovani and Michele Luchetti rightly point out, the two sense of coordination are often deeply intertwined in scientific practice because we use theoretical assumptions to construct measures, identify errors, and adjust models (Luchetti, 2020; Padovani, 2017). Framed in this vocabulary, hard problems of coordination occur if coordination in the broad sense involves idealised representations of the target system, and the perturbations resulting from these idealisations cannot be experimentally controlled or predicted based on independent theoretical assumptions. As a consequence, scientists are unable to predict or explain why measurement coordination in the narrow sense fails.

3. Operational pluralism as a guide to coordination

Geodesists gradually resolved their measurement problem between 1880 and 1924. On my reading, their success was predicated on following a particular methodology, which I will refer to as operational pluralism in what follows. In this section, I carve out the structure of this methodology by investigating how geodetic practice changed between 1880 and 1924. A good point of departure to understand this period of geodetic measurement is the work of German geodesist Robert Friedrich Helmert, who is widely considered as the "father" of modern physical geodesy. The period in question overlaps with the climax of Helmert's career, much of which was devoted to overcoming the discordances in ellipticity measurements (Reigber, 2017; Torge, 2005). Helmert published his two-volume classic *Mathematical and Physical Theories of Higher Geodesy* in the early 1880s. The book was widely translated and used as teaching resource across the world (Torge, 2009, pp. 237–38). Countering a growing frustration about the remaining discordances between I_P- and I_T-values. In the book, Helmert, argued "that the current practice of geodesists, who treat the geoid[12] as an ellipsoid of revolution […] appears justified" (Helmert, 1884, p. 91). As we will see in what follows, he linked the ongoing discordance to insufficiently understood perturbations of the measurement process, rather than irresolvable flaws in the ellipsoid model or available measurement procedures.

[12] The geoid was the hypothetical successor model that Gauss first imagined in 1828 (see section 2). The Geoid represents an equipotential surface coinciding with mean sea level and offers a closer approximation of the earth's figure and gravity field than any ellipsoid.

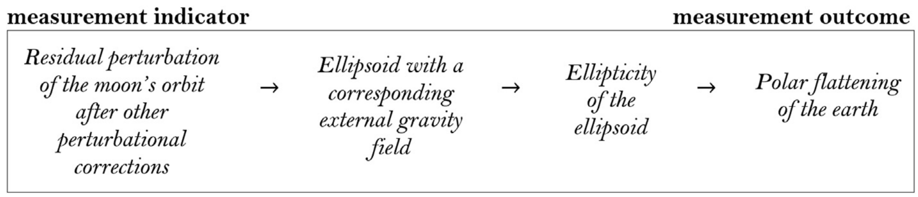

Fig. 4. Measurement inference from lunar perturbations to polar flattening.

Helmert's programmatic claims alone did of course not yet offer any new empirical evidence. As things stood at the beginning of the 1880s, there was no single ellipsoid model consistent with arc *and* pendulum measurements. Consequently, Helmert needed to show that the different measures could be coordinated successfully. In 1886, two years after the publication of his second textbook, he left his former post as Geodesy professor in Aachen to become the new head of the *Royal Prussian Geodetic Institute* (RPGI). He would soon be one of the most influential figures in the discipline, owing to the extent of his theoretical and empirical contributions and the leading role of the *RPGI* in international research. Among other things, the institute hosted the headquarters of the *International Geodetic Association*, with Helmert operating as head of its central bureau (Torge, 2005, pp. 564–65). Now, he had the means at his disposal to provide empirical evidence for his conjectural claims and pursue the long sought-after measurement convergence.

In what follows I reconstruct how Helmert and fellow geodesists, geophysicists, and astronomers, finally achieved convergent measurements of polar flattening. The key to this empirical success was the use of a more diverse range of measurement procedures. As will become clear from my exposition, only the last of the three "new" measures was entirely new. Moreover, the basic assumptions underpinning the first two measurements had been all been discussed in Pierre-Simon Laplace's *Mécanique Céleste*, published between 1789 and 1825 (Laplace, 1832, pp. 853-932, esp. 924-932; 1834, pp. 642-665). In all cases, however, geodesists only managed to conduct empirically informative measurements after further instrumental and perturbation-theoretic advances throughout the nineteenth century. I begin by surveying the different kinds of measurement procedures involved. Most of them were popularised or outlined in Helmert's canonical 1880 and 1884 textbooks, making them a good starting point for us.

3.1. *Astronomical measurements of ellipticity*

Helmert devoted fifty pages of his 1884 textbook to the relationship between the earth's flattening and astronomical quantities. Principally, there were two such quantities with suitable nomic links to the extent of the earth's flattening: the magnitude of a specific pair of perturbation in the moon's orbit, and the magnitude of the earth's precessional constant.[13] Let us denote these two indicators as I_M (magnitude of perturbations ins the moon's orbit) and I_{PC} (magnitude of the precessional constant). To employ them, geodesists needed mathematical expressions of the nomic links between the earth's ellipticity and I_M and I_{PC} that accounted for the gravitational attractions of other celestial bodies. If other phenomena affecting the moon's orbit and earth's precessional constant cannot be sufficiently accounted for by theory, there is no way to isolate the nomic relation of the earth's ellipticity to either the moon's orbit (I_M) or the earth's rotation (I_{PC}). Positively put, the use of both measurement indicators requires well-developed *perturbation theories*. As a consequence, it was only after significant advances in eighteenth- and nineteenth-century mechanics and astronomy that they became attractive for geodesists. Additionally, there was considerably disciplinary inertia among geodesists, so that even seminal nineteenth-century textbooks did not touch upon the relationship between ellipticity and either precession, the moon's parallax, or lunar orbital perturbations (Clarke, 1880; Fischer, 1845, 1846a, 1846b).[14]

3.1.1. *Perturbations of the moon's orbit*

The first systematic attempt to determine the earth's ellipticity from its perturbational effects on the moon's orbit had been undertaken long before Helmert's time, in the third volume of Laplace's *Mécanique Céleste*. Between the 1760s and 1780s, the German astronomer Tobias Meyer and the American astronomer Charles Mason observed a perturbation in the moon's longitude that was proportional to the sine of the longitude of the moon's node. The two nodes of the moon's orbit mark the points at which it comes closest to the earth's equator (Fig. 5) Thus, the effect varied with the proximity of the moon's orbit to the flattened earth's equatorial bulge. In 1783, Laplace found a corresponding latitudinal perturbation and argued that both phenomena can be explained by the impact that the earth's polar flattening has on the earth's gravity field (Chapin, 1995, p. 33). If all other parameters characterising the moon's orbit are known, the magnitude of these remaining perturbations could thus be employed as a measurement indicator for the earth's ellipticity. Alexander Bürg subsequently determined the first numerical values for these longitudinal and latitudinal perturbation coefficients based on his moon tables, from which Laplace derived ellipticity values of 1/304.6 and 1/305.5. Given the contemporary scarcity of empirical data, Laplace considered this sufficiently close to his I_P-values of 1/321.5 and 1/335.8 from the previous volume of the *Mécanique Céleste*, as well as the I_T-value of 1/334.5 that had just been adopted by the French *Commission Générale* (Chapin, 1995, p. 34). For reasons of bookkeeping, the outline of I_M-inferences is noted down in Fig. 4:

In the first three decades of the nineteenth century, some notable astronomers followed Laplace's example (Airy, 1861; Bürg, 1825, p. 15). Most importantly, the head of Gotha's internationally acclaimed observatory and former member of the *European Arc Measurement's* standing committee, Andreas Hansen, published a monumental two-volume study of the lunar perturbations in 1862 and 1864. In it, he made the following two significant discoveries: (i) Other planets of the solar system have notable perturbational effects on the moon that need to be subtracted from the ellipticity-related perturbations, thus decreasing the implied value of ellipticity (Hansen, 1862, pp. 481–97). (ii) A development of the known expression of the perturbation to higher powers entails different values for the ellipticity-related perturbations that account better for the moon's observed orbit (Hansen, 1864, pp. 272–322). In light of these two insights, Hansen determined the longitudinal and latitudinal perturbation coefficients caused by the earth's ellipticity with the highest accuracy so far.

In his 1884 textbook, Helmert derived two values of (i) 1/295.6 and (ii) 1/300.0 from the numerical outcomes that Hansen's gave for the ellipticity-related (i) longitudinal and (ii) latitudinal perturbations. He recommended their mean (1/297.8) as the preferred ellipticity outcome

[13] Johann Albrecht Euler (the oldest child of Leonhard Euler) also considered latitudinal variations in the moon's elevation as a possible astronomical indicator (Euler, 1768). However, geodesists overwhelmingly agreed that it could not be measured with sufficient precision to be serve as indicator of ellipticity (Bruns, 1878, p. 32; Helmert, 1884, p. 460; Todhunter, 1873, p. 447).

[14] The famous and influential exception is Louis Puissant's *Traité de géodésie* (1842a, 1842b), earlier versions of which had been published in 1805 and 1819. Even Puissant, however, only discusses precession and lunar perturbations on a purely theoretical level, drawing mostly on Laplace's earlier work.

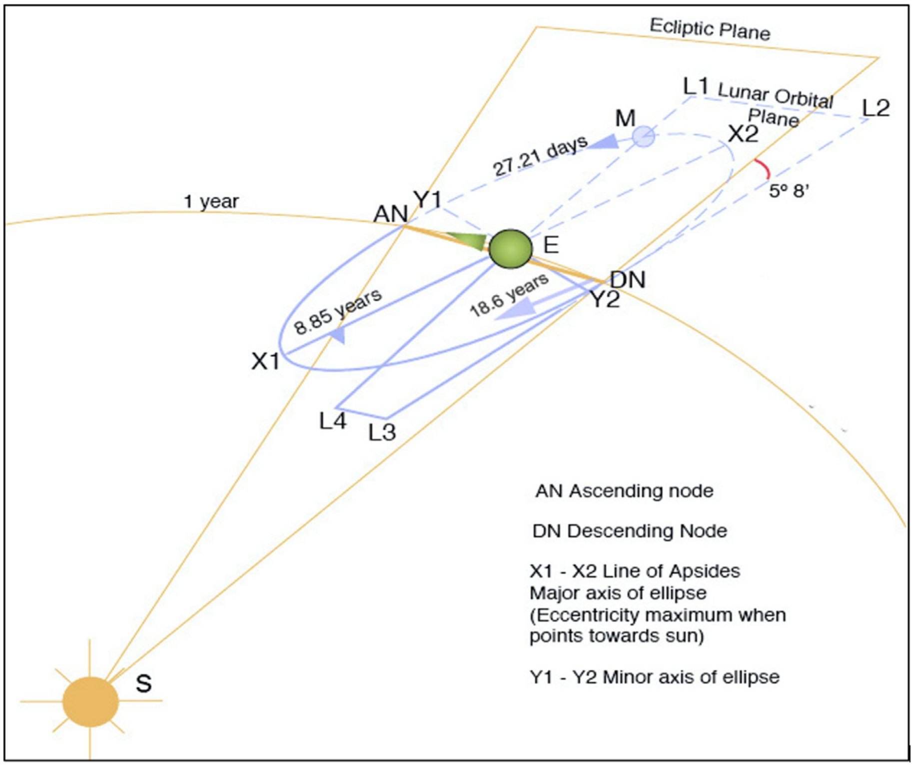

Fig. 5. Sketch of the moon's orbit, illustrating the position of its nodes AN and DN relative to the ecliptic, i.e., the hypothetical plane that intersects with the earth's orbit at every point. Wikimedia commons.

and, in light of the low uncertainties associated with Hansen's results, assigned a mean error of ±2.2. Helmert even remarked that “this mean error estimate, in our view, is rather too large than too small”, given its concordance with the I_P measurements discussed in the volume (see 3.3). Hansen himself had only derived an ellipticity value from the longitudinal coefficient, thus reaching an outcome at the lower end of Helmert's error bar (Helmert, 1884, pp. 468–73). Three years later, the French geodesists and astronomer François Félix Tisserand also employed I_M and derived a flattening of 1/297.2 from Hansen's lunar observations. He presented these results in a talk at the IAG's conference in Paris in 1889, raising further international attention to the geodetic utility of the astronomical measure (Tisserand, 1890, pp. 8–9).

