

## **We Should not Align Quantitative Measures with Stakeholder Values**

Miguel Ohnesorge

Department of History and Philosophy of Science, University of Cambridge, UK.

Email: [mo459@cam.ac.uk](mailto:mo459@cam.ac.uk);

---

### **Abstract**

There is a growing consensus among philosophers that quantifying value-laden concepts can be epistemically successful and politically legitimate if all value-laden choices in the process of quantification are aligned with stakeholder values. I argue that proponents of this view have failed to argue for its basic premise: successful quantification is sufficiently unconstrained to be achievable along multiple, stakeholder-specific pathways. I then challenge this premise by considering a rare example of successful value-laden quantification in seismology. Seismologists quantified earthquake size precisely by excluding stakeholder values from measure design and testing.

---

### **1. Introduction**

Quantification is a divisive topic. Traditional proponents of quantification stress its epistemic value and historical importance in physical science. Traditional sceptics highlight how quantification can blackbox value judgements and transfer illegitimate power to scientists and technocrats. Recent work in philosophy of measurement aims for an optimistic compromise. We should pursue the epistemic benefits of quantification but ensure its political legitimacy by aligning value-laden choices with the values of relevant stakeholders. In what follows, I argue that this compromise fails.

My argument consists of two parts. The first part illustrates what I call the alignment approach to quantification and shows that it presumes a basic but unsubstantiated premise. Proponents of the alignment approach argue that political legitimacy should be added “as an additional layer of security [...] to the familiar scientific process covered in textbooks on measurement” (Fabian and Alexandrova 2022, 6; similarly: Duque, Tal, and Barbic 2024, 9).

While they disagree on how to operationalize legitimacy, they agree that it requires a procedure for aligning value-laden choices with the values of relevant stakeholders. All operationalizations of the alignment approach presume that the empirical constraints on quantitative measurement are loose enough to reach quantification along multiple, stakeholder-specific pathways. Proponents of the alignment approach have yet to provide evidence that this basic premise is correct.

The second part of my argument casts doubts on the basic premise. I use seismology as an exemplary case in which a value-laden concept was quantified successfully. To design and test quantitative scales, seismologists had to disregard stakeholder values for overwhelmingly theoretical concerns. The case study suggest that value alignment will frequently be incompatible with successful quantification. Hence, the alignment approach falls short of its goal. To achieve politically legitimate *and* successful quantification, we need to look for alternative sources of legitimacy.

The plan is as follows. Section 2 introduces the problem of value-ladenness in quantification. Section 3 introduces the alignment approach to quantification and argues that all its proponents are committed to an unsubstantiated premise. Section 4 reconstructs quantification in seismology to challenge that premise. I conclude with an optimistic but merely promissory outlook on alternative sources of political legitimacy in measurement.

## **2. Quantifying Value-Laden Concepts**

Philosophers speak about quantification in at least two different senses. In a descriptive sense, quantifying simply means representing an attribute (mass, time, biodiversity, ...) in quantitative terms. I am here concerned with a second, normative meaning of quantification. To quantify an attribute, in this second sense, means *being justified* in describing that attribute in quantitative terms (following Helmholtz 1887, Michell 1997, and Trendler 2009).

To justify quantitative descriptions of an attribute, we need to justify quantitative scales for comparing instances of that attribute. Scales are quantitative if they produce measurement outcomes that contain sufficient relational structure for scientists to calculate differences between these outcomes via addition or subtraction. The basic quantitative scale is an interval scale, which presumes that all instances of an attribute stand in transitive and monotonic distance relations. To justify propositions like “If object X is 5°C warmer than object Y and

10°C warmer than object Z, then object Y is 5° warmer than object Z,” we need to justify the Celsius (interval) scale. Hence, quantifying means being justified in measuring an attribute on an interval scale.<sup>1</sup>

I am here not concerned with the general epistemological debate about how quantitative scales can be tested empirically. Scientists and philosophers continue to disagree about the exact theoretical and experimental presuppositions of such tests (Trendler 2009; 2019; Sherry 2011; Michell 2019; Tal 2019; Vessonen 2020; Thalos 2023). It suffices to note that there is sufficient common ground to identify clear examples of successfully quantified attributes in the history of science, ranging from temperature to electric current. The distance relations between specific instances of these attributes were shown to (i) remain stable in controlled conditions and (ii) converge across instruments and background conditions to a sufficiently high level of precision. In many successful cases, presupposing quantitative scales further allowed scientists to discover new and highly specific details of the empirical world (Smith 2014; Miyake 2017; Smith and Seth 2020). It is also commonly agreed that quantitative scales – if empirically justified – are epistemically valuable. They allow us to mathematically derive and empirically test predictions at a higher degree of specificity.

