Journal of Economics Library
www.kspjournals.org
Volume 3 March 2016 Issue 1

Marcel Boumans, ***Science Outside the Laboratory: Measurement in Field Science and Economics***, **Oxford University Press**, 2015, 216 pp. $59 Hardcover.

By Attilia RUZZENE [†]

Abstract. Marcel Boumans' *Science outside the Laboratory* revolves around the distinction between laboratory and field science, and the challenges that the latter faces in the measurement of scientific phenomena. Boumans raises a methodological puzzle, the possibility of reliable measurement *in the field*, and he gradually resolves it throughout an excursus in the history of science that brings to light episodes of methodological significance. The book starts with Oskar Morgenstern's warning about the peril of scientific observation in economics; touches upon Gaussian's theory of error and its uses in meteorology; discusses Haavelmo's intuition about the problem of passive observation; and concludes with a survey of contemporary methods for aggregating experts' judgments. Ultimately, *Science outside the Laboratory* is a call for expert knowledge as a complementary source of evidence that, if carefully integrated with the traditional tools of field sciences, can eventually lead us to more reliable, and in this sense more objective, measurement. In what follows, I will first outline what I take to be the main theses of the book and discuss some of its main tenets. I will then illustrate some of the crucial steps in Boumans' argument in detail. I will finally conclude with some general comments about the book.
Keywords. Economic Thought, Economic methodology.
JEL. A10, B30, B40.

1. The Lab and the Field

Overall *Science outside the Laboratory* can be understood as a long argument aimed at defending two related theses. First, the ideal of scientific objectivity is worth pursuing not only in its "natural" domain, that is the *laboratory*, but also in the *field,* where it has often been regarded as unachievable. Rehabilitating scientific objectivity as a legitimate ideal in the *field,* however, should not come at the cost of denying the fundamental differences between the two domains. On the contrary, it is only by first acknowledging the differences that we can overcome the challenges that confront the pursuit of objectivity in the *field*. Second, the personal and subjective is part and parcel of scientific inquiry in the *field*; but, far from being an obstacle to that inquiry, it can further the accomplishment of the ideal of objectivity. In line with these theses, chapter after chapter Boumans builds a case for the necessity of the subjective to achieve eventually a form of objectivity. The construction of this argument reveals Boumans' unconventional style and expertise, which combine the interest of the philosopher and the love for details that is typical of the historian. Indeed, while the

† Erasmus University Rotterdam, Netherlands.
☎.
✉. attilia.ruzzene@gmail.com

theses outlined above are never argued for in an explicit and systematic manner, Boumans does hint to the unavoidability of those conclusions through a number of case studies that unambiguously point in that direction.

To understand Boumans' argument and its implications, it might first be useful to spell out some of the assumptions that I believe lurk in the background of the discussion and remain to some extent implicit. As said at the very beginning of this article, the distinction between the *lab* and the *field* is the *fil rouge* throughout the book. Boumans sets himself up to the task of "developing an account of measurement for the field sciences" (p. 24), refers back to this distinction at different stages of his analysis, and interprets and assesses alternative approaches and authors in light of it. Despite its centrality, the character and reach of this distinction remain quite elusive. Boumans' association of concepts, allusions, and examples to the distinction between *lab* and *field* is very liberal, and sometimes approximate. In my view the fundamental idea behind it is that, far from signaling a mere methodological distinction, the *lab* and the *field* are separate domains of inquiry where different types of phenomena are studied. Phenomena in the former domain of investigation, that is the *lab*, are liable to fall under the control of the investigator, and are thereby *replicable*. The scientist can tinker with the phenomenon, interfere with it, modify it, and eventually master it. By difference, phenomena in the *field* resist control and are not amenable to regimentation: scientists must thus surrender to their *unreplicability*.