3.1.2. *The earth's precession*

The second astronomical quantity that Helmert reintroduced to the wider geodetic community is the *lunisolar precession*[15] (Helmert, 1884, pp. 426–38). Precession refers to a periodic circular movement in the orientation of the earth's rotational axis relative to the ecliptic (the plane that intersects with the earth's orbit around the sun at every point) (Fig. 6). Astronomers can observe the precession by recording periodic changes in the celestial coordinates of fixed stars. While the effects of precession had been observed for centuries, book three of Newton's *Principia* contained its first quantitative explanation (Newton, 1729, Prop. 3). In line with his theory of the earth's figure, Newton explained the magnitude of precession by appealing to the lunar and solar attraction on the equatorial bulge of the ellipsoidal earth. D'Alambert and Euler's first provided a precise expression to that explanation, using their advances in rigid-body mechanics. Since then, the precession is usually denoted by the precessional constant $C - \frac{A}{C}$, where C denotes the moment of inertia around the earth's equatorial axis and A the moment of inertia around its polar axis (Wilson, 1987). The respective magnitudes of A and

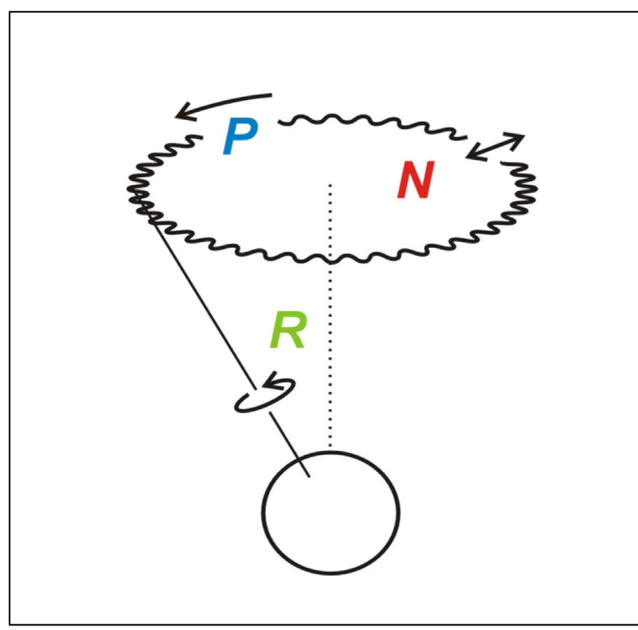

Fig. 6. Sketch illustrating the rotation (*R*), precession (*P*), and nutation (*N*)[16] of a solid body. Wikimedia commons.

C can only be derived with the help of additional hypotheses about the planet's interior density distribution. The difference in moments of inertia indicates the different impact that the luni-solar “drag” has on the rotational motion the earth at different latitudes.

[15] Since only the luni-solar precession matters for our concerns, I will simply refer to it as *precession* in what follows.

[16] The nutation is a wave-like perturbation of the precession that was discovered and mechanically characterised in the mid-eighteenth century (Bradley, 1748, p. 1; d'Alambert, 1749, pp. 73–80).

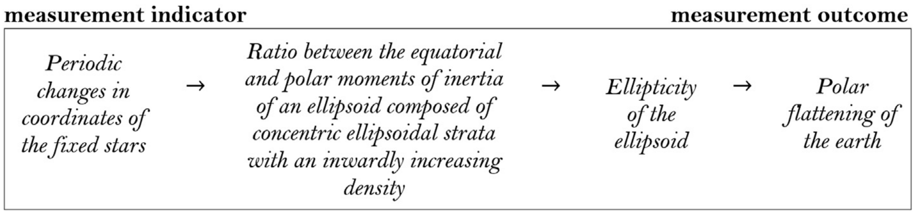

Fig. 7. Measurement inference from precession of fixed stars to polar flattening.

Yet again, the second volume of Laplace's *Mécanique Céleste* was the pioneering work in exploring the possibilities of using the precession as a measure of ellipticity (Fig. 7). Such efforts are complicated by the fact that the magnitude of precession was not uniquely determined by the earth's ellipticity. Strictly speaking, the movement of the rotational axis is not explained by the earth's flattening, but by the inequality between the moments of inertia around its polar and equatorial axis. For a solid ellipsoid with a homogenous density, this inequality varies with the difference in the length of its two axes (i.e., its ellipticity). If applied to the real earth, however, the magnitude of the inequality is affected by heterogeneities in its interior density distribution. While this implies that the contemporary ignorance about the earth's interior affected the reliability of I_{PC}-inferences, Laplace still considered the measure informative. More precisely, he used it to determine an upper ellipticity limit of 1/304. For that, he assumed that the earth's interior consists of spheroidal strata whose density increases gradually from its surface to its core and took the earth's central density to be 4.761 times higher than at sea level. His result significantly narrowed down the ellipticity range (1/230 and 1/578) permitted by Clairaut's hydrostatic theorem alone (Laplace, 1829, p. 930).

Similar to the lunar perturbations, the study of precession returned to the canon of physical geodesy through Helmert's textbooks (Helmert, 1884, pp. 426–38). Yet, he did not revive Laplace's attempt to employ it as a measure of ellipticity. The crucial steps to measure ellipticity from precession were only taken by prominent geophysicists in the late 1880s and 1890s. The two decades marked the beginning of modern seismological measurement and saw a hitherto unknown interest in the earth's interior (Miyake, 2017b; Schweitzer, 2008). First, Paris-based Rudolphe Radau showed that the different internal density variations proposed by Laplace, Helmert, and others, only result in a negligible change in the corresponding moments of inertia and, ipso facto, the earth's precession (Radau, 1885, 1890). Indeed, Octave Callandreau (Paris), Emil Wiechert (Göttingen), and George Darwin (Cambridge) consequently inferred nearly concordant ellipticities while using different hypotheses about the earth's density distribution and only assuming that its interior is composed of concentric ellipsoidal strata, whose density increases towards the center. Their outcomes were: 1/297.4 (Callandreau, 1889, p. 83), 297 (Wiechert, 1897, p. 241), and 296.4 (Darwin, 1899, p. 119). Darwin, in particular, argued at length that I_{PC}-inferences are of a high value "as an independent means to establish the ellipticity of the earth's surface" (Darwin, 1899, p. 123).

3.2. *Deflections of the vertical as a measure of ellipticity*

The fifth and final procedure that came to prominence in the early century measures the earth's polar flattening based on the deflections of the vertical (i.e., the direction of the gravity) across a triangulation network. Deflections of the vertical give quantitative estimates of how much the gradient of the terrestrial gravity field in a certain network differs from that of an ellipsoid. Such deflections are stated as angular quantities and can be determined by comparing locally determined astronomical coordinates at a certain point in a triangulation network with the coordinates that the same point is assigned to an ellipsoidal reference surface fitted to the triangulation network as a whole. To measure the earth's ellipticity based on such deflections, geodesists have to adjust the ellipticity of the reference ellipsoid so that sum of the squares of all deflections in a sufficiently large network is minimised. Areal deflections of the vertical constitute our last and fifth measurement indicator, which we will denote as I_{DV}. Contrary to I_M- and I_{PC}-inferences, this new measure was not connected to astronomical theory but evolved organically from geodetic measurement practice.

After finishing the triangulation of Eastern Prussia in the 1830s, the Prussian astronomer Friedrich Wilhelm Bessel was the first to systematically discuss how to quantify the errors introduced into a triangulation network by scattered gravity anomalies. Let us quickly run through the technicalities. When setting up a triangulation network, you need to orient astronomical telescopes and theodolites. Both are supposed to be fixated on the same equipotential ellipsoidal surface, which stands perpendicular to the direction of surface gravity at every point. If there are any gravity anomalies in the vicinity, the different observation points cannot be projected onto the same ellipsoidal surface. Thus, integrating multiple of such points into one network used in the measurements of ellipticity can introduce systematic errors corresponding to irregular deflections of the vertical throughout the triangulation network. Similar errors occur when multiple triangulation networks are used in an ellipticity measurement, but they cannot be fitted onto the same global ellipsoid (Fig. 8).

In his seminal paper Bessel attempted to effectively quantify and anticipate those errors in a procedure later dubbed *astrogeodetic network adjustment*.[17] While Legendre and Laplace were the first to lay out such a procedure (Legendre, 1805, Appendix; Laplace, 1829, pp. 358–70), Bessel credits his inspiration to Gauss's more recent analysis of a triangulation network in Hanover (see ch. 2). Bessel proposed to multiply the number of astronomical stations across triangulation networks, so that the ellipsoid might be orientated in such a way that the total amount of residual deviations (in latitude and longitude) can be statistically minimised (Bessel, 1837, pp. 295–304). For Bessel, the study of deflection aimed solely at the statistical minimisation of residual error and had no further inferential function. *The Ordnance Survey of Great Britain and Ireland* employed astronomical measurements for the same purpose (Clarke & James, 1858, p. 606).

The eventual move that turned these network adjustments into a novel measure of ellipticity was taken on the other side of the Atlantic, by the American geodesist William Hayford, chief Computer of the *US Coast and Geodetic Survey*. The mammoth project involved 509 (!) astronomical control stations spread across the entire North American triangulation network and aimed to test the isostatic compensation of the region. Remember that theories of isostasy entail that there are significantly fewer surface gravity irregularities than previously assumed since topographic mass surpluses were compensated by subterranean density deficits. The result of Hayford's network adjustment offered the most powerful and detailed evidence for regional isostatic compensation available worldwide. In his calculation, he compared how effective different hypotheses about the earth's subterranean density distribution were in minimising residual deflections. He not only concluded that the US-American continental crust is isostatically compensated but also proposed an exact

[17] This name continues to be used by twenty-first century geodesists. Individual stations that control for azimuth, latitude, *and* longitude are often referred to as "Laplace points": Torge (2001, p. 10).