In many scientific disciplines, the general problem of testing quantitative scales is compounded by a second, more specific problem: *(how) should quantification be pursued if it involves value-laden choices?* I assume that choices can be considered value-laden if and only if they are epistemically unforced and have morally significant consequences. This definition excludes epistemically unforced choices that do not have morally significant consequences (e.g., whether we should use a Fahrenheit or Celsius temperature scale), and consequential choices that are clearly determined by the evidence (e.g., whether global temperature should be measured through instruments or subjective reports).

Value-laden choices in designing quantitative scales abound in many areas of science. Prominent cases typically fall into two overlapping classes. A first class of cases concerns the quantification of “thick” qualitative concepts that serve descriptive *and* evaluative purposes (Alexandrova 2018; Alexandrova and Fabian 2022). Here, researcher make choices about the definition of their measurand and its indicators that determinate which (and whose) moral values are reflected in their measure. Should our questionnaires assume that well-being is simply the ratio of reported positive to negative emotions in a standardized questionnaire or is

---

<sup>1</sup> In some cases (e.g. mass, length, or thermodynamic temperature) we may be further justified in employing a ratio scale, which allows comparison via division and multiplication.

it a complex aggregate of disposable income, employment rate, self-reported health, and several other indicators? If the measure should preserve these different evaluative components, how much weight should scientists assign to each component when calculating a final well-being number?

A second, related class of cases concerns choices between alternative indicators and definitions that lead to highly differential consequences for different stakeholders. Such choices can occur even if our pre-quantitative concepts are purely descriptive and do not usually express moral valuations. For example, carbon emissions per country can be defined and measured in alternative, purely descriptive ways, but these alternative definitions will lead to highly different policy outcomes. The CO<sub>2</sub> emitted in the lifecycle of a commodity may be assigned to the country it was produced or consumed in, very likely leading to different policy decisions about who is held politically accountable for these emissions (Karakaya, Yılmaz, and Alataş 2019).

Two proposals for dealing with value-laden choices in quantification have long dominated the scientific and philosophical literature. The first proposal is a simple extension of an influential account of value-neutrality in scientific research (e.g., Nagel 1961, 492–93): scientists can and should avoid value judgements by making them conditional on a standardized measurement procedure and a standardized definition of the measurand. Science funders or policy makers may worry about “appraising” such standards on ethical grounds but scientists themselves merely “estimate” how well they are approximated by empirical phenomena. The appeal of this approach is apparent: we can pursue the epistemic benefits of quantification without deferring undue authority to scientists.

The second influential proposal can be read as a response to the first proposal: reject attempts at quantification because they confer illegitimate power to scientific experts. Drawing on historical and contemporary cases, many researchers have become convinced that value judgements cannot be avoided by standardization and, for the sake of democracy or justice, should not be black-boxed by seemingly objective standards. Newfield, Alexandrova, and John (2022) call this the “original critique” of quantification. It is implicit (though rarely endorsed) in classic scholarship in history and social studies of science (Hacking 1990; Rusnock 1995; Porter 1996; Muller 2018).

Few philosophers today endorse either of the two canonical proposals. Instead, recent work has taken inspiration in a growing literature on the role of moral and political values in

science. The result has been a new-found optimism that we can exploit the epistemic benefits of quantification while avoiding its political risks. The proposed solution is what I will call the alignment approach.

### 3. The Alignment Approach

If quantification involves value-laden choices, institutions designing quantitative measures wield power over those affected by the outcomes of measurements. We frequently defer power to institutions like governments, courts, or even supranational entities if we consider these institutions politically legitimate. It has been a key aim of recent work on quantification to spell out how quantification may be legitimized.<sup>2</sup> The result is a growing literature aimed at combining the pursuit of quantitative measurement with norms governing the epistemically unforced and consequential choices made during measure construction. Recent contributions to this literature explicitly or implicitly commit to a specific criterion for political legitimacy:

**Legitimacy<sub>A</sub>:** Measures are politically legitimate if the value-judgements during their design were subject to procedural constraints that aligned those judgements with the values of relevant stakeholders.

Legitimacy<sub>A</sub> forms the core of what I will call the “alignment approach” to value-laden quantification – adopting a notion used widely in debates about the role of moral and political values in science (Parker and Lusk 2019; Schroeder 2021; Elabbar 2023). Legitimacy<sub>A</sub> can be operationalized in different ways, depending on how narrowly the class of appropriate procedures is drawn and who is considered as the relevant stakeholders of measurements. I will spell out the different proposal for operationalizing it in what follows, if only to show that all of them presume a basic but unsubstantiated premise to be criticised in the following section.

