Erasmus Journal for Philosophy and Economics,
Volume 8, Issue 2,
Autumn 2015, pp. 102-110.
http://ejpe.org/pdf/8-2-br-3.pdf

Review of Marcel Boumans's *Science outside the laboratory: measurement in field science and economics. New York:* Oxford University Press, 2015, 198 pp.

ALESSANDRA BASSO
University of Helsinki

Science outside the laboratory brings the reader along an exploration into the epistemology of scientific investigation in the field. The book is organized in such a way as to move from general epistemological questions to more specific methodological issues, with each chapter diving deeper into a specific problem related to the previous chapters. In this review, I focus on a line of argument which I think cuts across the chapters and allows me to highlight the main contributions of the book.

Starting from the identification of the specific problems one faces when conducting scientific research in the field, *Science outside the laboratory* sets up to provide a positive answer to the question of whether it is possible to obtain reliable knowledge when laboratory investigation is not possible. In order to do so, Boumans provides an account of measurement of field phenomena and addresses methodological issues related to its validation. From this discussion, it emerges that the validation of field investigation requires field-specific assessment methods and, moreover, that expert judgment is needed in order to make reliable field investigation. The book, therefore, moves on to discuss what kind of judgment is needed, and what methods can be employed to validate experts judgments in the context of field investigation. This nicely closes the book's narrative circle and encourages the reader to look back at the opening question with an optimistic attitude: observation and measurement outside the laboratory can be rigorous, but it requires field-specific criteria for evaluating epistemic value. The suggestions offered for developing assessment methods that can be used to validate field investigation are, I believe, among the book's main contributions.

RELIABLE SCIENCE OUTSIDE THE LABORATORY

The methodological investigation offered in *Science outside the laboratory* is motivated by the attempt to deal with some of the problems that arise when investigating phenomena that, for practical, technical or ethical reasons, cannot be studied in the laboratory. The book uses the term ‘field science’ to refer to the variety of research practices outside the laboratory, which studies phenomena that cannot be isolated from their environment and hence cannot be investigated by means of manipulation or intervention (p. 2). Field science, therefore, seems to be a category of scientific practice that intersects the natural and the social sciences. However, some of the considerations made in the book appear to be related more to the opposition between natural and social sciences than to the contrast between field and laboratory science.

One aspect of the book where the issues at hand seem to cut across the social science/natural science divide rather than the field/laboratory distinction is the discussion of the challenges that emerge when conducting scientific research outside the laboratory. Boumans highlights in particular three problems, but only one of them appears to be specific to field science. First, a specific problem of field science is that, in contrast to the results of laboratory experiments, field results are not reproducible, and hence they are less objective, because they rely more on the credibility of the observer. Second, the gathered data is often based on evasive answers, lies or misunderstandings and hence research practice has to take into account the possibility of deliberate deceit from the individuals under investigation. This problem is not specific to field science: it does not arise when conducting field investigation in the physical sciences and, instead, it may come about in the social and behavioral sciences, whether conducted in the field or not. Nature does not lie, but individuals may well do so even if they are in the controlled environment of laboratory experiments.

Third, Boumans argues that field science is more inexact than laboratory science, leading to more incomplete theories and models (p. 3). As an example, he provides a thorough discussion of the difficulties related to the selection of the right set of variables to be included in economic models. Based also on the work of Trygve Haavelmo, Boumans argues that selecting the right set of variables is especially difficult due to the inexactness of economic theory. It seems to me that this kind of problem is not specific to field science. It may affect both the natural and the social sciences, although arguably the social sciences are more commonly affected by it. The incompleteness of theories and models in certain areas of science, however, might influence both laboratory and field investigation in these contexts. For instance, it might be argued that rational choice theory is unable to guide the choice of the right set of variables to be included in models of field phenomena like the business cycle, as well as in models of laboratory phenomena like preference change.