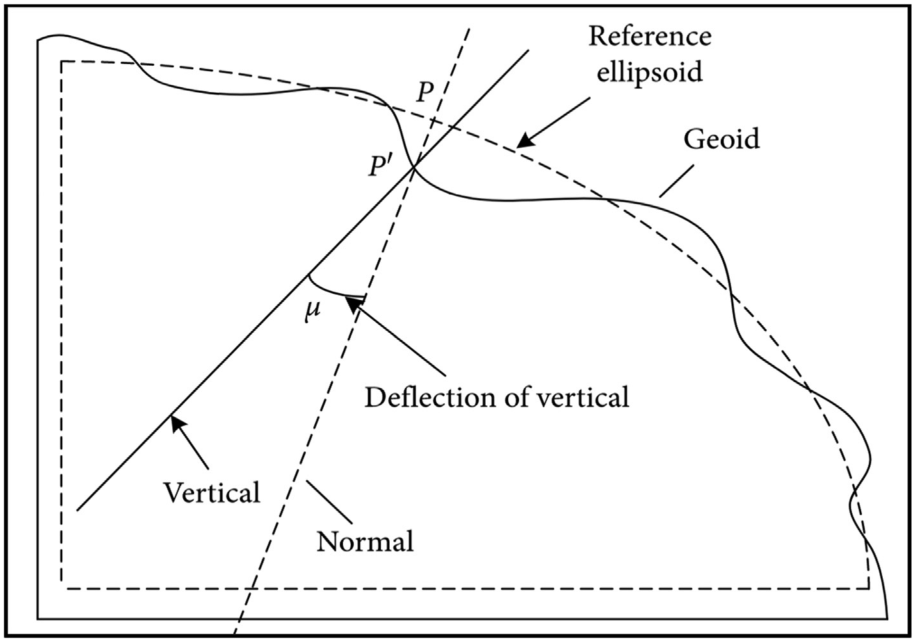

Fig. 8. Illustration of the deflection of the vertical by a gravity anomaly, where "anomaly" refers to any departure from the best-fitting elliptic meridian. The "geoid" represents an equipotential surface that is perpendicular to the direction of gravity at any point.

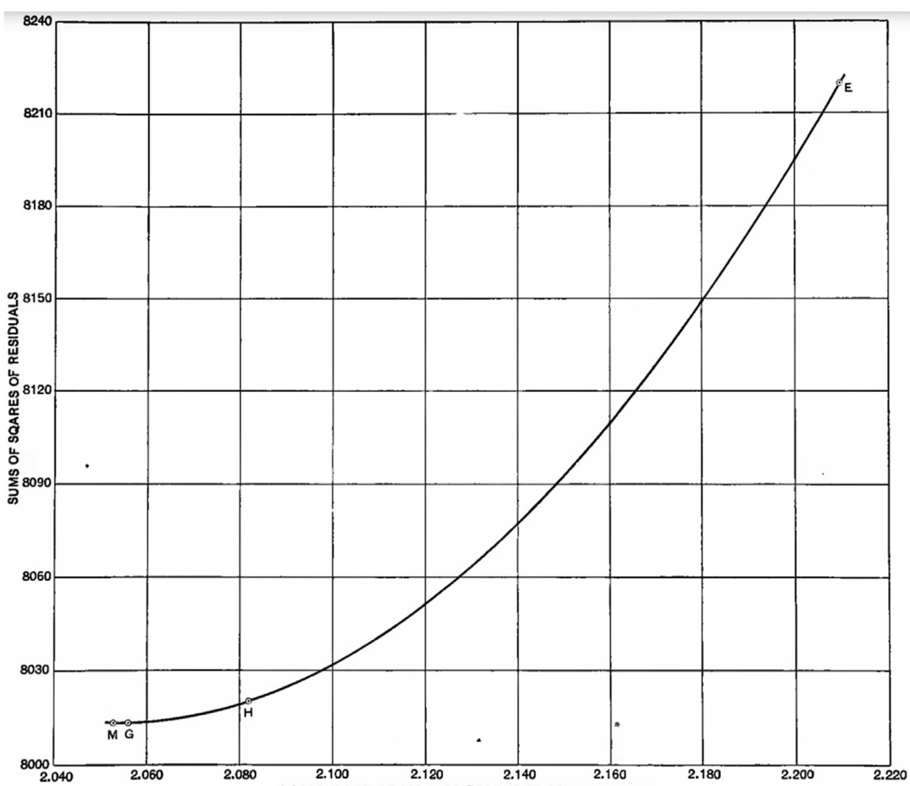

Fig. 9. Hayford's correlation of isostatic compensation depth and deflection residuals, showing that least-squares error minimisation is achieved with hypothesis G (113.8 km). Taken from: John F. Hayford, *The Figure of the Earth and Isostasy from Measurements in the United States* (Washington 1909), 145.

equilibrium depth (113.8 km). This procedure visibly outperformed corrections for topographical mass surpluses and alternative isostatic compensation depths in minimising residual deflections (Fig. 9).

Notably, Hayford's new measure of ellipticity presumed these isostasy results both in content and method. If similar compensations exist across the world, isostatic reductions could guide the astronomical orientation of triangulations and safeguard I_T measurements from systematic errors. Taking this thought one step further, an ellipsoid that accounts best for regional areal deflections after applying isostatic reduction can be expected to do so all across the earth's surface (Fig. 10). Thus, Hayford argued, a single very large triangulation network, such as the Northern American, is sufficient to reliably measure the earth's ellipticity. To do so,

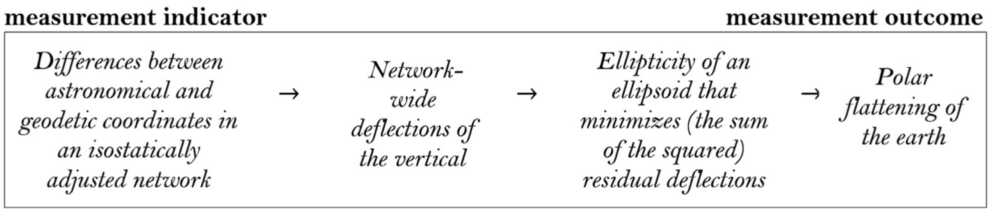

Fig. 10. Measurement inference from areal network adjustments to polar flattening.

geodesists need to adjust the parameters of an ellipsoidal reference model in such a way that the remaining deflections across an isostatically compensated and suitably extensive triangulation network are minimised. Hayford gives an admirably clear illustration of this new method and the epistemic benefits it offers, comparing it to manual model-making:

> The area method is illustrated by supposing that the model maker is given a piece of sheet metal cut to the outline of the continuous triangulation, which is supplied with the necessary astronomic observations, and accurately molded to fit the curvatures of the geoid as shown by the astronomic observations, and that he is then requested to construct the ellipsoid of revolution which will conform most accurately to the bent sheet. (Hayford, 1909, pp. 169–70).

While Hayford rightly stressed the benefits of determining ellipticity through his procedure, his promise of superior accuracy has a strong conjectural element. Generalising from only one isostatically adjusted network in Northern America presumes that similar isostatic compensations exist across other continents.

3.3. Reasoning with multiple measures

Having surveyed the new measurement procedures available at the end of the nineteenth and beginning of the twentieth century, we can attend to how they contributed to resolving physical geodesy's hard problem of coordination. Recall that the severity of problems of coordination varies according to scientists' ability to predict or experimentally control the perturbations arising from the shortcomings of their model of the measurement process. In what follows, I show that increasing the number of physically distinct measurement indicators plays a crucial role in resolving hard problems of coordination. In particular, they allow scientists to isolate previously overlapping perturbation by investigating their different impact on physically distinct measurement indicators. I begin by giving a descriptive account of geodesists methodology before systematising it in the subsequent section.

It is illustrative to start with Helmert's textbooks once again. Beyond the comprehensive overview of the different physical quantities relating to the earth's polar flattening, he had proposed a new I_P-value. Recall that throughout the nineteenth century, the pendulum values were always significantly larger (~1/288–1/290) than the ones inferred from meridional arcs (~1/297–1/300). Controlled pendulum measurements had enjoyed a higher epistemic standing than triangulations since they were less likely to be affected by unnoticed deflections of the vertical (Bruns, 1878; Fischer, 1868). As we have seen, Helmert could now compare both procedures to the ellipticity value of 297.8 $\pm$ 2.2 determined through the lunar perturbations (I_M) and the ellipticity limits implied by relationship between the earth's ellipticity and its precession (I_{PC}). He noted that I_T, I_M, and I_{PC} converged quite closely while disagreeing with the pendulum values (Helmert, 1884, viii). This motivated Helmert to suspect that the supposedly superior pendulum measurements have been missing the target, and articulate hypotheses about which unique sources of error might explain this discordance.

Helmert proposed two crucial sources of error to explain the discordant pendulum measurements: the effects of subterranean compensation and the insufficient distribution of pendulum stations. Building on these hypotheses, he proposed two new systematic error corrections: (i) a different altitude reduction procedure, and (ii) a more evenly and widely distributed selection of pendulum stations, leading him to an alternative I_P result of 1/299.6 (Table 1). While (ii) is self-explanatory, (i) deserves some explication. For the purpose of measuring ellipticity, pendulum stations on the earth's irregular surface always needed to be reduced to mean sea level, which was supposed to be approximated by a smooth ellipsoidal surface. Helmert stopped using the standard "Bouguer reduction", in which the raw pendulum measurement outcomes at higher altitudes were corrected for the supposed surplus attraction of the additional topographical mass between them and mean sea level. Rather, he applied a "condensation reduction", where higher stations are reduced to mean sea level as if the topography between measurement altitude and target surface would be "condensed" into the target surface (Helmert, 1884, 2: 225). This explanation for previous errors and the corresponding new correction are linked to Helmert's belief in global isostastic compensation. The still quite conjectural theory of isostasy implied that "the effects of the continental masses are more or less compensated by a lower density beneath the earth's crust" (Helmert, 1884, p. 364).[18] Results supporting the (at least local) existence of isostasy had been recorded in the Himalayas, Caucasus, Harz, Pyrenees, and close to Moscow (Helmert, 1884, pp. 378–79). If isostastic compensation holds globally, it would explain why previous reduction corrections for topographic surplus attraction had introduced systematic errors.

Notably, Helmert admitted that all three indicators are subject to nontrivial forms of error. Inferences from the moon's orbit are subject to many other theoretical corrections, meaning that their reliability depended on the correctness of a whole class of astronomical background assumptions.[19] His pendulum (I_P) value derived from 160 stations overwhelmingly concentrated in the Northern Hemisphere and South Asia, potentially ignoring large regional variations. His new reduction procedure for the different pendulum stations was also still conjectural, presuming *suspected* isostatic compensation effects. Finally, inferences from the precessional constant (I_{PC}) relied on a conjectural model of the earth's interior as consisting of concentric spheroidal strata whose density increases inwardly.[20] Hence Helmert did not argue for the superiority of particular measurement procedures. Rather, he used their convergence to motivate a new *hypothesis* for explaining the discordance of I_P. He subsequently justified this hypothesis by showing how its application to the I_P data led to value that aligned with the other measures much more closely:

> My value receives very good confirmation from the lunar perturbations and the precessional constant. One would have to seriously change the latter to make it consistent with the 1/289 flattening values that are accepted by some, while likewise subscribing to the existence of a density law for the earth's interior in form of a simple power series since this law cannot exist with such a flattening and the observed value for the precessional constant (Helmert, 1884, viii).

US Coast and Geodetic Survey's astronomer William Harkness approached the problem in a similar manner in 1891. He assembled the

[18] For more detailed discussions of the origins of the theory of isostasy see: Oreskes (1999, ch. 1), Howarth (2008), and Ohnesorge (2021).