#### 3.1 *Participative versus indirect alignment*

Proponents of alignment approaches disagree on how directly stakeholders have to confer legitimacy on quantitative measures. Some philosophers argue that scientists can legitimize value-laden choices by relying on surveys of stakeholder values. Others call for alignment

---

<sup>2</sup> This desideratum is spelled out most clearly in Alexandrova 2018 and Fabian and Alexandrova 2022.

procedures that directly involve the relevant stakeholders in designing, executing, and evaluating measurements. I will illustrate the distinction by discussing some exemplary approaches that fall on different ends of the spectrum.

Alexandrova's (2018, 440) "deliberative polls of normative presuppositions" illustrate an indirect alignment procedure. She argues that such polls can ground the procedural objectivity of value-laden measures. Her example is the UK Office of National Statistics's (ONS) "What matters to you?" online poll, in which citizens could judge the normative aptness of different well-being measures. The ONS then consulted the survey responses to make value-laden choices between alternative measures. For Alexandrova "the honest effort to canvass the diverse views shows that the value presuppositions on this measure have arguably passed the sort of test I have in mind." (Alexandrova 2018, 439).

There are two recent pilot projects that attempt to align quantitative measures with stakeholder values by directly involving them in the measurement process. The first is Fabian and Alexandrova's (2022) proposal for an "ideal of participatory measurement." Their participatory measure of well-being involved employees and beneficiaries of a UK charity. Beneficiaries contributed directly to the choice of a definition of well-being and helped choosing questionnaire items to calculate a well-being value. Duque, Tal, and Barbic (2024) have also argued for a direct involvement of stakeholders in psychosocial measurement. They launched an exemplary project with a Canadian mental health services provider, which aimed to implement collective "ethical iterations." In such iterations, scientists, service providers, and participants collectively deliberate the appropriate values for making epistemically unforced choices and revise these choices in light of service outcomes.

### *3.2 Wide versus narrow scoping of stakeholders*

Proponents of the alignment approach also disagree how to best scope the relevant group of stakeholders. This disagreement, essentially, concerns whether the values of users and subjects of measurement deserve any privileged consideration over and above the values of the general public. Philosophers linking to legitimacy to alignment with user and subject values endorse *narrow scoping*, while philosophers urging alignment with public values endorse *wide scoping*.

The two pilot projects of participatory measurements are excellent examples of narrow scoping (Alexandrova and Fabian 2022; Rodriguez Duque, Tal, and Barbic 2024). In both cases, the processes used to align epistemically unforced choices with stakeholder values involved scientists, service providers, and, most importantly, subjects of quantitative well-being or mental health assessments.

Schroeder (2019) endorses wide scoping for economic measures. For him, epistemically unforced choices in quality-of-life measures should be based on egalitarian considerations because most citizens hold egalitarian views about health. Schroeder has argued repeatedly that wide scoping is strongly preferable to narrow scoping in and beyond quantitative measurement (Schroeder 2017, 2021). If scientists adapt their value-laden choices to specific groups of stakeholders, they risk undercutting public trust in science. With few exceptions (explicitly anti-democratic values such as racism), value-laden choices should therefore be aligned with democratic values – i.e., the values held by the public of a particular democratic society or its elected representatives (Schroeder 2022).

It should now be clear that there are several influential proposals for how value-laden choices in quantification may be legitimized by aligning them with stakeholder values. Philosophers disagree which procedure for value alignment can secure political legitimacy and scope the class of relevant stakeholders differently. My goal is not to intervene in these debates. Instead, I will argue that they share a basic but unsubstantiated premise.

### *3.3 The basic premise of the alignment approach*

Defenders of the alignment approach value quantitative measurement but share traditional critics' sense that quantification can be highly consequential and, hence, requires political legitimacy. They are optimistic that their favoured procedures for value alignment can confer political legitimacy on quantitative measures. This optimism raises a basic but crucial question that, to my knowledge, has not been addressed in the literature. Values differ across groups, countries, and time. Quantitative measures need to be adjusted to different groups, countries, or historical periods to not sacrifice their political legitimacy. Clearly, however, quantification is incredibly hard, and even physical scientists often took decades or centuries to develop a quantitative scale that passes empirical scrutiny (Yoder 1989; Chang 2004; Schlaudt 2009; Sherry 2011; Luchetti 2020). The question, then, is whether quantification is empirically unconstrained enough to be achievable along such stakeholder-specific pathways.

Surely, we can define and operationalize mental health or well-being in line with stakeholder values, but can the quantitative scales thus constructed scales pass empirical scrutiny?

**The basic premise of the alignment approach to quantification:** The process of quantification is sufficiently unconstrained empirically for quantitative measurement to be achievable along multiple, stakeholder-specific pathways.

Proponents of the alignment approach presume that the above question can be answered positively but provide no evidence for such a proposition. The resulting optimism about legitimate and successful quantification rests on a premise in need of scrutiny. It is, indeed, the *basic premise* of the alignment approach since it sets it apart from both traditional proponents and traditional critics of quantification.