This admittedly minimalist definition enables us to understand why Boumans associates the *lab* with the natural sciences and the *field* with the social sciences. Even if the match is approximate, social phenomena, that is phenomena involving the *human factor*, undeniably tend to belong to, and make up most of, the latter domain of inquiry. The human factor clearly complicates things further insofar as, unlike Nature, the social world is neither benevolent nor simple: Nature, says Boumans quoting Morgenstern, might hold back information, but she does not lie deliberately (p.8-9). Other interesting properties happen to emerge when phenomena can be regimented in such a way as to be rendered replicable. First, by creating the conditions under which the phenomenon repeats itself, one makes it also amenable to *law-like* descriptions. Furthermore, once empirical regularities start populating our domain of investigation, it is likely to become easier to *theorize* in a principled and systematic way about it. Finally, tinkering and interfering with the phenomenon, that is *intervening* (Hacking, 1983), is a prominent strategy to bring about change, and thus a powerful way to start building up causal knowledge.

In my view, Boumans understands the *field* as the *lieu* where, since phenomena are less than mastered regularities are *quasi*-law-like, measurement is *in*accurate, our theories are *in*complete, and causal knowledge is likely spurious. The *field* falls short of *objectivity* because error is pervasive in the senses defined above. Aiming at objectivity as an ideal thus involves finding strategies to reduce error (that is, making our measurements more accurate), define more precisely the scope of our empirical regularities, and reduce the incompleteness of our theories. In seeing science in the *field* as a departure from exactness, Boumans' view is not very distant from that of Oskar Morgenstern, John Stuart Mill, and Alfred Marshall. The main respects in which he differs from the authors above is probably his belief that knowledge that is subjective, or personal, not only is not a hindrance, but can actually help the pursuit of objectivity as an ideal. Whereas his predecessors might have been somehow open to this idea, they never fully embraced it. This is instead Boumans' solution: "The question is not how to exclude subjective judgment, but rather where do we allow it, how much, and in what sense?"(p.120)

While the first part of the book is a preliminary methodological discussion of measurement in the *field*, the bulk of Boumans' argument is explicated in chapters 4, 5, and 6. Here, Boumans tries to vindicate his account by showing us that: traditional strategies of measurement in the *field* have proven insufficient to *eliminate* error; arguments that dispute the validity of subjective judgments are highly controversial and ultimately *weak*; and there are positive reasons in favor of subjective judgments as a complementary strategy for reducing error in the *field*. As already mentioned, Boumans' argumentative strategy is rather unconventional. It has a strongly historicist flavor while retaining an obviously prescriptive component. The argument hangs on a number of methodological considerations that are illustrated rather than proven. This represents somehow a challenge for the reader who sometimes feels like staring at a moving target. It is however clear that Boumans is not drawing conclusions through systematic derivation from a set of principles; and, moreover, to gauge his analysis according to these standards would be misguided. Rather, Boumans opts for what I would call a *casuistic*. In other words, he constructs cases directed to extract a methodological lesson from a historical episode.

These cases back up the crucial steps in his reasoning. Boumans thus makes a case for the *pervasiveness of error* in the field sciences. He makes a case for the *validity of clinical judgments*. He makes a case for the *relevance of expert knowledge* for achieving an error-free field science. In what follows, I will briefly review some of these cases to give the reader a sense of what to expect when engaging with Boumans' work.

2. The Problem of Passive Observation

Boumans discusses the *problem of passive observation* in the context of an exchange between Jan Tinbergen and Trygve Haavelmo in the late-Thirties and early-Forties of the past century. The exchange revolves around the problem of how to establish the causal significance of a variable by means of linear regression models. In particular, the problem is whether, and to what extent, the notion of influence (the variable coefficient) and strength of influence (the coefficient time the standard deviation) are valid constructs for causal significance. In his work for the League of Nations, Tinbergen concluded that the rate of interest had at best a negligible explanatory power on investments based on the value of these constructs being not significantly different from zero. Haavelmo, however, regarded this conclusion as erroneous: he pointed out how this fact could be simply a consequence of the limited variation of the variable in question (the rate of interest) in the time interval represented in the data. Haavelmo then generalized this problem by introducing a distinction between *factual* and *potential influence* of a causal variable.