[19] For a detailed discussion see Airy (1845, p. 237).

[20] For a detailed discussion see Airy (1845, p. 236) and Darwin (1899, pp. 120–23).

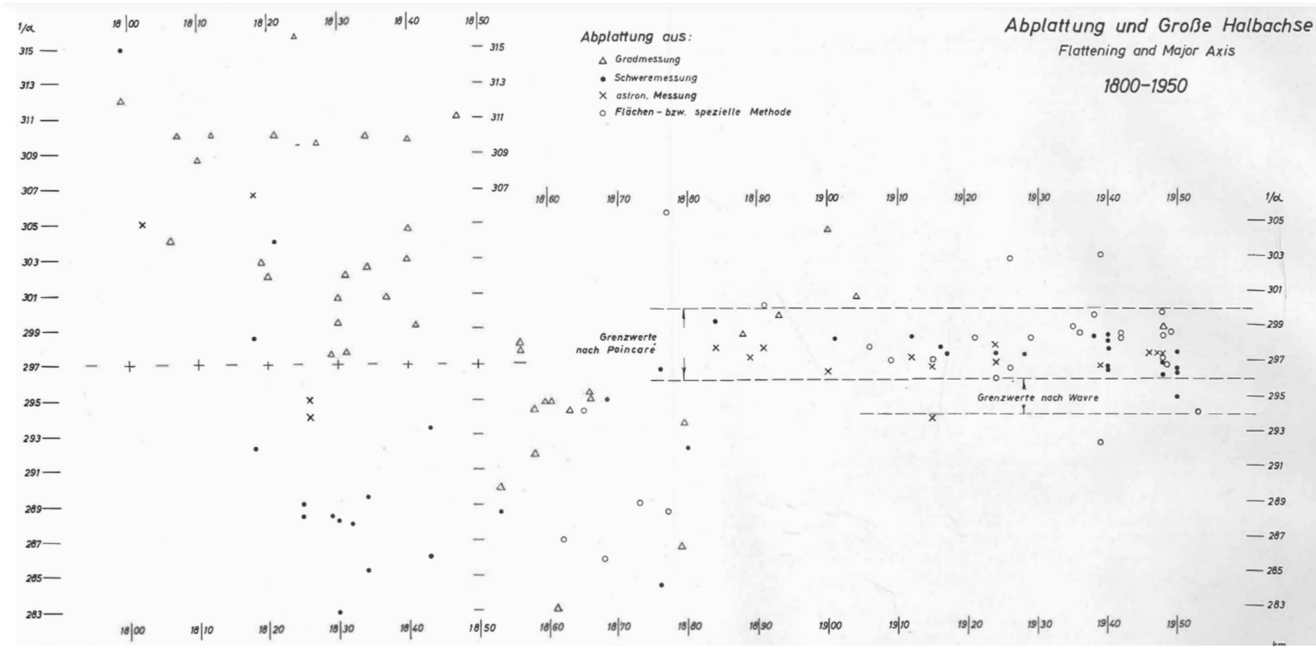

Fig. 11. Inverse-flattening/time graph illustrating the conflict and post 1880 convergence between measurements of polar flattening, in which, '•' denotes an I_P, 'Δ' an I_T, 'x' an astronomical (I_M or I_{PC}) outcome, while '₀' groups together any outcomes obtained from I_{DV} and other non-standard triangulation measurements. Reproduced with permission from: Strasser (1957, Appendix, 95).

Table 1
Most important I_P-values published before and in Helmert's textbook, showing clearly how he departed from earlier results. Taken from: Georg Strasser, *Ellipsoidische Parameter der Erdfigur*, Munich 1957, Appendix.

No.	Year	Inverse Ellipticity	Researcher
1	1818	292.3	Ilmari Bonsdorff
2	1825	289.1	Edward Sabine
3	1829	288.45	Eduard Schmitt
4	1830	288.1	George Biddell Airy
5	1832	288	Nathaniel Bowditch
7	1834	285.26	Francis Baily
8	1843	286.1	Henrik G. Borenius
9	1868	294.1	Philipp Fischer
10	1874	284.4	Amandus Fischer
11	1880	292.2	Alexander R. Clarke
11	**1884**	**299.26**	**Friedrich R. Helmert**

most widely discussed I_P, I_T, I_M, and I_{PC} outcomes of the last decades and inferred a best estimate via the method of least square. His goal was not merely to find the best estimate, however, but to identify how much the different measures contributed to the total probable error associated with the result. Harkness noted that I_P values (around 1/289) as well as the only I_T outlier above 1/297 were responsible for more than half of the probable error. After excluding these values, he determined 1/300.20 as the best estimate, with a probable error of ±2.96. Like Helmert, he thus used convergence in light of different sources of error to argue against discordant values. In his case, the target is an older and widely received I_T value from British geodesist Alexander Ross Clarke:

> In short, the general adjustment, the pendulum experiments, and precession and nutation give a flattening differing little from 1:300, the result from lunar perturbations is uncertain within rather wide limits,[21] and Clarke's geodetic arc gives 1:293.5. Thus, it appears that the geodetic value stands quite alone, and **as it is almost certainly erroneous**, the probable error of the observed value of *e* could be largely diminished by making it depend solely upon the results from pendulum experiments and precession (Harkness, 1891, p. 143).

Harkness himself did not discuss which possible source of error could explain Clarke's individual I_T outlier. Fortunately, Helmert only offered such an explanation in 1884: the arc data used in his calculation was unduly centred on Britain, Central Europe, and the Indian subcontinent (Helmert, 1884, 2: vii).

The convergence towards ellipticities around 1/297 (approximating the long-standing of I_T values discussed in section 2) and Helmert's explanations of previous discordances were further substantiated when Callandreau's, Wiechert's, and Darwin's results from the precession inferences (I_{PC}) were published during the 1890s. As we have seen, these all fell into 1/297 ± 0.4. Darwin, again, explicitly highlights the convergence between I_{PC} and the other indicators, as well as their different sources of possible error:

> The precessional constant could be used to determine the ellipticity of the earth with perhaps as little error as any other method. The uncertainty is, indeed, **of a different kind**, being dependent on our ignorance of the interior of the earth. […] This estimate of the ellipticity agrees well with the results of all other methods, except that of the pendulum, from which it is concluded that the ellipticity is about 1/299 (Darwin, 1899, p. 123).

Note that Darwin already takes for granted that the older pendulum values of about 1/289 are not worthy of consideration, agreeing with Helmert and Harkness. The "disagreement" that he mentions is the slight departure from Helmert's new I_P-outcome that we discussed above (1/299.26).

This last residual conflict was alleviated when Helmert published a new analysis of 1600 (!) pendulum stations in 1901, explicitly responding to Darwin and Wiechert and determining an ellipticity of 1/298.3 ± 1.1. In response to Darwin's and Wiechert's work he had proposed three joint hypotheses to explain the remaining deviation, which he tried to correct for in his calculation. Beyond drastically increasing the number and geographic distribution of stations, Helmert had now developed the

[21] As mentioned earlier, Harkness's did not consider both of Hansen's perturbations, while Helmert and Tisserand did. I suspect that this results from a lack of available data in the US at that point, which I cannot substantiate yet. With both perturbations in mind, Harkness's result would have been even clearer.

Table 2
Overview of convergent outcomes towards the beginning of the twentieth century. Values that were disqualified based on comparisons across measurement procedures and the causes of their discordance are italicised.

Indicators	Potential Sources of Systematic Error	Inverse Ellipticity	Computer	Year
Latitudinal Variations in Pendulum lengths	*Concentration in specific regions, altitude reduction, island, and coast anomalies*	*~289*	*mean value*	*1820–80*
		299.6	Robert F. Helmert	1884
		298.3	Robert F. Helmert	1901
Latitudinal Variations in Lengths of Meridional Arcs	Deflections of the vertical during orientation, *concentration in specific regions*	299	Friedrich W. Bessel	1838
		300	George Everest	1842
		298	Henry James	1858
		293.5	*Alexander R. Clarke*	*1878*
Lunar Perturbations	Data collected over large periods of time and at different observatories, unaccounted perturbations of the lunar orbit	297.8	Robert F. Helmert	1884
		297.2	Felix Tisserand	1887
Precessional Constant	Heterogenous interior density distribution of the earth	297.4	Octave Callandreau	1889
		297	Emil Wiechert	1897
		296.4	George H. Darwin	1899

function describing the latitudinal surface gravity variation on an ellipsoid to a higher order, and predicting errors resulting from the unique mass composition around small islands (Helmert, 1901, pp. 331–36). At the beginning of the twentieth century, geodesists had thus reached a first tentative consensus on the earth's polar flattening, involving not only two but four convergent measurement procedures with different sources of error. The following table tries to sum up these developments more concisely, highlighting which values were successively disqualified for departing from the approximate range in which the different measures converged (Table 2).

In 1906, finally, Hayford applied his new measurement procedure based on I_{DV} and arrived at an ellipticity of $1/297.8 \pm 0.9$. After the further extension of the North American triangulation, he corrected the value to $1/297.0 \pm 0.5$. As we have seen, Hayford went much beyond Helmert in accounting for the effects that subterranean density distributions have on geodetic measurements. He not only stopped correcting for topographic surplus masses but established an exact depth at which irregularities earth crust and mantle are fully balanced out. The success of such a determination across North America added further weight to Helmert's account of the errors in earlier pendulum measurements. Hayford, again, did not justify his result by pointing out the superior reliability of his measure. As we have seen, Hayford's procedure relied on the assumption that the North American isostatic compensation could be generalised across the entire earth. Indeed, Hayford points out that "it is important to note the close agreement [of Helmert's value] with the C. & G. Survey 1906 value for the reciprocal of the flattening, namely, 297.8 ± 0.9 […], though the two values depend on different kinds of observations made in different parts of the earth" (Hayford, 1909, p. 173).

Hayford's results meant that all major ellipticity measurements since 1884 – involving four different indicators – had converged in the range of 1/296.6 to 1/298.3. Thereby, geodesists had successfully reduced the maximum divergence between measured inverse ellipticities from about 15 across two measurement procedures to about 1.7 across four measurement procedures. Notably, the currently accepted value for the earth's polar flattening is 1/298.257223563 (WGS 84) and falls within that convergence interval. While geodesists had not conducted new arc measurements of ellipticity between 1884 and 1909,[22] several of the most important arc measurements from before 1884 also align closely with the new consensus (Table 2). After facing an underdetermined choice about the appropriate flattening magnitude and uncertainty about the adequacy of their model before the 1880s, geodesists had now identified a unique convergence interval and established well-supported hypotheses for explaining earlier discrepancies (Fig. 11).