Existing arguments for aligning quantitative measures are, in principle, not sufficient to substantiate the basic premise. Such arguments take two forms. First, some proponents of value alignment are content with criticising the harmful consequences of aspiring towards an unattainable value-free ideal (Bocchi forthcoming) or arguing for the superiority of their specific alignment procedures (Schroeder 2021). Such work provides valuable insights into merits of specific alignment procedures vis-à-vis the value-free ideal or competing proposals. It does not bear on the epistemic warrant for such stakeholder-dependent scales correctness and, ipso facto, the basic premise of value alignment.

The second type of argument provided for value-alignment takes the form of practical demonstrations (Alexandrova and Fabian 2022; Duque, Tal, and Barbic 2024). Even concrete pilot projects for aligning measures, however, fail to substantiate the basic premise, because they concern psychosocial attributes whose quantitative status remains highly questionable on empirical grounds (Michell 1999; Trendler 2009; 2019; Wolff 2023). Contrary to cases of successful quantification, (quantitative) distance relations postulated between well-being or mental health have not been tested via stable, convergent, and sufficiently precise measurements, nor have such measurement led to the discovery of further quantitative details of the empirical world. The pilot projects demonstrate realistic procedures for aligning value-laden choices in quantification scales with stakeholder values. They do not demonstrate that the resulting scales of well-being and mental health will eventually satisfy the empirical requirements of quantitative measurement.

## 4. Against Alignment

In the previous section, I have argued that existing versions of the alignment approach do not substantiate that approach's basic premise. The most developed proposals for value-alignment focus on measures whose quantitative status remains disputed. My approach in what follows is the inverse. I focus on a rare case of a value-laden but epistemically successful quantification: seismological scales of earthquake size. Seismology's historical development motivates my counterfactual contention that we would have not been able to quantify earthquake size along stakeholder specific pathways. This suggests that the basic premise is, at least sometimes, false.

### 4.1 Quantification in Seismology

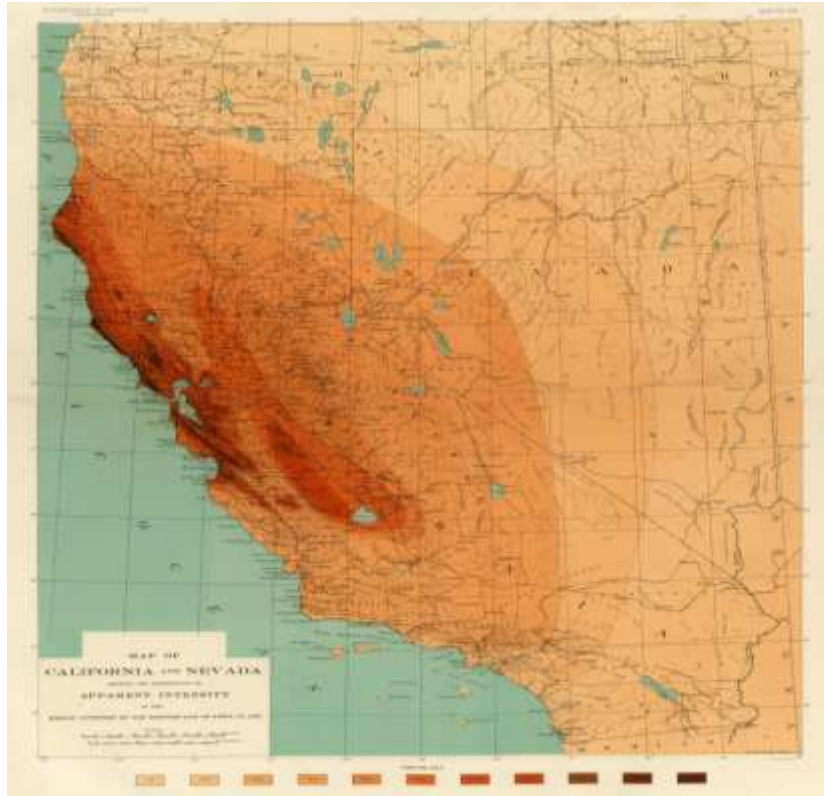
It is a central aim of modern seismology to measure 'earthquake size'. The first widely used scales of earthquake size were intensity scales. Their basic assumption is that earthquake size is proportional to ground motion and its direct effects. For example, the famous Rossi-Forel Scale (fig. 1) ranks earthquakes by sorting them into ten different levels of intensity based on criteria ranging from seismograph readings to "general panic" or the occurrence of a "great disaster." Intensity is a spatial quantity, meaning that the same earthquake will have different intensities at different places.