The intersection between the categories of field science and social science can be seen as providing potentially interesting perspectives for investigating the epistemology of field science and its relation to the debates regarding the social sciences. On the one hand, Boumans's focus on field science can offer a somehow new perspective to address some of the problems that affect the social sciences, and, on the other hand, some of his conclusions can be interpreted as contributing to general debates regarding the epistemology of the social sciences. However, to fully exploit this potential, it would have been interesting to see a clearer distinction between these two categories of scientific practice and a discussion of the relations between them.

The philosophical debates concerning the reliability of social scientific investigation are not foreign to social scientists themselves. As an example of how these problems are examined within social scientific disciplines, Boumans discusses Oskar Morgenstern's considerations about how economic observation is different from observation in the natural sciences. Two of Morgenstern's ideas are particularly relevant. First, the common assumption made in statistical regressions that errors average out over a large number of observations is not warranted for economic data, and neither is there good evidence to support it. Due to the specific nature of economic observations, and more generally of social phenomena, divergence in the distribution of data is to be expected, and hence one cannot assume that the data is normally distributed. Second, Morgenstern argues that scientific observation should be guided by theory, but economic theory is highly inexact and hence cannot adequately fulfill this task.

Morgenstern's analysis has a pessimistic tone and ends up being skeptical about the possibility of making reliable observations in economics. Boumans does not share this pessimism. For him, Morgenstern's skeptical conclusions are brought about by the economist's propensity to take the natural sciences as the gold standard of reliable observation, the reference against which economic observation should be compared. Boumans's analysis, instead, is based on the idea that the criteria for evaluating the reliability of field science should be local: the methods of field investigation should obey the standards of social science (p. 16).

AN ACCOUNT OF MEASUREMENT

In order to explore the methods that can be used to validate field investigation, Boumans begins by providing an account of measurement that can deal with the peculiar characteristics of field science. The account he develops is based on a prominent theoretical approach, the representational theory of measurement, and on the authoritative methodological guidance offered by current metrology. Since neither one is satisfactory when dealing with measurement of field phenomena, Boumans extends and adjusts them so as to arrive at a representational account of measurement that is consistent with inexact theory and incomplete models.

Representational accounts are based on the idea that measurement outcomes represent certain qualitative features of the system under measurement. The crucial question that these accounts have to address is what underpins the claim that the outcomes represent the measured property. The representational theory of measurement developed by Krantz, Luce, Suppes, and Tversky in the three-volume *Foundations of measurement*, takes an axiomatic approach to answer this question. The representational relation is established by providing a set of formal axioms about the structure of the property under measurement and a logical proof that, under these assumptions, the measurement outcomes are a homomorphic mapping of the qualitative structure of that property.

A criticism often raised against the axiomatic approach is that it lacks empirical bases, because the assumptions about the property under measurement are not tested empirically. To discuss this issue, *Science outside the laboratory* provides a brief historical excursus into the development of the axiomatic approach, which throws light on the progressive shift from an empirical understanding of the axioms to the essentially formal perspective of the representational theory of measurement. With Norman R. Campbell, and partially also with Stanley Smith Stevens, measurement is interpreted as being grounded on certain characteristics of the property of interest, like order and additivity, which are understood as hypotheses about facts of the world that have to be tested and proven empirically. With Suppes and coauthors, instead, the foundations of measurement are established under the condition that the empirical system has certain characteristics, without bothering with the empirical justification of these assumptions.

Although I agree with Boumans that the historical development of the axiomatic approach reveals a shift from an empirical to a formal understanding of the axioms, it should be noticed that, apart from relatively simple cases like the measurement of length, providing an empirical interpretation of the axioms does not necessarily mean that they are in fact tested empirically. Testing whether certain quantities, like loudness or happiness, satisfy the assumption of additivity, for instance, can be challenging or even impossible. Indeed, realist representational accounts of measurement, which argue for an empirical understanding of the axioms, maintain that there is little point in testing the axioms one by one, because a measurement is not required to rely only on true assumptions about a phenomenon in order for its outcomes to provide some true information about that phenomenon (Swoyer 1987).