After the end of the first world war, the IAG's 1924 general assembly in Madrid formally declared Hayford's outcome as the polar flattening of the first global standard ellipsoid. During the assembly, George Tyrrell McCaw, secretary of the British Colonial Survey and Geophysical Committee, presented a least-squares analysis of the major ellipticities measured through different meridional arcs (I_T) and informally compared the outcome ($1/296.2 \pm 1.3$) with the other three indicators. Unfortunately, McCaw did not actually include the other indicators in his least-squares analysis and solely weighed the different arcs according to their length, not considering their individual error sensitivity (McCaw, 1924). After a controversial discussion on the scientific and practical purposes of the new standard model of the earth's figure, the committe instead accepted Hayford's 1909 results - completely corrected for isostasy and in closer agreement with the other indicators - as the model's defining parameters (Dundas, 1924; Lambert, 1926). This marked the first international consensus on the accuracy and appropriate quantitative dimensions of the ellipsoid model of the earth. In 1930, another IAG general assembly in Stockholm settled on a corresponding global standard equation for the variation of surface gravity with the latitudes of the Hayford ellipsoid (Torge, 2017, p. 50). To see how significant this achievement was, keep in mind that the IAG meetings had settled a scientific problem that had been discussed since the *Principia* and provided important evidence for Newton's universal inverse square law. As Samuel Oppenheim, professor of Astronomy at the University of Vienna, notes in his contribution to the seminal *Encyklopädie der Mathematischen Wissenschaften*:

> To determine the exact validity of the Newtonian law based on the distance between attracting bodies, the communicated [geodetic] results matter in multiple ways. The calculation of the earth's flattening from pendulum measurements on its surface and its comparison with geodetic [triangulation] measurements do not show a full agreement but fall sufficiently into their respective error limits. […] The perturbation of the earth's flattening on the movement of the moon is also nearly fully anticipated by the inverse-square law of attraction (Sommerfeld & Oppenheim 1926, pp. 122–23).

We should add to Oppenheim's analysis that geodesists did *not only* manage to make the different measurements converge within a narrow interval that is compatible with the inverse square law of gravitational attraction. Rather, they also identified plausible physical sources behind earlier discrepancies – most crucially the isostatic compensation of the topographic attraction on pendulum stations. These achievements

[22] The new ellipsoid parameters that some geodesists derived had only aimed at regional best-fit surfaces for the Russian and Australian survey, not at a globally adjusted ellipsoid. See: Strasser (1957, pp. 55–57).

offered substantial evidence for Newton's law used in derivations of the ellipsoid model and settled the most prominent measurement problem in contemporary geoscience, vindicating the pluralistic methodology followed by geodesists.

4. Articulating and defending operational pluralism

4.1. Methodological outlines

Above, I have sketched how geodesists solved their hard problem of coordination, reaching a convergence between measurements of the earth's polar flattening and explaining previous errors in pendulum and arc measurements. As I put the matter above, problems of coordination are *hard* if (i) scientists do not have empirical access to a parameter without relying on a idealised theoretical model, and (ii) their measurement indicators are subject to multiple overlapping perturbations they can neither predict theoretically nor shield their measurements against. After struggling with two discordant measurement procedures for two centuries, geodesists finally overcame their measurement problem by introducing additional measures based on different physical indicators. In what follows, I distil a distinctive methodology from geodesists successful approach, which I refer to as operational pluralism. After systematically articulating operational pluralism, I show how it applies my descriptive of geodetic practice and discuss its epistemological significance in light of canonical objections.

The methodology followed by geodesists resembles a three-step heuristic that is repeated throughout multiple iterations:

(1) Introduce measurement procedures with a physically distinct indicator that you believe to be subject to different sources of systematic error.
(2) Identify which measures cause outliers from a shared convergence interval.
(3) Analyse the perturbations uniquely affecting the discordant measure to explain and anticipate the lack of coordination.

Introducing different kinds of measures enlarges the kinds of measurement results which scientists can intercompare. If you compare a greater number of different measures, you increase your chance of having some of them converge in a sufficiently narrow interval. As a consequence, you can identify outliers from that interval. These become the targets of further inquiry that aims to identify their unique sources of error. If two existing measures M_1 and M_2 conflict, and one or more additional procedures $M_3 - M_n$ agree with M_2, the above methodology urges you to further inspect the sources of error associated with M_1. By introducing M_3, we have moved beyond an underdetermination between conflicting parameter values and can articulate and investigate a hypothesis about the operative source of error. This hypothesis is then used to predict errors in further iterations of the methodology and assessed based on how well it increases future measurement coherence. In the best possible case, the adjusted convergence interval leads to the identification of new outliers, which again motivate new hypotheses about the corresponding sources of error. Coordination is achieved iteratively, as hypotheses about sources of error and initial numerical convergence intervals are justified based on their contributions to this process of mutual adjustment. Epistemologically, operational pluralism thus remains firmly rooted in the iterative model of justification originally developed by Chang (2004, ch. 5) and subsequently adopted by most epistemologists of measurement.

The application and utility of operational pluralism depends on several indicators being sufficiently different in their physical constitution and resulting contextual applicability. It is for that reason that I dubbed it "operational", whereby I refer to the physical operations that scientists perform determine the magnitude of a particular indicator. In our case, geodesists relied on such diverse practices as triangulations or pendulum networks stretching hundreds of kilometres, equally extensive networks of astrogeodetic control stations, or dozens of telescopes across several observatories. All of these measures involve sophisticated operations (or systems of operations, if you like) that are affected by the earth's figure and constitution in subtly different ways. These physical differences initially motivate the hypothesis that a new procedure might be sensitive to different sources of systematic error (step 1). Theoretical or adhoc hypotheses about the specific error sensitivities of the discordant measure are then used to explain the detected discordances (step 3) and anticipate them in further measurements (step 1 in the next iteration). Both numerical convergence intervals and hypotheses about distinct sources of error are justified diachronically, based on their contribution to subsequent iterations steps 1 to 3.[23] Contrary to canonical instances of epistemic iteration (e.g. Chang, 2004, ch. 4) iterative corrections are pursued horizontally between different measurement procedures and associated hypotheses about indicator-specific sources of error, not necessarily involving vertical iterations between procedures and general theoretical models of their target.

How does this map onto geodetic practice? As we have seen, geodesists started to strongly rely on additional measures from around 1880. These efforts guided a gradual convergence between five measurement procedures, which we previously identified by their five different indicators. We can nicely organise geodesists' approach into three different iterations. During the first iteration, Helmert and Harkness intercompared the existing measures with two novel procedures based on lunar perturbations (I_M) and the earth's precession (I_{PC}) (*step 1*). Both indicators were believed to be sensitive to different sources of error than arc and pendulum measurements. Helmert and Harkness identified problems with the earlier I_P values around 1/289 and Clarke's I_T outcome of 1/294 since they departed significantly from a convergence interval between roughly 1/297 and 1/300 (*step 2*). In the case of I_P, the error was explained by an inferior altitude reduction procedure, while Clarke's I_T value was explained by a flawed distribution of his arc data across the globe (*step 3*). In the subsequent iteration, Darwin could further narrow down the convergence interval to 297 $\pm$ 0.5, identifying a potential problem with Helmert's I_P value of 1/299.6 (step 2). At the end of this step, Helmert explained this remaining discordance by three joint hypotheses about an unrepresentative geographical distribution of stations, island anomalies, and a mathematical shortcoming in the older versions of Clairaut's theorem. After applying these new corrections, he arrived an I_P value of 1/298.3 $\pm$ 1.1, whose associated error bar overlapped with Darwin's result. Finally, Hayford further substantiated the earlier corrections in iteration 3, when he introduced the fifth measure with yet another complementary source of error. The error bar of his 1/297.0 $\pm$ 0.5 overlapped with Helmert's and Darwin's.

Of course, this methodology for solving hard coordination problems can fail. Similar forms of reasoning and assessment may indicate a mistaken convergence interval and, subsequently, fail to lead to any fruitful hypotheses about the operative source of errors. Likewise, there may be cases in which even a highly diverse group of measures does not indicate any convergence interval to start with. Nonetheless, our findings show that operational pluralism *can* serve as a powerful heuristic for solving difficult cases of a recurrent epistemic problem in theory-mediated measurement. Keep in mind that incredibly extensive and costly ellipticity measurements had been pursued for more than 200 years before the above consensus was reached.

4.2. New responses to old worries

With this informal reconstruction at hand, we can anticipate some objections. In particular, I will argue that the methodology I distilled from geodetic practice is immune to two pertinent objections against

[23] If scientists do not have any new measurement procedure at their disposal (step 1), they might also repeat step 2 and 3 to assess produce evidence for their explanations of previously discovered outliers.

similar methodologies. Both objections identify the fallible nature of comparing the outcomes inferred from distinct procedures and take this fallibility to undermine their evidential significance or physical meaningfulness. These objections lose their force once we understand individual comparative inferences as parts of a larger diachronic and iterative process. A third and final worry concerns the problems that persistent discordances pose for the methodology of convergent, theory-mediated measurements proposed by George Smith and William Harper. I argue that my proposal can be read as an addition to this methodology, laying out a possible response to persistent discordances.

4.2.1. Measurement robustness and evidential objections

There are clear parallels between my account and discussions of *measurement robustness*.[24] Measurements are *robust* if "several different procedures yield closely similar results for a certain quantity under measurement" (Basso, 2017, p. 57). While I showed how operational pluralism can resolve hard problems of coordination, philosophers working on robustness point out how converging measures facilitate strong *evidence*. Donald Campbell's classical work on "multiple determination", for example, stresses that multiple agreeing measurement indicators increase the "inferential strength" of theories (Campbell & Fiske, 1959). William Wimsatt, similarly, takes measurement robustness to provide evidence of the predictive reliability of the theories (Wimsatt, 1981, p. 67). Recent work on robustness in the philosophy of metrology has focussed on the inverse evidential relation, i.e., on the evidence that robustness provides for measurement reliability. Eran Tal and Alessandro Basso have both argued that measuring the same value for an idealised parameter across multiple measurement procedures provides evidence for the reliability of these procedures (Basso, 2017; Tal, 2017a).

The classical worries about evidential appeals to robustness are so-called "independence objections" (Basso, 2017, p. 3). Independence objections highlight how difficult it is to assess whether different measures of the same parameter are *independent enough* from each other (Cartwright, 1991, p. 154; Stegenga, 2012, p. 209). Sensitivity to the same perturbational effects could turn out to introduce a common systematic error that we had failed to anticipate. In that case, the use of an additional procedure could reinforce the flaws of some of our initial measures and does not offer additional evidence of any kind.

My case study suggests that the focus of such debates has been too narrow. Robustness arguments and independence objections are usually presented as claims about the reliability of a particular comparative evidence assessment.[25] As I presented my case study, geodesists employed multiple indicators in a *diachronic* and *iterative* methodology. The aim of such a methodology is not to offer a static evidential criterion of a measurement's reliability but to gradually improve it (Chang, 2004, ch. 5). This severely raises the burden of argument for advocates of independence objections. Critics need to argue for more than the fallibility of individual convergence intervals. Rather, they need to show that drawing inferences from convergence intervals to operative sources of error cannot *improve* one's epistemic situation or might even *worsen* it. Like any kind of inference, inferences from a supposed convergence interval between multiple measures might sometimes be affected by an unaccounted systematic error and numerical convergence is no general criterion for "accuracy" or "truth". However, the self-corrective character of operational pluralism allows that such shortcomings can themselves become targets of inquiry in future iterations of the methodology.

