



SCIENCE OUTSIDE THE LABORATORY

MARCEL BOUMANS

Reviewed by Catherine Herfeld

Science outside the Laboratory: Measurement in Field Science and Economics

Marcel Boumans

Oxford: Oxford University Press, 2015, £38.99

ISBN 9780199388288

In this book, Marcel Boumans takes up an under-researched topic in philosophy of science, namely, the reliability of measurement in field science and the rules that need to be met to ensure such reliability outside the laboratory. Generally, Boumans understands measurement as the attempt to acquire quantitative knowledge about a phenomenon of interest—an object or event called the ‘measurand’—by assigning numbers to its properties. To ensure that such numbers give us reliable information, specific rules are needed; for instance, to map a phenomenon by way of a model or formula requires procedures that ensure the reliability of this mapping. It is important to note that measurement encompasses both assigning numbers to a property of a phenomenon and the rules according to which we do this (p. 2). Boumans’s goal is to develop ‘an account of measurement for field science’ (p. 24), which he understands very broadly as ‘the varied range of research practices outside the laboratory’ (p. 2). Thereby, he takes seriously the idea that measurement in field sciences such as economics and meteorology is different from measurement in laboratory sciences such as physics and that the former must proceed according to rules and standards distinct from those of the latter.

Boumans introduces the concept of ‘clinical measurement’ to make clear one of the major problems of measurement in field sciences: to arrive at reliable measurements, the scientist has to strike the optimal balance between objectivity

and subjectivity (p. 116). Clinical measurement meets this challenge because it involves model-based procedures to ensure precision and attain calibration of measurement, on the one hand, and it requires that rational consensus be established among institutional experts, on the other. With this model-based account, Boumans's book is in the good company of recent work in philosophy of science that provide arguments for such approaches to measurement (see, for example, Tal [2015]). While Boumans does not engage with this literature extensively, *Science outside the Laboratory* offers many historical and methodological details from social field-sciences, and from economics in particular; Boumans remains faithful to his previous work that mostly involved modelling and measurement in economics. His approach in this book is inspired by what has been labelled historically informed 'philosophy of science in practice' (Ankeny *et al.* [2011]). This allows him to substantiate his account with sufficient detail about measurement practices that are normally 'inferred from almanacs, dictionaries, guides, handbooks, instructions, reports, teaching materials, tutorials, and yearbooks' (p. 173).

The major premise underlying Boumans's account is that the assessment of reliability in field science is different from the assessment of reliability in laboratory science because, unlike the latter, field science—for practical, technical, or ethical reasons—cannot isolate a phenomenon from its natural environment, which makes controlling for confounding factors and intervention hard or even impossible. Instead, the field scientist has to rely upon passive observation—observation without intervention or control—which, to be objective, must be managed in a certain way. The impossibility of conducting laboratory experiments implies that measurement in field science does not proceed by 'universal standards', where the implementation of an experiment is independent from context (p. 173). Rather, the standards and rules required for reliable measurement outside the laboratory are context-dependent, sensitive to local conditions. However, according to Boumans, such measurement standards are no less rigorous than those in the laboratory. Rather, the lack of control and systematic intervention in field science measurement is compensated for by the use of mathematical models representing the respective phenomenon and its environment, which serve as a 'virtual laboratory' for the field scientist (p. 24).

At times, it seems as if Boumans sees the 'methodological tension' between field and laboratory sciences clearly reflected in the difference in standards between natural and social sciences (p. 174). While natural science laboratories are 'scientific domains', the social sciences often study unique and non-reproducible phenomena, the result of historical processes. The 'field' is in the public domain and open to various actors having distinct interests. Those interests are often in conflict with the goals commonly ascribed to science and can have a significant influence on passive observation of the phenomenon in the field, qualifying those observations as subjective and prone to error. Boumans takes social statistics as being one example of this.

He begins his introduction with Oskar Morgenstern's views on the accuracy of such statistics and of economic observation more generally. Morgenstern had become sceptical about the possibility of arriving at accurate and reliable social statistics. One reason for this was the inexactness of economic theory used in passive observation; another was the hostile social and political environment during Nazi Germany, whose institutions often produced self-serving statistics, exemplifying for Morgenstern the subjective dimension of observation in the social sciences. In line with Morgenstern, Boumans sees this problem originating in the social scientist's dependence on data collected by institutions not following or concerned with good social scientific practice. The sheer amount of data that is needed to measure social phenomena prevents the field scientist relying on the observations of scientists alone. Social statistics that are collected by institutions with interests that differ from those of the social scientist can potentially lead to 'deliberate deceit' (p. 174), presenting an additional challenge.

While Boumans does not share Morgenstern's pessimism about improving the reliability of social statistics (p. 16), he acknowledges the challenges involved. He takes passive observation to be 'personal', which he understands as relying on rational credence, similar to Leonard Savage's approach. This personal—perhaps better, subjective—element in observation is the reason why the credibility of the observer, instead of the reproducibility of the results, is crucial for reliable measurement outside the lab. Here it becomes even more obvious that Boumans, like Morgenstern, takes the

difference between field and laboratory sciences to be one between inexact social sciences and exact natural sciences. (These distinctions are not the same, however: the distinction between field and laboratory sciences does not fully correspond to the distinction between the natural and social sciences.) According to Boumans, exactness presupposes complete causal knowledge that field science, which has to cope with a potentially large number of unknown unknowns, can never obtain. Thus, there is an inexactness in field science that must be dealt with, which the natural sciences do not confront.

