

**Marcel Boumans.** *Science Outside the Laboratory. Measurement in Field Science and Economics.* xi + 198 pp., figs., tables, bibl., index. New York: Oxford University Press, 2015. £38.99 (cloth).

Book Review published in *Isis*, 108, 4, 880-881.

The accurate measurement of phenomena in field sciences, where controlled laboratory conditions are generally unavailable, has been a challenging subject for generations of philosophers and scientists. Boumans' book situates this subject within a detailed historical and philosophical enquiry. Drawing on the work of historically important figures such as the actuary Corneille Landré, and economists like Tjalling Koopmans and Trygve Haavelmo, the book sets out the development of philosophical, scientific enquiry into the nature of measurement and statistics for science conducted outside the laboratory, particularly in economics.

The book is first and foremost a philosophical text. Boumans' uses this historical cast of actors to argue that measurement in field sciences can be assessed in terms of precision, reliability and rigour only with the additional incorporation of expert judgement. This is because measurands have both known and unknown components determined through what he describes as a 'calculus of observations' in which objective and subjective judgement are enrolled. Given that statistics therefore cannot be sufficient to explain all the possible influences on a measurand, expert judgement is critical to defining reliable measurement. This opens up an important question: are all experts equal? Boumans' argues that experts can be validated like models with consensual agreements (aggregation of expert views) produced within teams of experts within specific scientific institutions. This is the central contribution of the book.

The book made a timely appearance on my desk just as the U.K. Royal Statistical Society approved the formal establishment of a History Section within the Society. Boumans' work demonstrates the value of thinking historically about statistical questions. Whether in considering Landré and Guillaume Mounier's critique of the applicability of least squares method to actuarial science, or in discussing Haavelmo's development of significance, or in Amos Tversky and Daniel Kahneman's account of representativeness and bias, there is plenty to engage with for those interested in the history of statistics.

The book is mostly written towards an audience of academics interested in the philosophy of measurement though field scientists and economists should appreciate the extensive discussion of the reliability of observation and the challenge of justifying empirics produced through expert and therefore subjective judgement. It unfolds with a discussion of the evaluation of measurement (chapter 2) before establishing the centrality of expert judgement in field observations, drawing on for example Buys Ballot work in meteorology (chapter 3). Chapter 4 focuses on economics and the two century struggle to define if economic science could be purely inductive, an answer that is largely negative. Theory is therefore needed and Boumans ponders how expert judgement vis-à-vis theory should be presented. Drawing on medical science and the parallels with clinical judgement, in chapter 5 Boumans suggests that subjective judgement is essential to field measurement, but can be empirically validated not least through assessing the production of expert consensus within scientific institutions (the subject of chapter 6).

As a reader more interested in historical aspects, I felt the book might have expanded the discussion of historical context further. While some actors are given considerable historical treatment, not least among them the 19<sup>th</sup> century polymath Buys Ballot, other later figures are given a more cursory historical treatment. I kept thinking about the kinds of context that Philip Mirowski would wish to highlight, including the role of institutions such as RAND, the development of economics as a discipline with its allied political projects, the treatment of knowledge within a marketplace of ideas (a rather different form of consensus 'judgement'), and the cultural resonance of physics-envy within econometrics. These are minor quibbles, but would help enhance the historical relevance of the book.

Boumans' book prompted a couple of other important questions for me. First, I was interested to know more about whether expert judgement of the success of past predictions in economics might be problematic given that economic predictions can shape economic outcomes. Second, I wondered about the kinds of expert judgements made in scientific institutions and I would have liked to hear more about what kinds of institutions were envisaged. For instance, does the Intergovernmental Panel on Climate Change (IPCC) count as such an institution? Climate scientists regularly review their predictions and well-established scientific research is given a higher weight in the internal aggregation of expertise within the IPCC reports. At the same time, the IPCC reports are also subjective relying on the (inclusion or exclusion) of particular (kinds of) experts in each panel. A rather different model is preferred in the Intergovernmental Platform on Biodiversity

and Ecosystem Services (IPBES) that takes a more inclusive approach to expertise globally and which invites open, external review. Drawing on existing cases of scientific institutions like these, it would be interesting to consider how Boumans' analysis represent the ideals envisaged for these expert judgements and how this philosophical enquiry might translate into the everyday practice of these institutions.

Samuel Randalls