Note that my response is much more optimistic than existing arguments presented by William Wimsatt (1987), Jaakko Kuorikoski et al. (2010), Kuorikoski and Marchionni (2016), and Alessandra Basso (2017). Their common idea is that practicing scientists can sufficiently avoid making wrong judgements about independence if they stick to what one might call an "error proviso". According to that proviso, one should only draw evidential conclusions from a supposedly independent convergence if one *knows* that the respective measures are subject to different sources of error.[26] This claim amounts to a quite rigid precondition, presupposing independent means for individuating and predicting operative sources of error. I have defined hard problems of coordination as precisely those situations in which scientists' lack the theoretical tools for doing so. In operational pluralism, comparative inferences only need to pick out targets for further inquiry. Hence the reference to *beliefs* about different sources of error (rather than *knowledge*) in step 1 of my heuristic. Scientists do not need to independently confirm the operative sources of error affecting particular indicators *before* drawing inferences from their convergence or disagreement. Rather, they can increase the credence they give to a numerical convergence interval *and* hypotheses about specific source of errors in a complementary and iterative manner: convergence intervals may be justified and adjusted by scientists' ability to explain diverging results, while hypotheses about causes of error are justified based on how well they can be used to increase coherence in future measurements.

4.2.2. Physical meaningfulness and spurious convergences

A second long-standing worry one might raise against operational pluralism can be traced back to Percy Bridgman's *Logic of Modern Physics* (1927). For Bridgman, contemporary developments in physics had exposed the danger of taking theoretical parameters to be *physically meaningful* outside of their domain of operational determinability. Quantum and relativity theory had to "save" physics from a crisis caused by generalising quantity terms to situations in which they were never physically measured (e.g. at very high velocities or subatomic scale). Taking inspiration in these events, Bridgman proposed to radically restrict the inferential domain of theoretical parameters to the specific physical operations that are used to determine their magnitude. If two such operations associated with the same theoretical parameter cannot be employed in the same domain, we have no reason to think agreements or disagreements between them have any physical significance, Hence, we cannot draw justified inferences about the physical world from such comparisons. Bridgman drew quite radical conclusions about the inferential scope of measurements, becoming (in)famous for endorsing claims to the extent that astronomical telescopes and rods do not measure the same attribute (Bridgman, 1927, pp. 17–18).

The literature is rife in rebuttals to Bridgman's rigid account of the inferential scope of measurements. Almost every commentator agrees that it is *overly* restrictive and incongruent with scientific concept formation (Chang, 2017; Feest, 2005; Hempel, 1966, ch, 7). Instead of adding another rebuttal to the list, I want to take operationalism seriously and show that its basic commitment to a notion of physical

[24] Measurement robustness constitutes a separate problem from derivational robustness, which many take to be a guiding principle of theorising and modelling. The virtues of derivational robustness have been discussed extensively by Richard Levins (1966) and, more recently, by Weisberg (2006); Kuorikoski et al. (2010), and Odenbaugh and Alexandrova (2011). A taxonomy of the different notions of "robustness" in modelling and measurement is provided in Wimmsatt (1981).

[25] Bayesian accounts – of which Schupbach's (2018) explanatory account might be the most sophisticated example – go even further and specify *how much* a successful comparison of "independent" results should increase the evidential support for some hypothesis – where "independent" is specified in terms of probabilistic, confirmational, or explanatory independence (depending on the account). Such projects are much more ambitious than my efforts here. I am only defending the salience of a methodology for *improving* measurement coordination and do not attempt to quantify this salience according to any particular metric.

[26] Basso weakens this proviso further to include cases in which the same source of error causes discrepancies of different magnitude across different measures. The difference is not central to my argument here because it also requires antecedent knowledge of the errors uniquely affecting certain measures.

meaningfulness is compatible with the methodology I defend.[27] Even if we accept the inferential use of quantities cannot be categorically limited to specific operations – as Bridgman himself conceded later (Chang, 2017) – we can rescue a genuine worry about *spurious convergences* from his account. I understand a convergence between measures of a quantity as spurious iff some of these measures are subject to domain-specific and unaccounted sources of error. Some remarks in the *Logic* already even indicate that Bridgman primarily intended his notion of physical meaningfulness as a useful tool for identifying spurious convergences, as he noted that the "essential point" of his views is to differentiate "constructs" from "physically meaningful quantities",[28] where the latter "can be defined [i.e. operationalised] in several alternative ways in terms of physically distinct operations" (Bridgman, 1927, pp. 55–56).[29]

Bridgman's (charitably interpreted) worry about spurious convergences can straightforwardly be applied to our case. Triangulations (I_T), pendulum stations (I_P), and network-wide deflections of the vertical (I_{DV}) were assessed in different places or regions and involved different theoretical corrections. Bridgman's worry becomes even more acute when we compare any one of those three indicators with observations of the lunar orbit (I_M) or the movement of fixed stars (I_{PC}). If his account is correct, my operational pluralism is susceptible to the following objection: It is wrong to first suppose that different indicators permit inferences to the magnitude of the same theoretical parameter and then account for indicator-specific errors afterwards. This supposition of my account makes scientists unnecessarily vulnerable to being led astray by the mathematical structure of their models and face confusion once measurement results stop cohering (Bridgman, 1927, p. 42). Consequently, they should have taken their measurement procedures to have different local targets until the different perturbations acting in the specific domains of all indicators were fully understood.

Geodesists detailed concern about domain specific perturbations shows that they were acutely aware of the problem of spurious convergences. However, they integrated this awareness into a productive methodology for isolating and predicting the effects of such perturbations on their measurement. Fleshing out *how* such awareness guides successful measurement practice is the key to blocking the above objection and rescuing a notion of physical meaningfulness that is compatible with operational pluralism. Pace Bridgman, geodesists did not take the danger of spurious convergence to *restrict* comparisons between measurements with distinct physical indicators and domains. Rather, it is precisely through diachronic and iterative comparisons that they gradually identified and anticipated the operative sources of measurement error. In such a methodology, physical meaningfulness does not act as a static criterion for avoiding the risk of spurious convergences. Rather, it plays a dynamic role in the diachronic improvement of measurements, in which scientists gradually *reduce* the risk of spurious convergences. This is reflected in the third step of the iterative heuristic proposed in section 4.1. To successfully apply operational pluralism, scientists should not only aim at achieving convergence, but develop hypotheses to explain outliers in virtue of perturbations uniquely affecting specific measurement indicators. These hypotheses may then be used to predict domain-specific errors and are assessed based on their ability to increase future measurement coherence (cf. 4.2.1).

To sum up, I happily concede that Bridgman's worry about spurious convergences is genuine and that a criterion for the physical meaningfulness of theoretical parameters is practically warranted. However, I think that physical meaningfulness should be conceptualised as a dynamic criterion, gradually realised in form of scientists' increasing ability to anticipate domain-specific errors.[30] For example, we can say that the theoretical parameter "ellipticity" was less physically meaningful in respect to its intended target domain (the physical properties of the earth, such as its attraction, rotation, curvature, or density) before 1880 than it was around 1924. During that period geodesists gradually learned to anticipate the perturbations unique to the subdomains of their particular physical indicators. Conceptualised in this manner, physical meaningfulness offers a useful normative notion that urges scientists to gradually anticipate domain-specific measurement errors and minimise the risk of spurious convergences.

In fact, the subsequent history of geodesy provides a nice example of how operational pluralism guides the pursuit of further discordances and increased physical meaningfulness *after* scientists achieved an initial convergence. During the twentieth century, geodesists discovered more local deviations from their initial reference model and tied them to a growing body of knowledge about the form and dynamics of the earth's internal strata. Similar to the period investigated in this paper, geodesists discovered, and anticipated such deviations by reasoning with multiple measures. Most notably, they would employ seismological measurements of the varying viscosity in the earth's interior and satellite measurements of the earth's gravity field, which provided new, partially discrepant results. This led to the construction of finer-grained models of the earth's internal constitution and gravity field but also further increased the meaningfulness of the ellipsoid – now backed by various new corrections for the resulting perturbations (Fischer, 1975, pp. 35–42).

4.2.3. Convergent theory-mediated measurement and resistant discrepancies

Before closing, I need to address an influential account of physical measurement, which, among other things, has directly inspired various aspects of this proposal. Over the last decades, George Smith and William Harper have re-habilitated the methodology of Newton's *Principia*, which reserves a central role for convergent theory-mediated measurements (Harper, 2012; Smith, 2002, 2014). According to that methodology, scientists generate high-quality evidence by (i) inductively generalising observed regularities, (ii) deriving new phenomena from them, and (iii) comparing these phenomena with physical measurements. On the Smith-Harper account, it is not merely the convergence and precision of these measurements that facilitate evidential support. Rather, high quality evidence is generated if any robust deviations from predicted values can be explained by the initial theory. Thus, pursuing converging theory-mediated measurements produces robust discrepancies, whose subsequent explanation maximises the evidential demand on theory. It comes as no surprise that celestial mechanics is both Smith's and Harper's prime example for gradually converging theory-mediated measurements, arguably constituting one of the most successful research programmes in the history of science. Newtonian gravity allowed astronomers to explain an incredible number empirical discrepancies by postulating additional perturbing bodies that conformed with their theoretical expectations – a legacy since picked up by General Relativity (Harper, 2012, pp. 378–84).

Harper and Smith's focus on the paradigm case of celestial mechanism raises the question of how scientists should handle cases of theory-

[27] For a similar, constructive approach to Bridgman's operationalism see: Chang (2004, 2017).

[28] This conception of physical meaningfulness is not coextensive with the formal definitions of meaningfulness as invariance across unit transformation defended in the *Foundations of Measurement* or, more recently, by Louis Narens (Krantz et al., 2006; Narens, 2007). This is not the place to discuss their differences in detail, but it suffices to note that the operationalist notion of meaningfulness focusses on invariance across *actual* physical operations performable by scientists. In contrasts, the "units" in the formal definitions are referring to ideal atomic operations that still need to be coordinated with physical measuring devices.

[29] This aspect of Bridgman's thought has also been picked up by Mahmoud Jalloh (draft), who is fleshing out operational invariance in terms of dimensionality.

[30] Thus, I agree with Teru Miyake and George Smith who define "physically meaningful representations and parameterizations", as those that "yield deviations that have physically identifiable sources". I only insist that physical meaningfulness is best conceptualised as a matter of degrees, increasing gradually throughout the process of inquiry (Miyake and Smith 2021, p. 172).

mediated measurements that remain discordant over time (Miyake, 2013, pp. 313–14; Ohnesorge, 2021). I take it that hard problems of coordination are such cases. Here, scientists cannot shield measurement procedures against the physical causes of discordances, nor can they uniquely identify and predict them by relying on background theory. On my account, such problems can sometimes be overcome by increasing the variety of physically distinct measurement operations, while aiming to isolate operative sources of error. Scientist can then correct persistent discordances between measurements in a process of diachronic revision, in which they gradually introduce new, measurement-specific corrections based on their ability to facilitate measurement convergence. This offers an instance of "epistemic iteration" on the horizontal level (Chang, 2004, ch. 5). Hence, I hope that operational pluralism offers a welcome addition to the Smith-Harper account of convergent measurement.