We can understand the *factual influence* of a variable as its explanatory power with respect to the outcome of interest in a given data set. *Potential influence* is the variation we would observe in the outcome had the variable changed in a controlled environment. What we are ultimately interested in is the potential influence, which we take as capturing the *real* causal structure underlying the data. However, we can only know about the potential influence of a variable through its factual influence, as the former cannot be directly observed whereas the latter actually can. Haavelmo convincingly demonstrated how imperfect a guide the latter is to the former. Mere absence of variation in the current data set might in fact lead us to the erroneous conclusion that the variable in question has at best a negligible causal influence on the outcome because its factual influence would turn up as close to zero.

We can now see how salient the *problem of passive observation* is for measurement in the *field* as it incarnates, and thus exemplifies, the limits of *any* method germane to it. This case makes the point that there is a fundamental difference between how data are generated in the *lab*, the lieu of so to speak "active observation", and how they are obtained in the *field*, the lieu of "passive observation"; and this fact has crucial consequences for what these data are evidence for. Moreover, insofar as we are interested in the underlying causal structure *and* we cannot manipulate it directly, passive observation (in all its forms) turns out to be a fallible strategy because it cannot possibly provide direct information about that causal structure. In other words, the evidence that more securely would take us closer to our epistemic goal, that is evidence of counterfactual dependence, is evidence we can only (or best) fabricate in the lab. What we harvest in the field is at best its imperfect surrogate.

Thus, it seems that Boumans does make a point. How strong a point is it, though? Even though the severity of the *problem of passive observation* is uncontroversial, it certainly loses some of its dramatic impact once we inscribe it within the more general issue of under-determination. Roughly, theory is under-determined by the data whenever the empirical evidence at our disposal is compatible with a multiplicity of hypotheses and thus insufficient to discriminate among those. In this case, the factual influence of a variable being close to zero in the data is compatible with at least two rival hypotheses. The causal (potential) influence is also negligible *and* the causal influence is significant though obscured in the data for some reasons. Under-determination of theory by data is a long known, and discussed, phenomenon. While it certainly is a pervasive phenomenon in the *field*, the *lab* is not immune from it. Actually, it is exactly in the latter context that the physicist Pierre Duhem first discussed it.[2] Thus understood, the *problem of passive observation* becomes a less compelling case because, once reformulated in slightly more abstract terms, it challenges the distinction it was initially meant to underpin.

3. On the (Ir)Rationality of Human Judgments

In chapter 5, Boumans introduces the notion of *judgment* and distinguishes between *rational judgments* and *considered judgments*. The distinction becomes meaningful once we consider that to find solutions to real life problems agents model the situations they face. Rational judgments are assessments of the probability of a certain occurrence formulated against the background of a probabilistic model. We thus speak of a judgment as *rational* vis-à-vis a given model: provided the model is correct the judgment it licenses is rational if valid. Considered judgments (aka clinical, human, or Kantian judgments) express the adequacy of a given model to represent the problem faced in a certain situation. They thus rely on a broader set of considerations that, among other facts, take into account the specifics of the case. This distinction enables Boumans to raise the following points. Each situation can be conceptualized as a different problem by using different models. Hence, rational judgments are correct or mistaken only within a given model; in other words, different models elicit different rational judgments. Furthermore, it might be the case that there is no unique, or uniquely justified, solution to the problem of what is the right model for a given situation.

Boumans illustrates these points by way of rather detailed case studies. The "hot hand" case revisits the debate about whether *hot hands* among basketball players do exist or are merely a cognitive illusion. It shows how people model differently

[2] How threatening the under-determination problem really is, is actually matter for discussion (see Laudan, 2012)

the same simple real-life situation and thereby achieve different rational judgments. The Monty Hall case tells us how the popular problem of "the three doors" has kept the discussion alive for decades without leading to an agreed-upon solution. Hence, besides showing as the former case does that there are different ways of modeling the same real-life problem, it also points out that there is no obvious solution to what the best model for a given situation is. Unlike the Monty Hall case, the case of medical judgments shows that even though different rational judgments of the same situation are simultaneously possible, sometimes a model *is* best suited than others to a given real-life problem. Whether this is the case depends on how *considered* the judgments formulated in its support are. In this way Boumans convincingly shows that *practitioners* themselves can be in the position of formulating considered judgments reliably.