- 
- 
- I. Recorded by a single seismograph or by some seismographs of the same model but not by several seismographs of different kinds; the shock felt by an experienced observer.
  - II. Recorded by seismographs of different kinds; felt by a small number of people at rest.
  - III. Felt by persons at rest; strong enough for the duration or the direction of the shaking to be appreciable.
  - IV. Felt by persons in motion; disturbance of movable objects, doors, windows; cracking of ceilings.
  - V. Felt generally by everyone; disturbance of furniture and beds; ringing of some bells.
  - VI. General awakening of those asleep; general ringing of bells, oscillation of chandeliers, stopping of clocks; visible disturbance of trees and shrubs. Some startled persons leave their dwellings.
  - VII. Overthrow of movable objects, fall of plaster, ringing of church bells, general panic, without damage to buildings.
  - VIII. Fall of chimneys, cracks in the walls of buildings.
  - IX. Partial or total destruction of some buildings.
  - X. Great disaster, ruins, disturbance of strata, fissures in the earth's crust, rock falls from mountains.
- 
- 

**Fig.1:** Rossi-Forel Scale of Earthquake Intensity. Reproduced from Howell 2005.

Intensity scales clearly do not pass empirical tests of quantitateness. Even conceptually, a close reading of the Rossi-Forel Scale suggests that there is no clear way to compare distances between different intensities additively. Why would we assume that the increase in intensity from level III to V is of the same size as that from levels VII to VII? If it is not of the same size, we cannot compare earthquakes by adding or subtracting their intensities. As seismologists soon realised, the different indications in the scale also do not covary monotonically, meaning that seismograph recordings, social reactions, and architectural destruction may lead to different intensity assignments for the same earthquake at the same location.

Pre-quantitative scales were nonetheless very valuable to stakeholders and often designed in close collaborations with citizens in earthquake-prone regions (Coen 2013). The indications for different intensity levels were based on events that stakeholders cared about and could verify themselves. Recording these indications (through seismographs *and* subject reports by affected citizens) allowed scientists to track and understand the physical and social effects most threatening to them. Mapping of intensity across different locations provided an excellent method for locating seismic faults and identifying future risk areas. For that reason, the main media for communicating intensity were so-called isoseismal map, which can successfully guide future hazard mitigation through relocating citizens or implementing special construction laws (fig. 2). Intensity scales are a fantastic example of successful value-alignment in measurement, where stakeholder values affected the choice of indicators and directly participated in the process of measurement by submitting intensity reports and qualitative commentary to scientists.



**Fig. 2:** Isoseismal Intensity Map of the 1906 San Francisco Earthquake, where different shadings represent different intensities on the Rossi-Forel Scale. Reproduced from Lawson et. al 1908.

Intensity scales gradually lost importance after Charles Richter published the first magnitude scale of earthquake size in 1935. Magnitude scales assign a single value to each earthquake, which can be inferred from the maximum amplitude  $A$  of a seismogram at a distance  $d$ , to the epicentre of an earthquake:

$$M = \log_{10} \left[ \frac{A}{A_0(d)} \right], \quad (1)$$

where  $A_0$  is the conventionally chosen amplitude of a standard shock (assigned 0M) at  $d$  (Richter 1935). Initially the scale was only used at epicentral distance up to about 600km in Southern California and, by implication, based on amplitudes of specific seismic waves (mostly shear body waves) originating in earthquakes with a specific hypocentral depth. Richter and his colleague Beno Gutenberg subsequently extended the scale for use across the world. They achieved this feat by modelling the effects of variations in Earth's internal structure and crust, earthquake focus depth, wave-type, and instrument design on recorded amplitude (Gutenberg and Richter 1942; 1956).

Seismologists' hope was that the magnitude scale approximately tracked the energy released during an earthquake, providing a single physical standard for comparing earthquakes across time and space. This hope was eventually realised by the scale's successor, Keiiti Aki's moment scale, which directly links a specific spectrum in the seismogram to an earthquake's moment  $M_0$  – the product of the area, displacement, and rigidity of the fault dislocation – a reasonable approximation of the released energy:

$$M_0 = 4\pi p v_s^3 \Delta \Omega_0, \quad (2)$$

where  $\Omega_0$  is a particular shear wave signature in the seismogram,  $p$  is the density of the transmission media in the Earth's body or crust,  $v_s$  is the shear wave pulse's velocity, and  $\Delta$  is the distance to the hypocentre. Earthquakes could now be compared based on a (modeled) parameter of their physical mechanism rather than their heterogenous effects at Earth's surface. The moment of specific earthquakes can be measured consistently across regions and instruments, there is a clear physical meaning to distances between values on the scale, and the scale has been successfully used to discover quantitative details of Earth's internal structure (Miyake 2017; Di Giacomo, Harris, and Storchak 2021).