5. Conclusion

Recent work in the philosophy of measurement has highlighted that establishing derived measurements involves problems of coordination. I have suggested that such problems are not merely a general predicament of derived measurement, but that their difficulty varies according to scientists' predictive and experimental control over perturbations of the measurement process. Such perturbations arise from discrepancies between idealised models and the de facto physical features of the measurement indicators, target, and context. To identify how scientists might respond to hard problems of coordination, I analysed a measurement problem in physical geodesy that had persisted for more than two hundred years after its first discussion in Newton's *Principia*. This problem persisted so stubbornly because geodesists were unable to theoretically predict or experimentally control the perturbations resulting from the earth's heterogenous topographic and subterranean density distribution. Notably, they *did* eventually resolve their measurement problem, indicating that their methodology can provide lessons for resolving similar problems in and beyond physical geoscience. This methodology – which I dubbed operational pluralism – can be pursued by (i) increasing the number of physically distinct measurement indicators at disposal, (ii) identifying outliers from a shared convergence interval, and (iii) explaining their discordance based on perturbations uniquely affecting measurement with specific physical indicators. In repeatedly following steps 1 to 3, scientists can iteratively adjust previous convergence intervals and develop increasingly specific hypotheses about the sources of measurement error.

I have argued that operational pluralism offers a *generalizable* and *strong* methodology. The methodology is *generalizable* because it provides a response to a common epistemic problem affecting theory-mediated measurements of large and partially inaccessible physical systems. It is *strong* because it is immune to canonical objections that have been raised in the literatures on operationalism and measurement robustness. Far from falling prey to these objections, operational pluralism indicates that the focus of the respective debates might have been cast too narrow. While philosophers have questioned the physical meaningfulness and evidential value of *individual* comparisons between different measures, such comparisons become methodologically salient when used *repeatedly* and *iteratively*.

CRediT author statement

The author confirms being the sole contributor to this work and has approved it for publication.

Funding

This work was supported by the Cambridge Trust [Grant number: 10556463].

Acknowledgements

I thank Hasok Chang for his invaluable guidance in writing this paper and the Ph.D dissertation that I developed it from. I thank George Smith for his generous help in navigating the history of Newtonian gravity. I thank the members of the 2021 Du Châtelet price committee (Katherine Brading, Alisa Bokulich, Daniel Mitchell, and Wendy Parker) for detailed and very insightful comments on earlier versions of this paper. Finally, I thank Cristian Larroulet Philippi, Aditya Jha, Ahmad Elabbar, Mahmoud Jalloh, and Aja Watkins for very helpful feedback throughout.

References

Airy, G. B. (1826). On the figure of the earth. *Transactions of the Royal Society, 13*, 548–579.

Airy, G. B. (1845). Figure of the Earth. In E. Smedley, Hugh J. Rose, & Henry J. Rose (Eds.)*, Vol. 5. Encyclopaedia Metropolitana* (pp. 165–240). London: Fellowes & Rivington.

Airy, G. B. (1861). Corrections of the Elements of the Moon's Orbit, deduced from the Lunar Observations made at the Royal Observatory of Greenwich from 1750 to 1851. Being an extension of a preceding memoir entitled Corrections to the Elements of the Moon's Orbit deduced from the Lunar Observations made at the Royal Observatory of Greenwich from 1750 to 1830. *Memoirs of the Royal Astronomical Society, 29*, 1.

d'Alambert, J.-B. R. (1749). *Recherches sur la précession des équinoxes et sur la nutation de l'axe de la terre dans le système newtonien ([Reprod. en fac-sim.])*. https://gallica.bnf.fr/ark:/12148/bpt6k3804r.

Basso, A. (2017). The appeal to robustness in measurement practice. *Studies in History and Philosophy of Science, 65–66*(December), 57–66. https://doi.org/10.1016/j.shpsa.2017.02.001

Belot, G. (2015). Down to earth underdetermination. *Philosophy and Phenomenological Research, 91*(2), 456–464. https://doi.org/10.1111/phpr.12096

Bessel, F. W. (1837). Ueber den Einfluss der Unregelmässigkeiten der Figur der Erde auf geodätische Arbeiten Und ihre Vergleichung mit den astronomischen Bestimmungen. *Astronomische Nachrichten, 14*(19–21), 269–312. https://doi.org/10.1002/asna.18370141901

Bessel, F. W. (1838). *Gradmessung in Ostpreußen und ihre Verbindung mit Preußischen und Russischen Dreieckskettten*. Berlin.

BIPM (International Bureau of Weights and Measures). (2006). The International System of Units (SI) Brochure. 8th edition. http://www.bipm.org/en/si/si_brochure/. (Accessed 3 September 2022).

Bokulich, A. (2018). Using models to correct data: Paleodiversity and the fossil record. *Synthese, 198*(24), 5919–5940 (2021). https://doi.org/10.1007/s11229-018-1820-x.

Bokulich, A. (2020). Calibration, coherence, and consilience in radiometric measures of geologic time. *Philosophy of Science, 87*(3), 425–456.

Bokulich, A., & Oreskes, N. (2017). Models in geosciences. In L. Magnani, & T. Bertolotti (Eds.), *Springer HandbooksSpringer handbook of model-based science* (pp. 891–911). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-30526-4_41.

Boumans, M. (2007). Invariance and calibration. In M. Boumans (Ed.), *Measurement in economics: A handbook* (pp. 19–40). Amsterdam: Elsevier.

Bradley, J. (1748). A letter to the Right Honourable George Earl of Macclesfield concerning an apparent motion observed in some of the fixed stars; by James Bradley D. D. Astronomer Royal, and F. R. S. *Philosophical Transactions (1683-1775), 45*, 1–43.

Bridgman, P. W. (1927). *The logic of modern physics*. New York: Beaufort Books.

Bruns, H. (1878). *Die Figur der Erde: Ein Beitrag zur Europäischen Gradmessung*. Berlin.

Bürg, J. T. (1825). Bürg Epoche Der Mittleren Länge Des Mondes Für 1779 Jährliche Aenderung Derselben, Gleicjung Der Länge Etc. *Astronomische Nachrichten, 4*(March), 9.

Callandreau, O. (1889). Mémoire Sur La Théorie de La Figure Des Planètes. *Annales de l'Observatoire Impérial de Paris. Mémoires, 19*(5), 1–84.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. https://doi.org/10.1037/h0046016

Cartwright, N. (1991). Replicability, reproducibility, and robustness: Comments on Harry Collins. *History of Political Economy, 23*(1), 143–155. https://doi.org/10.1215/00182702-23-1-143

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

Chang, H. (2017). Operationalism: Old lessons and new challenges. In A. Nordmann, & N. Mössner (Eds.), *Reasoning in measurement* (pp. 25–38). Routledge.

Chapin, S. (1995). The shape of the earth. In *The general history of astronomy: Planetary astronomy from the renaissance to the rise of astrophysics* (pp. 22–34). Cambridge: Cambridge University Press.

Clarke, A. R. (1880). *Geodesy*. Oxford: Clarendon Press.

Clarke, A. R., & James, H. (1858). *Ordnance trigonometrical survey of Great Britain and Ireland: Account of the observations and calculations of the principal triangulation, and of the figure, dimensions and mean specific gravity of the earth as derived therefrom*. London: Ordnance Survey of Great Britain.

Darwin, G. H. (1899). The theory of the figure of the earth carried to the second order of small quantities. *Monthly Notices of the Royal Astronomical Society, 60*(2), 82–124. https://doi.org/10.1093/mnras/60.2.82