Boumans argues that the axiomatic approach is not suitable for field measurement, independently of whether the axioms are tested or simply assumed. In his view, homomorphic representation requires a white-box model based on complete knowledge of the property under measurement. However, complete knowledge is not possible in field science, and hence a representational account of field measurement cannot be based on homomorphism. As an alternative, Boumans suggests extending the representational theory of measurement so as to encompass also cases in which one can aspire only to gray-box models, that is, modular models where the modules are black-boxes (p. 50).

Since gray-box models do not allow establishing homomorphism, the question of what grounds the claim that the outcomes are reliable representations should be addressed in a different way. Boumans's approach to answer this question is by looking at the practical guidance for the assessment of measurement reliability offered by current metrology.

In current metrology, the assessment of measurement is based on the evaluation of uncertainty. This is done partially by means of statistical methods and partially by other means, like skilled judgments and calibration against externally accepted references. The statistical methods are based on the evaluation of outcomes' stability under controlled variations of the measurement conditions, and therefore, according to Boumans, they are made for exploiting the controlled conditions of the laboratory. Since controlled variations might be impossible in the field, Boumans argues that the role of statistical methods in the assessment of field measurement should be narrower than in laboratory science. To compensate, non-statistical methods should have a broader role. Moreover, since field science lacks authoritative institutions able to fix standard references for calibration, the only available source of information for non-statistical assessments of field measurements is skilled judgment. As a consequence, the assessment of field measurement is more subjective than the one of laboratory measurement, but this does not mean that it is less rigorous. On the contrary, according to Boumans field measurement simply requires different assessment methods.

On Boumans's account, an objective assessment of field measurement requires constructing a model of the measurement process, which should include all relevant factors suggested by theory, expert judgment and background knowledge, and, then, testing the validity of this encompassing model. Due to the inexactness of theory, this model might still be incomplete, but the validation test will tell whether a significant factor is missing. The validation of these gray-box models should be done by means of 'behavior pattern tests': instead of testing the validity of the model structure, the assessment is based on how accurately the model can reproduce the major patterns exhibited by the target phenomenon.

By emphasizing the role of models in underpinning claims about the epistemic reliability of measurement, Boumans's suggestion contributes to a recent body of literature on model-based accounts of measurement (see Tal 2013, 2015). In particular, the book can be seen as defending the idea that theoretical models of the measurement process are necessary in order to obtain outcomes that can be considered as reliable representations of the properties under measurement. Different kinds of models can serve this purpose, ranging from detailed white-box analytical models to gray-box models based on inexact theory. I believe that model-based accounts of measurement can offer specific advantages in the context of field science, because they allow making claims about measurement reliability also in cases where it is not possible to control the actual system under measurement. However, it is worth noticing that the relevance of model-based accounts of measurement is not limited to field science. Models can play a role in underwriting reliability claims also in the measurement of physical quantities in the laboratory (Tal 2011).

The philosophical discussion of the role of models in measurement practice is still at an early stage and many questions remain open, concerning for instance what kind of model is employed and how, as well as what connections can be made with the wide philosophical literature on scientific modeling. Therefore, I think that Boumans's model-based account could come to full life in the investigation of particular cases of successful measurement practice in the field. Although *Science outside the laboratory* is filled with historically detailed examples, which give an overview of the methodology of scientific practice in the field, it would have been useful if Boumans had provided a specific case study of measurement practice which could help to illuminate some of these issues.

THE NEED FOR EXPERTS AND THEIR ASSESSMENT

Drawing on a number of examples, Boumans argues that reliable measurement in field science always requires expert judgment. Building models, selecting the relevant variables, choosing between alternative models, and assessing them require theory as well as additional information that comes from expert judgment.

By focusing in particular on Haavelmo's works, the book provides a thorough analysis of a methodological debate in econometrics about the possibility of testing the completeness of models. Haavelmo identifies an important problem faced by econometricians when testing the significance of the causal factors to be included in their models. Econometric models are tested against specific bodies of data. If the residual errors are large, then it is inferred that something is wrong with the model. But, if the residuals are small, Haavelmo argues that one can only assume that the model includes the right set of variables, without being able to prove it. According to Haavelmo, in case of small residuals, it is impossible to discern between two alternative cases: either the model is complete, or the data set happens to come from a contingent situation in which certain determinant variables are constant or uninfluential, and hence remain latent. In other words, a variable might appear uninfluential simply because the variation it displays in the data set is too small to show its effective potential influence. As a consequence, it might happen that the test yields small residuals even if certain relevant causal factors are not included in the model.