Let me spend a few words on this interesting case. The "Harvard Medical School Test" has been used to conclude that physicians often suffer of what is known in the literature as the base-rate fallacy. The test proves that when assessing the probability of a given outcome, for example the chance that a person found to have a positive test result actually has the disease, physicians erroneously overlook the base rate or pre-test probability. Their judgments thus count as *irrational* if assessed according to the Bayesian model for revising beliefs in light of new evidence. However, a more detailed analysis of the practice shows that physicians' responses are compatible with the use of heuristics that are tailored to the situations the practitioner typically faces when formulating this type of decision. In particular, these heuristics are developed on the basis of complex models of decision making (e.g. the threshold model by Pauker and Kassirer), which take into consideration crucial pieces of information such as the risk and benefits of administering medical tests like biopsy. Treating these judgments as irrational because they are not in line with Bayesian reasoning is thus misconceived since the relevant model behind the heuristics in use is a different one.

Boumans seems to draw the additional conclusion that "the kind of subjective knowledge that is needed to complement objective knowledge is knowledge that is not crystallized in models and is *personal*. It is knowledge that is part of Karl Popper's "World 2": *the mental or psychological world, the world of our feelings of pain and pleasure, of our thoughts, of our decisions, of our perceptions and our observations; in other words, the world of our mental or psychological states or processes or of subjective experiences* (p.147). I found these assertions rather puzzling. In my view what the case above shows is that the "subjectivity" of practical reasoning —for example, in medical decision-making—is anything but personal (in the sense above). It is instead practice-related knowledge embodied in heuristics, the effectiveness and rationality of which is testified by the model from which the heuristics originated. The heuristics thus codify background knowledge that needs not be entirely appropriated by the individual agent. This fact, however, does not make it part of our mental or psychological states. It does require, though, that it be validated in different modes, and invites us to inquire about the model behind the heuristics and the process through which the latter originated from the former.

4. Conclusion Remarks

Science outside the Laboratory can be seen as a solution to another problem that had a lasting influence on the philosophical debate about field science in the past decades. That is the primordial tension between theory and observation. Boumans restates the old learned lesson that the field sciences have theoretical weaknesses and the empirical obstacles they face are to some extent insurmountable. Boumans seems decisively skeptical of the theoretical alternative. Despite the plea of early econometricians, it is not at all clear where knowledge of fundamental causal structures can eventually be found. On the other hand, he seems to suggest that the solution to the problem cannot be empirical (read statistical) either. The data we collect in the *field* depend on the whims of Nature; moreover, there is no reason to expect that Nature's experiments would meet our epistemic need, if not by sheer chance. His case of *passive observation* is, after all, devoted to illuminate exactly this point. The impasse Boumans describes seems as real and threatening as ever.

In this debate, Boumans wisely leans towards some form of epistemic modesty and mild empiricism. If theories are incomplete and data insufficient to retrieve the nomological machine responsible for their creation, maybe we should revise our epistemic ambitions. Maybe the data we have should be used to establish local regularities, that is, regularities that hold quite robustly for only a limited set of circumstances. And we should forget altogether about nomological machines. This idea is saliently captured by the notion of *grey boxes* (p. 50-52). *Grey boxes* are an intermediate between *black* and *white boxes*. A *white box* is a set of causal-descriptive statements on how a real system actually operates. *Grey boxes* are modular designed models where the modules are *black boxes.* Unlike *white boxes*, *grey boxes* do not require direct test of the causal structure. They instead test the causal structure indirectly through stress test, extreme-conditions test, or test the model capacity to replicate the data through behavior tests.