Dundas, L. J. L., et al. (1924). The proposed adoption of a standard figure of the Earth: discussion. *The Geographical Journal, 64*, 129–139.
Euler, J. A. (1768). Versuch die Figur der Erden durch Beobachtungen des Mondes zu bestimmen. *Abhandlungen Der Bayerischen Akademie Der Wissenschaften, 5*(4), 197–214.
Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences, 41*(2), 131–149. https://doi.org/10.1002/jhbs.20079
Fischer, P. (1845). *Lehrbuch der höheren Geodäsie: Die Theorie der Beobachtungsfehler und ihre Ausgleichung durch die Methode der kleinsten Quadrate* (Vol. 1). Leske.
Fischer, P. (1846a). *Lehrbuch der höheren Geodäsie: Die Beobachtungsarbeiten und Instrumente*. Leske.
Fischer, P. (1846b). *Lehrbuch der höheren Geodäsie: Die Berechnungen*. Leske.
Fischer, P. (1868). *Untersuchungen über die Gestalt der Erde*. Joh. Ph. Diehl.
Fischer, I. (1975). The figure of the earth — Changes in concepts. *Geophysical Surveys, 2*(1), 3–54. https://doi.org/10.1007/BF01447937
van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective* (1st publ). Oxford: Clarendon Press.
Gauss, C. F. (1828). Bestimmung des Breitenunterschiede zwischen den Sternwarten Göttingen Und Altona durch Beobachtungen am Ramsdenschen Zenithsector. In C. F. Gauss (Ed.), *Cambridge Library Collection. Mathematics: Vol. 9. Werke* (pp. 5–64). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139058308.001.
Good, G. (2011). Measuring the inaccessible earth: Geomagnetism, in situ measurements, remote sensing, and proxy data. *Centaurus, 53*(2), 176–189.
Greenberg, J. L. (1988). Breaking a "vicous circle": Unscrambling A.-C. Clairaut's iterative method of 1743. *Historia Mathematica, 15*(3), 228–239. https://doi.org/10.1016/0315-0860(88)90063-8
Greenberg, J. L. (1995). *The problem of the earth's shape from Newton to Clairaut: The rise of mathematical science in 18th century Paris and the fall of 'normal' science*. Cambridge: Cambridge University Press.
Greenberg, J. L. (1996). Isaac Newton and the problem of the earth's shape. *Archive for History of Exact Sciences, 49*(4), 371–391. https://doi.org/10.1007/BF00374704
Hansen, P. A. (1862). *Darlegung der theoretischen Berechnung der in den Mondtafeln angewandten Störungen* (Vol. 1). S. Hirzel.
Hansen, P. A. (1864). *Darlegung der theoretischen Berechnung der in den Mondtafeln angewandten Störungen* (Vol. 2). S. Hirzel.
Harkness, W. (1891). *The solar parallax and its related constants, including the figure and density of the earth*. Washington, Govt. Print. Off.
Harper, W. (2012). *Isaac Newton's scientific method: Turning data into evidence about gravity and cosmology*. Oxford; New York: Oxford University Press.
Hayford, J. F. (1909). *The figure of the earth and isostasy from measurements in the United States*. Washington DC: U.S. Government Printing Office.
Heidelberger, M. (1994). Alternative Interpretationen der Repräsentationstheorie der Messung. In G. Meggle, & U. Wessels (Eds.), *Proceedings of the 1st conference "Perspectives in analytical philosophy"* (pp. 310–324). Berlin; New York: Walter de Gruyter.
Helmert, F. R. (1884). *Die mathematischen und physikalischen Theorieen der höheren Geodäsie* (Vol. 2. 2 vols). Leipzig: B. G. Teubner.
Helmert, F. R. (1901). Der normale Teil der Schwerkraft im Meeresniveau. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 328–336.
Hempel, C. G. (1966). In*Philosophy of natural science*. Upper Saddle River, NJ: Prentice Hall.
Howarth, R. (2008). Gravity surveying in early geophysics. II. From mountains to salt domes. *Earth Sciences History, 26*(2), 229–261. https://doi.org/10.17704/eshi.26.2.f04281625w2w7614
Jalloh, M.. unpublished. Scientific polysemy, semantic detoxification, and sophisticated operationalism.
Krantz, D. H., Duncan Luce, R., Suppes, P., & Tversky, A. (2006). *Foundations of measurement volume I: Additive and polynomial representations*. Mineola, NY: Dover Publications.
Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science, 61*(3), 541–567. https://doi.org/10.1093/bjps/axp049
Kuorikoski, J., & Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science, 83*(April), 227–247. https://doi.org/10.1086/684960
Lambert, W. D. (1926). The figure of the Earth and the new International ellipsoid of rreference. *Science, 63*(1627), 242–248.
Laplace, P. S. (1796). *Exposition Du système Du Monde*. I. Paris.
Laplace, P. S. (1829). *Mécanique Céleste* (Vol. 1). Boston: Hillard, Gray, Little, and Wilkins (N. Bowditch, Trans.).
Laplace, P. S. (1832). *Mécanique Céleste* (Vol. 2). Boston: Hillard, Gray, Little, and Wilkins (N. Bowditch, Trans.).
Laplace, P. S. (1834)*, Vol. 3. Mécanique Céleste*. Boston: Hillar, Gray, Little, and Wilkins (N. Bowditch, Trans.).
Legendre, A. M. (1805). *Nouvelles Méthodes Pour La Détermination des Orbites des Comètes*. Paris: F. Didot.
Levins, R. (1966). The strategy of model building in population biology. *American Scientist, 54*(4), 421–431.
Luchetti, M. (2020). From successful measurement to the birth of a law: Disentangling coordination in Ohm's scientific practice. *Studies in History and Philosophy of Science Part A, 84*(December), 119–131. https://doi.org/10.1016/j.shpsa.2020.09.005
Mach, E. (1986). Principles of the theory of heat: Historically and critically elucidated (B. McGuinness, Trans.). In *Vienna Circle Collection* (Vol. 17). Dordrecht: Springer. https://doi.org/10.1007/978-94-009-4622-4.
McCaw, G. T. (1924). The proposed adoption of a standard figure of the Earth. *The Geographical Journal, 64*(2), 120–129.
Miyake, T. (2011). *Underdetermination and indirect measurement*. Stanford University.
Miyake, T. (2013). Essay Review: Isaac Newton's Scientific Method. Edited by William Harper *Philosophy of Science, 80*(2), 310–316. https://doi.org/10.1086/670660.
Miyake, T. (2015). Reference models: Using models to turn data into evidence. *Philosophy of Science, 82*(5), 822–832. https://doi.org/10.1086/683322
Miyake, T. (2017a). Modelling and uncertainty in seismology. In N. Mössner, & A. Nordmann (Eds.), *Reasoning in measurement* (pp. 220–232). New York: Routledge.
Miyake, T. (2017b). Magnitude, moment, and measurement: The seismic mechanism controversy and its resolution. *Studies in History and Philosophy of Science, 65–66*(December), 112–120. https://doi.org/10.1016/j.shpsa.2017.02.002
Miyake, T., & Smith, G. E. (2021). Realism, Physical Meaningfulness, and Molecular Spectroscopy. In T. D. Lyons, & P. Vickers (Eds.), *Contemporary Scientific Realism: The challenge from the history of science* (pp. 159–180). Oxford: Oxford University Press.
Narens, L. (2007). *Introduction to the theories of measurement and meaningfulness and the use of symmetry in science* (1st ed.). Mahwah, NJ: Taylor & Francis Inc.
Odenbaugh, J., & Alexandrova, A. (2011). Buyer beware: Robustness analyses in economics and biology. *Biology and Philosophy, 26*(5), 757–771. https://doi.org/10.1007/s10539-011-9278-y
Ohnesorge, M. (2021). How incoherent measurement succeeds: Coordination and success in the measurement of the earth's polar flattening. *Studies in History and Philosophy of Science Part A, 88*(August), 245–262. https://doi.org/10.1016/j.shpsa.2021.06.006
Oreskes, N. (1999). *The rejection of continental drift: Theory and method in American earth science*. New York: Oxford University Press USA.
Padovani, F. (2017). Coordination and measurement: What we get wrong about what Reichenbach got right. In , *European studies in philosophy of scienceEPSA15 selected papers* (pp. 49–60). Cham: Springer International Publishing.
Parker, W. S. (2014). Simulation and understanding in the study of weather and climate. *Perspectives on Science, 22*(3), 336–356. https://doi.org/10.1162/POSC_ a_ 00137
Puissant, L. (1842a). *Traité de géodésie, ou, Exposition des méthodes trigonométriques et astronomiques, applicables à la mesure de la terre, et à la construction du canevas des cartes topographiques* (Vol. 1. 2 vols). Bachelier.
Puissant, L. (1842b). *Traité de géodésie: ou, Exposition des méthodes trigonométriques et astronomiques, applicables à la mesure de la terre, et à la construction du canevas des cartes topographiques* (Vol. 2. 2 vols). Bachelier.
Radau, R. (1885). Mémoires et Observations. Remarques Sur La Théorie de La Figure de La Terre. *Bulletin Astronomique, Serie I, 2*, 157–161.
Radau, R. (1890). Mémoires et Observations. Sur La Loi Des Densités à l'intérieur de La Terre. *Bulletin Astronomique, Serie I, 7*, 76–92.
Reichenbach, H. (1920). *Relativitätstheorie und Erkenntnis Apriori*. Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-50774-8
Reichenbach, H. (1932). Die Kausalbehauptung und die Möglichkeit ihrer empirischen Nachprüfung. *Erkenntnis, 3*, 32–64.
Reigber, C. (2017). Friedrich Robert Helmert 1843–1917: Ausschnitte aus seinem Lebensweg und seinem internationalen Wirken. In W. Freeden, & R. Rummel (Eds.), *Springer Reference NaturwissenschaftenHandbuch der Geodäsie: 6 Bände* (pp. 1–38). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-46900-2_90-1.
Sabine, E. (1825). *An account of experiments to determine the figure of the earth by means of the pendulum vibrating seconds in different latitudes: As well as on various other subjects of philosophical inquiry*. London: John Murray.
Schupbach, J. N. (2018). Robustness analysis as explanatory reasoning. *The British Journal for the Philosophy of Science, 69*(1), 275–300. https://doi.org/10.1093/bjps/axw008
Schweitzer, J. (2008). The birth of modern seismology in the nineteenth and twentieth centuries. *Earth Sciences History, 26*(2), 263–280. https://doi.org/10.17704/eshi.26.2.q412767051228155
Smith, G. E. (2002). The methodology of the Principia. In G. E. Smith, & I. B. Cohen (Eds.), *Cambridge Companions to PhilosophyThe Cambridge Companion to Newton* (pp. 138–173). Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521651778.005.
Smith, G. E. (2007). Gaining access: Using seismology to probe the earth's insides. Https://Web.Stanford.Edu/Dept/Cisst/Events0506.Html.
Smith, G. E. (2014). Closing the loop: Testing Newtonian gravity, then and now. In Z. Biener, & E. Schliesser (Eds.), *Newton and empiricism* (Vols. 262–353). Oxford: Oxford Univ. Press.
Smith, G. E., & Seth, R. (2020). Brownian motion and molecular reality. In *Oxford Studies in philosophy of science*. Oxford, New York: Oxford University Press.
Sommerfeld, A., & Oppenheim, S. (1926). Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen: Fünfter Band: Physik. In *Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*. Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-663-16029-8
Stegenga, J. (2012). Rerum concordia discors: Robustness and discordant multimodal evidence. In L. Soler, E. Trizio, T. Nickles, & W. Wimsatt (Eds.), *Boston studies in the philosophy of scienceCharacterizing the robustness of science: After the practice turn in philosophy of science* (pp. 207–226). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2759-5_9.
Strasser, G. (1957). *Ellipsoidische Parameter Der Erdfigur, 1800-1950*. Munich: Verlag der Bayerischen Akademie der Wissenschaften.
Stump, D. J. (2015). *Conceptual change and the philosophy of science: Alternative interpretations of the a priori*. Routledge.
Tal, E. (2016). Making time: A study in the epistemology of measurement. *The British Journal for the Philosophy of Science, 67*(1), 297–335. https://doi.org/10.1093/bjps/axu037
Tal, E. (2017a). A model-based epistemology of measurement. In N. Mößner, & A. Nordmann (Eds.), *Reasoning in measurement* (pp. 233–253). New York: Routledge.
Tal, E. (2017b). Calibration: Modelling the measurement process. *Studies in History and Philosophy of Science Part A, 65*, 33–45. https://doi.org/10.1016/j.shpsa.2017.09.001

Tisserand, F. F. (1890). Annex A X: Réflexions Au Sujet Dela Constitution de La Terre Suposée Fluide Dans Un Interior, et Remarques Sur Les Diverses Méthodes Employées Pour Déterminer l'aplatissement Superficiel. In *C.R. Der 9. Generalversammlung der IAG in Paris 1899* (pp. 8–9). Berlin: Königlich-Preussisches Geodätisches Institut.

Todhunter, I. (1873). *A history of the mathematical theories of attraction and the figure of the earth, from the time of Newton to that of Laplace* (Vol. 1). London: Macmillan.

Torge, W. (2001). *Geodesy*, 3rd compl. rev. and Extend. ed. Berlin; New York: De Gruyter.

Torge, W. (2005). The international association of geodesy 1862 to 1922: From a regional project to an international organization. *Journal of Geodesy, 78*(9), 558–568. https://doi.org/10.1007/s00190-004-0423-0

Torge, W. (2009). Geschichte Der Geodäsie in Deutschland. 2. In *Durchges. und korrigierte Aufl.* Berlin: De Gruyter.

Torge, W. (2017). Geschichte der Erdmessung. In R. Rummel (Ed.), *Erdmessung und Satellitengeodäsie: Handbuch der Geodäsie, herausgegeben von Willi Freeden und Reiner Rummel* (pp. 1–71). Berlin, Heidelberg: Springer Reference Naturwissenschaften. https://doi.org/10.1007/978-3-662-47100-5_2. Springer.

Watkins, A. (2021). The epistemic value of the living fossils concept. *Philosophy of Science*. https://doi.org/10.1086/714875. May.

Weisberg, M. (2006). Robustness analysis. *Philosophy of Science, 73*(5), 730–742. https://doi.org/10.1086/518628

Wiechert, E. (1897). Ueber die Massenverteilung im Inneren der Erde. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1897*(3), 221–243.

Wilson, C. (1987). D'Alembert versus Euler on the precession of the equinoxes and the mechanics of rigid bodies. *Archive for History of Exact Sciences, 37*(3), 233–273.

Wimsatt, W. (1987). False models as means to truer theories. In M. H. Nitecki, & A. Hoffman (Eds.), *Neutral models in biology* (pp. 33–55). New York; Oxford, England: Oxford University Press.

Wimsatt, W. (2012). Robustness, reliability, and overdetermination (1981). In L. Soler, E. Trizio, T. Nickles, & W. Wimsatt (Eds.), *Boston studies in the philosophy of scienceCharacterizing the robustness of science: After the practice turn in philosophy of science* (pp. 61–87). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2759-5_2.