The key takeaway of this history is that quantifying earthquake size involved redefining what “earthquake size” meant. The moment of an earthquakes refers to the physical mechanism at the earthquake source, while intensity referred to ground motion and its destructiveness. This semantic change went hand in hand with a change in what data the scales produced and whose interests these data served. While intensity is indexed to particular places across an affected area, its quantitative successor scales assign only one number to each earthquake that is indexed to its hypocentre. The new and elegant quantitative values for earthquake size were of little use in mapping hazard risk and, ipso facto, mitigating those risks in future earthquakes. Indeed, twenty-first century seismologists had to revive intensity scales based on subjective reports to make progress in mitigating earthquake risk and better understand the physical links between earthquake moment and local ground motion (Hough 2000; Atkinson and Wald 2007).

#### *4.2 The Alignment Dilemma*

The development of seismology illustrates that heavily value-laden concepts can be quantified successfully. It suggests, however, that such quantification might require scientists

to make epistemically unforced choices that do *not* align with stakeholder values, hence sacrificing legitimacy<sub>A</sub>. There is no purely epistemic reason for pursuing a quantitative scale instead of qualitatively recording, understanding, and mitigating earthquake effects. Yet, if quantification is the goal, such a redefinition appeared necessary – after all, intensity had previously been studied for about century without any clear progress towards quantification. Seismologists excluded stakeholders and their preferences from measurement design and evaluation, and redefined earthquake size to refer to a property of overwhelmingly theoretical and surprisingly little practical interest.

It is a widely known insight from other historical studies that quantitative measurement often involves serious meaning change and that this meaning change is heavily constrained empirically (Chang 2004; Smith and Seth 2020). What seismology illustrates is that such meaning change may well require scientists to abandon stakeholder values. Since we have no positive example of quantification along stakeholder-specific pathways, this observation calls into question the basic premise of the alignment approach to quantification. If the basic premise is not correct, proponents of the alignment approach are left with a dilemma, which can be put in argument form as follows.

**The Alignment Dilemma:**

**(P1)** Scientists should align value-laden choices in the design of quantitative measures with stakeholder values.

**(P2)** Quantifying attributes requires scientists to alter the definition and scope of pre-quantitative concepts.

**(P3)** Frequently, the stakeholder values served by pre-quantitative concepts cannot be served by their quantitative successors (the basic premise is, frequently, false).

**(C):** Frequently, Scientists should not quantify attributes.

Note that the point of this argument is not to convince you of its conclusion. Rather, it maps out a dilemma: if you want to pursue quantification (hence, not accept C), you cannot accept the argument illustrated by my case study (P2 & P3) *and* require scientists to align value-laden choices with stakeholder values (P1). The dilemma suggests that the alignment approach to quantification will often, if not always, be unfeasible on empirical grounds.

## 5. Conclusion

I have argued that the alignment approach to quantification fails because it relies on an unsubstantiated premise. Its proponents assume that quantification is sufficiently unconstrained to be achievable among multiple, stakeholder-specific pathways. The history of seismology – a rare success case of value-laden quantification – casts strong doubts on this premise. Quantification is difficult to achieve on empirical grounds and will frequently require choices that do not align with stakeholder values.

This conclusion does not call into question the basic intuition of recent work on quantification: we should insist that quantitative measures meet demands of political legitimacy. Twentieth-century seismology illustrates the shortcomings of the alignment approach to quantification, but is surely no model of how quantification *ought* to look like (for similar views: Hough 2000; Coen 2013). The challenge, as I see it, is to work out a criterion of legitimacy that grants researchers sufficient freedom to make non-aligned choices during quantification but ensures that quantitative scales serve the interests of stakeholders in the medium or long term. Such a criterion should draw from a recent interest in the diverse sources of institutional legitimacy beyond the nation state by political philosophers (Adams 2018). I lack the space to spell out such a criterion here but hope to have illustrated its relevance.

## Acknowledgements

Special thanks are due to Cristian Larroulet Philippi for many hours of discussing measurement in seismology together, and to Ahmad Elabbar for pushing me to write up my worries about value alignment in measurement. For feedback on earlier versions of this paper I thank Aja Watkins, Alisa Bokulich, Anna Alexandrova, Eran Tal, Federica Bocchi, Pieter Beck, and Ruward Moulder. Finally, I thank all other attendees of the Values in Science workshop at Pembroke College, University of Cambridge in March 2024 and the 'Science and the Public' session at the PSA 2024, members of the Philosophy of the Geosciences Research Group at Boston University and Hasok Chang's research group at the University of Cambridge, all of whom provided me with helpful questions and comments. The research for this paper was funded by the British Academy, PFSS24\240061.