In subsequent chapters, Boumans focuses on the implication of the fundamental distinction between field and laboratory sciences for questions about objectivity and rigour in the former (p. 16). How should we treat observations in field science in order to develop a methodology based upon standards from the social and not the natural sciences? To develop such an account of measurement, Boumans begins by discussing the prominent representational theory of measurement as first developed in its axiomatic form by Krantz *et al.* ([1971]). This representational account considers measurement to be a process of assigning numbers to properties of the measurand in such a way that relevant empirical qualitative relations among those properties are reflected in the numbers and in relevant properties of a numerical relational structure. The problem that such accounts have to address is the so-called representation problem, that is, formulating criteria to assess whether a model represents an empirical system and in what way it does so (p. 29). In Chapter 2, Boumans traces a shift in the historical development of the representational theory of measurement, away from an empirical interpretation of the axioms and towards their purely mathematical representation, that is, where their representation is articulated in terms of homomorphism. The problem with such a solution is that it requires white-box models and thus full information about the properties of the measurand. Such information is not possible in field science, which is why these models are not to be had there. Field science has to cope with uncertainty, which is why—according to Boumans—evaluations of such uncertainties require grey-box models: modular models whose modules are black boxes. The advantage of such models is that in order to test them, we do not need observations for each individual relationship specified by the model; our knowledge of the measurand thus neither has to be complete nor does it have to be exact. Yet, they nevertheless allow for validation tests to ‘identify and [...] estimate the magnitude of the uncertainty of neglected, ignored, or unknown influence quantities’ (p. 52). Thus, the resulting account of measurement is more a methodology for model building and model validation in cases of measurement outside the laboratory, rather than it is a systematic theory. Boumans arrives at his account in a piecemeal fashion, by taking from current theories and the measurement practices of metrology those tools and procedures useful for measurement outside the laboratory.

In Chapter 3, Boumans takes up the idea that measurement has two aspects, expressed in numbers. What we do know about the measurand is expressed in the measurement value and what we do not know is expressed in our assessment of the reliability of this value. This idea is embodied in various versions of the ‘calculus of observations’. Such a calculus is based on (1) a representation of the measurand and accompanying noise; (2) mechanical procedures for calibration and securing precision; (3) theoretical assumptions about regularities governing the measurand; and (4) trained judgement (p. 63). He discusses how, throughout the history of this calculus, those four elements changed as a result of failed attempts to arrive at objective measurement by replacing personal judgement with mechanical procedures. Such procedures can never fully deal with error and every model will always be underdetermined by the observations. That is why such procedures always have to be complemented by personal judgement.

Chapter 4 details some of the epistemological problems in field science that originate from its reliance on passive observation. By focusing on Trygve Haavelmo’s epistemology of econometrics, Boumans argues that statistics does not reliably arrive at a complete causal picture, because passively observed data does not necessarily display enough variation to allow field scientists to detect and measure the relevant causal variables of the mechanism behind a phenomenon of interest (p. 111). To make sure that one’s statistical model contains all relevant causal variables, guidance from theory is required but is not sufficient. Because social scientific theory is incomplete, additional information based upon expert judgement and experience is required to identify all the causally relevant variables.

Such expert judgement, so Boumans argues in Chapter 5, is not rational, which would require a general standard for determining the optimal outcome. Rather, to determine whether a measurement is biased, the field scientist must consider the target problem, the information structure, and the sample space; only then would the determination be 'scientific'. Scientific judgement is then based not only on objective but also on subjective knowledge. It requires expertise, knowledge about tools and techniques required for measurement, already established knowledge about the measurand, and theoretical knowledge. But at the same time, it requires imagination and intuition. For example, expert judgements are often grounded in thought experiments, which can help in the discovery of aspects of a phenomenon previous unobserved.

Judgements arrived at by way of objective but also subjective knowledge can obviously lead to disagreement in a scientific community; decisions have to be made about which experts should be trusted, as well as how to weigh their opinions. In the last chapter, in order to reach intersubjectivity in scientific communities, Boumans offers a methodology for such validation of subjective knowledge, alongside the validation of the measurement models. He labels his approach a 'model-based consensus'. It is grounded in the idea that diverging expert judgement should be assessed in the same way as black-box models. These judgements are based on how accurate the expert's judgements, grounded upon his or her thought experiments, have been in the past. On the basis of this evaluation, an expert's opinion is then weighted in the aggregation process of reaching consensus. To avoid problems related to lack of information about an individual expert's track record, Boumans concludes by suggesting that we should take the record of a team of experts affiliated with a scientific institution as the expert agency, and their judgement should become the subject of validation. Rational consensus among such experts in combination with model-based procedures to ensure precision and calibration is what constitutes clinical measurement. It is by way of such clinical measurement that we can achieve reliable measurement results outside the laboratory, according to Boumans.

This review omitted a lot of the historical detail Boumans carefully reconstructs to develop and justify his account of reliable measurement outside the laboratory; this book will be of interest to historian too, and not only to philosophers of science interested in the methodological debates around measurement in field science, in the challenges field sciences present, and in the various ways field scientists attempted to cope with those challenges. It is also an important book for philosophers with a specific interest in measurement practices in the social sciences and econometrics in particular, as well as for social field-scientists searching for new ways to ensure reliable measurement and ways to reach rational consensus.

Catherine Herfeld
University of Zurich
c.s.herfeld@gmail.com

References

Ankeny, R., Chang, H., Boumans, M. and Boon, M. [2011]: 'Introduction: Philosophy of Science in Practice', *European Journal of Philosophy of Science*, 1, pp. 303-7.

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. [1971]: *Foundations of Measurement*, Vol. 1, New York: Academic Press.

Tal, E. [2015]: 'Measurement in Science', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.