Boumans' message resonates well with several other positions in philosophy and in the sciences. It is enough to mention notions such as *mechanism* in the former camp (Elster, 1998) or guiding principles such as the *Marschak's Maxim* in the latter (Heckman & Vytlacil, 2007; Heckman & Urzua, 2010). Both have experienced increasing popularity in the last decades and both embody, in my view, some form of epistemic modesty and mild empiricism. This fact certainly provides additional plausibility to Boumans' take on the issue. However, it also adds to the feeling of surprise when one comes to glimpse the avenue he eventually pursues. As Boumans aptly acknowledges, Kevin Hoover makes a plea for a strikingly similar position when asserting that the goal of econometrics is to discover facts generated by unobservable nomological machines, but that do not presuppose knowledge of those machines (Hoover, 2002). If we look at the contemporary debate in econometrics, what we see is not only the harsh debate between so-called structuralists and experimentalists. We also see attempts at "building bridges" between the two sides that arguably respond to the Marshak's Maxim mentioned above (Heckman, 2010).

For some unexplained reason, however, this option is not contemplated by Boumans. His solution is searched elsewhere in the partly unexplored resources of expert knowledge. Thus we can say that Boumans' solution to the old dilemma between theory and observation consists in the advocacy of some form of methodological pluralism. We find the current methods wanting and we supplement their shortcomings by integrating the evidence they provide with evidence from other sources. This is certainly a plausible strategy and it would be interesting to see how it would play out more concretely. However, this is not the only solution and not necessarily the best. The examples above show that the dilemma can be overcome so to speak "from within". Each discipline might have its own, maybe yet unexplored, resources to move beyond the sharp opposition between theory-driven and data-driven practice, as Heckman's case illustrates. If we followed this route we might also end up advocating some form of pluralism, but it would be pluralism with a different face. For example, it might be based on the intuition that results are not commensurable across methods but should be instead taken as shedding lights on different facets of a complex reality.

Overall *Science outside the Laboratory* presents an overarching narrative that I find rather convincing and in a way original. The book provokes contradictory feelings in the reader, though. It is extremely rich in details that sometimes provide unexpected insights and other times blur the narrative. Sometimes the reader finds himself asking for more clarity and precision; yet, he may be captivated by the vividness of the narration. It is a challenging book that deserves credit for the salience of its theses and the originality of its style. It also deserves credit for another message that it, maybe unintendedly but forcefully, advocates. Boumans' craftsmanship brings to the fore the inherent complexity of the scientific practice against any easy stereotype that some philosophers, and sometimes scientists themselves, might be inclined to build. In particular, it shows that the rationality of scientific *judgment*, which isn't always immediately apparent to the external observer (and by this, I mean any observer that is an outsider to the particular scientific community), can be found in the specific blend of experience, codified practice, and "local" judgments that characterizes the work of the scientific practitioner. Boumans' contribution consists in showing at a painstaking level of detail where the solidity of these judgments rest. In this sense, we can say that his call for expert knowledge is thereby vindicated inch by inch.

References

Elster, J. (1998). A plea for mechanism, in Hedström, P. and R. Swedberg, eds. *Social Mechanisms. An Analytical Approach in Social Theory*. Cambridge University Press: Cambridge.

Hacking, I. (1983). *Representing and Intervening*. Cambridge University Press: Cambridge.

Heckman, J.J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. NBER Working Paper Series, No. 16110. doi. 10.3386/w16110

Heckman, J.J., & Vytlacil, E. (2007). Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation, in Heckman, J.J. and E. E. Leamer, eds. *Handbook of Econometrics*. Elsevier: Amsterdam.

Heckman, J.J., & Urzua, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics* 156(1), 27-37. doi. 10.1016/j.jeconom.2009.09.006

Hoover, K.D. (2002). Econometrics and reality. in Mäki, U., eds. *Fact and Fiction in Economics*. Cambridge University Press: Cambridge.

Laudan, L. (2012). Demistifying undetermination. in Curd, M., Cover, J.A. and C. Pincock, eds. *Philosophy of Science: the Central Issues*. Norton: New York.