This discussion sheds light on a problem that is of general interest for field science: because of the impossibility of enforcing an experimental design in the field, we are stuck with Nature's experiments. A field science cannot be an inductive science by relying only on the data coming from contingent situations and statistical methods to analyze it. According to Haavelmo, if Nature's experiments are insufficient to evaluate the potential significance of causal factors, theory may help. But, since econometric theory is inexact, additional sources of information are needed in order to choose the right set of variables. This leaves open the question of where this additional information comes from: although Haavelmo refers to this kind of knowledge, he never specifies what kind of knowledge this is. According to Boumans, this additional source of information is provided by expert judgment.

Since making inferences from observations to phenomena and building accurate models require experts, the reliability of field investigation depends, in part, on the reliability of expert judgment. *Science outside the laboratory* offers a perspective on how to validate expert judgments by drawing on methods developed for laboratory science and adapting them so as to fit the context of field science. To assess the reliability of field investigation, Boumans suggests, the experts should be validated in a similar way as the measurement models.

Boumans's proposal is based on a method of reaching consensus among experts which was developed for applications to engineering problems. According to the Cooke method, rational consensus among experts is based on a weighted average of the individual expert judgments, where the weights are determined by testing the experts' estimations against the known value of reference variables (p. 159). When attempting to apply this method to the social sciences, however, it might prove difficult to individuate suitable variables for the test. In the social sciences, the variables that enjoy wide consensus on their value are commonly well known to the experts, making it pointless to use them as references for the validation of their judgments.

As a solution to this problem, Boumans suggests to opt for a model-based forecasting test. In economics, he argues, there is much more consensus on the validity of specific empirical models than on the values of specific variables. Therefore, he suggests to test experts' judgments against the predictions made by valid models. More precisely, expert judgment should be evaluated by means of behavior pattern test, that is, not only against point predictions, but also against a broader range of predictions about behavioral patterns, like frequencies, trends, phases, etc. (p. 171). Experts who regularly make good predictions of behavioral patterns should be assigned higher weights in the aggregated judgment.

With this methodological suggestion, *Science outside the laboratory* concludes its exploration of the methodology of reliable field science with an optimistic attitude. Reliable investigation of field phenomena is possible, but it requires a context specific methodology of assessment. In particular, assessing the reliability of field measurement requires the validation of both the measurement models and the involved expert judgments. The proposals for field-specific methods of assessment for measurement and expert judgment are, I believe, among the main contributions of the book. Interestingly, in both cases, the recommended assessment method is model-based, which highlights a growing attention towards the multiple roles played by models in field scientific practice and, more generally, in the social sciences.

In conclusion, *Science outside the laboratory* is a recommended reading for philosophers or social scientists interested in scientific practice outside the laboratory, its challenges, its methodology, but also for anyone who is interested in understanding the role of models in underpinning reliability claims in the social sciences.

REFERENCES

Tal, Eran. 2011. How accurate is the standard second? *Philosophy of Science*, 78 (5): 1082-1096.

Tal, Eran. 2013. Old and new problems in philosophy of measurement. *Philosophy Compass*, 8 (12): 1159-1173.

Tal, Eran 2015. Measurement in science. In *The Stanford Encyclopedia of Philosophy* [summer 2015 ed.], ed. Edward N. Zalta. http://plato.stanford.edu/entries/measurement-science/.

Swoyer, Chris. 1987. The metaphysics of measurement. In *Measurement, realism and objectivity*, ed. John Forge. Dordrecht: Reidel, 235-290.

Alessandra Basso is a PhD student at the Academy of Finland Center of Excellence in the Philosophy of the Social Sciences. Her thesis is intended as a contribution to the philosophical investigation of measurement practice, and pays particular attention to measurement in the social and behavioral sciences.

Contact e-mail: <alessandra.basso@helsinki.fi>