## References

- Adams, N. P. 2018. 'Institutional Legitimacy'. *Journal of Political Philosophy* 26 (1): 84–102. <https://doi.org/10.1111/jopp.12122>.
- Alexandrova, Anna. 2018. 'Can the Science of Well-Being Be Objective?' *The British Journal for the Philosophy of Science* 69 (2): 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Alexandrova, Anna, and Mark Fabian. 2022. 'Democratising Measurement: Or Why Thick Concepts Call for Coproduction'. *European Journal for Philosophy of Science* 12 (1): 7. <https://doi.org/10.1007/s13194-021-00437-7>.
- Atkinson, G. M., and D. J. Wald. 2007. "'Did You Feel It?' Intensity Data: A Surprisingly Good Measure of Earthquake Ground Motion'. *Seismological Research Letters* 78 (3): 362–68. <https://doi.org/10.1785/gssrl.78.3.362>.
- Bocchi, Federica. forthcoming. 'Metrics in Biodiversity Conservation and the Value-Free Ideal'. *Synthese*. <https://philsci-archive.pitt.edu/23136/>.
- Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. New York, NY: Oxford Univ. Press.
- Coen, Deborah. 2013. *The Earthquake Observers – Disaster Science from Lisbon to Richter*. Illustrated edition. Chicago ; London: University of Chicago Press.
- Di Giacomo, Domenico, James Harris, and Dmitry A. Storchak. 2021. 'Complementing Regional Moment Magnitudes to GCMT: A Perspective from the Rebuilt International Seismological Centre Bulletin'. *Earth System Science Data* 13 (5): 1957–85. <https://doi.org/10.5194/essd-13-1957-2021>.
- Duque, Sebastian Rodriguez, Eran Tal, and Skye Pamela Barbic. 2024. 'The Role of Ethical and Social Values in Psychosocial Measurement'. *Measurement* 225 (February): 113993. <https://doi.org/10.1016/j.measurement.2023.113993>.
- Elabbar, Ahmad. 2023. 'The Curatorial View of Assessment and the Ethics of Scientific Advice: Beyond Decisional Autonomy towards Distributive Epistemic Justice'. IUHPST. 2023. <http://iuhpst.org/pages/inter-division-commissions/joint-commission/2023-essay-prize.php>.
- Gutenberg, B., and C. F. Richter. 1942. 'Earthquake Magnitude, Intensity, Energy, and Acceleration\*'. *Bulletin of the Seismological Society of America* 32 (3): 163–91. <https://doi.org/10.1785/BSSA0320030163>.

- . 1956. ‘Earthquake Magnitude, Intensity, Energy, and Acceleration: (Second Paper)’. *Bulletin of the Seismological Society of America* 46 (2): 105–45. <https://doi.org/10.1785/BSSA0460020105>.
- Hacking, Ian. 1990. *The Taming of Chance*. Ideas in Context. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511819766>.
- Helmholtz, Hermann. 1887. ‘Zählen und Messen, erkenntnisstheoretisch betrachtet’. In *Philosophische Aufsätze, Eduard Zeller zu seinem fünfzigjährigen Doctorjubiläum gewidmet.*, 17–52. Leipzig: Fues.
- Hough, Susan E. 2000. ‘On the Scientific Value of “Unscientific” Data’. *Seismological Research Letters* 71 (5): 483–85. <https://doi.org/10.1785/gssrl.71.5.483>.
- Howell, Benjamin F. 2005. *An Introduction to Seismological Research: History and Development*. Cambridge ; New York: Cambridge University Press.
- Karakaya, Etem, Burcu Yılmaz, and Sedat Alataş. 2019. ‘How Production-Based and Consumption-Based Emissions Accounting Systems Change Climate Policy Analysis: The Case of CO2 Convergence’. *Environmental Science and Pollution Research* 26 (16): 16682–94. <https://doi.org/10.1007/s11356-019-05007-2>.
- Lawson, Andrew Cowper, and California Earthquake Commission. 1908. *Atlas of maps and seismographs: accompanying the report of the State Earthquake Investigation Commission upon the California earthquake of April 18, 1906*. Washington, D.C: Carnegie Institution of Washington.
- Luchetti, Michele. 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): 119–31. <https://doi.org/10.1016/j.shpsa.2020.09.005>.
- Michell, Joel. 1997. ‘Quantitative Science and the Definition of Measurement in Psychology’. *British Journal of Psychology* 88 (3): 355–83. <https://doi.org/10.1111/j.2044-8295.1997.tb02641.x>.
- . 1999. *Measurement in Psychology: A Critical History of a Methodological Concept*. New. New York: Cambridge University Press.
- . 2019. ‘Conjoint Measurement Underdone: Comment on Günter Trendler (2019)’. *Theory & Psychology* 29 (1): 138–43. <https://doi.org/10.1177/0959354318814962>.
- Miyake, Teru. 2017. ‘Magnitude, Moment, and Measurement: The Seismic Mechanism Controversy and Its Resolution’. *Studies in History and Philosophy of Science* 65–66 (December): 112–20. <https://doi.org/10.1016/j.shpsa.2017.02.002>.
- Muller, Jerry Z. 2018. *The Tyranny of Metrics*. Princeton Oxford: Princeton University Press.

- Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World.
- Newfield, Christopher, Anna Alexandrova, and Stephen John, eds. 2022. *Limits of the Numerical: The Abuses and Uses of Quantification*. Chicago, IL: University of Chicago Press.  
<https://press.uchicago.edu/ucp/books/book/chicago/L/bo146791774.html>.
- Parker, Wendy S., and Greg Lusk. 2019. 'Incorporating User Values into Climate Services'. *Bulletin of the American Meteorological Society*, April. <https://doi.org/10.1175/bams-d-17-0325.1>.
- Porter, Theodore M. 1996. *Trust in Numbers – The Pursuit of Objectivity in Science and Public Life*. Reprint edition. Princeton, N.J: Princeton University Press.
- Richter, Charles F. 1935. 'An Instrumental Earthquake Magnitude Scale\*'. *Bulletin of the Seismological Society of America* 25 (1): 1–32.  
<https://doi.org/10.1785/BSSA0250010001>.
- Rusnock, Andrea. 1995. 'Quantification, Precision, and Accuracy: Determinations of Population in the Ancien Régime in France'. In *The Values of Precision*, edited by M. Norton Wise, 17–38. Princeton, NJ: Princeton Univ. Press.
- Schlaudt, Oliver. 2009. *Messung als konkrete Handlung: Eine kritische Untersuchung über die Grundlagen der Bildung quantitativer Begriffe in den Naturwissenschaften*. Würzburg: Königshausen u. Neumann.
- Schroeder, S. Andrew. 2017. 'Using Democratic Values in Science: An Objection and (Partial) Response'. *Philosophy of Science* 84 (5): 1044–54.  
<https://doi.org/10.1086/694006>.
- . 2019. 'Which Values Should Be Built Into Economic Measures?' *Economics and Philosophy* 35 (3): 521–36. <https://doi.org/10.1017/s0266267118000317>.
- . 2021. 'Democratic Values: A Better Foundation for Public Trust in Science'. *The British Journal for the Philosophy of Science* 72 (2): 545–62.  
<https://doi.org/10.1093/bjps/axz023>.
- . 2022. 'The Limits of Democratizing Science: When Scientists Should Ignore the Public'. *Philosophy of Science* 89 (5): 1034–43. <https://doi.org/10.1017/psa.2022.54>.
- Sherry, David. 2011. 'Thermoscopes, Thermometers, and the Foundations of Measurement'. *Studies in History and Philosophy of Science Part A* 42 (4): 509–24.  
<https://doi.org/10.1016/j.shpsa.2011.07.001>.

- Smith, George E. 2014. 'Closing the Loop: Testing Newtonian Gravity, Then and Now'. In *Newton and Empiricism*, edited by Zvi Biener and Eric Schliesser. Vol. 262–353. Oxford: Oxford Univ. Press.
- Smith, George E., and Raghav Seth. 2020. *Brownian Motion and Molecular Reality*. Oxford Studies in Philosophy of Science. Oxford, New York: Oxford University Press.
- Tal, Eran. 2019. 'Individuating Quantities'. *Philosophical Studies* 176 (4): 853–78. <https://doi.org/10.1007/s11098-018-1216-2>.
- Thalos, Mariam. 2023. 'The Logic of Measurement: A Defense of Foundationalist Empiricism'. *Episteme*, June, 1–26. <https://doi.org/10.1017/epi.2023.32>.
- Trendler, Günter. 2009. 'Measurement Theory, Psychology and the Revolution That Cannot Happen'. *Theory & Psychology* 19 (5): 579–99. <https://doi.org/10.1177/0959354309341926>.
- . 2019. 'Conjoint Measurement Undone'. *Theory & Psychology* 29: 100–128. <https://doi.org/10.1177/0959354318788729>.
- Vessonen, Elina. 2020. 'The Complementarity of Psychometrics and the Representational Theory of Measurement'. *The British Journal for the Philosophy of Science* 71 (2): 415–42. <https://doi.org/10.1093/bjps/axy032>.
- Wolff, J. E. 2023. 'Coordination and Obsolescence: A Response on Behalf of Measurement Realism'. *Synthese* 201 (3): 78. <https://doi.org/10.1007/s11229-023-04061-1>.
- Yoder, Joella G. 1989. *Unrolling Time: Christiaan Huygens and the Mathematization of Nature*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511622441>.