STUDIA METODOLOGICZNE
NR 36 • 2016, 27–41
DOI: 10.14746/sm.2016.36.2

MARCEL J. BOUMANS

Models of Clutter

ABSTRACT. In his seminal 'Models of Data,' Patrick Suppes [1962] proposes a 'hierarchy of models' to define a correspondence between abstract theories and the complex activities of conducting experiment and measurement. Although he nicely distinguishes a lowest level of 'ceteris paribus conditions', that is, a level of 'noises, lighting, odors, phases of the moon,' he does not provide a model for this level, and therefore is not able to connect this level to the upper levels. The level of ceteris paribus conditions aims at reducing clutter: to mute loud noises, to fresh the air from bad "odors", or to re-organize the schedule for observations. These attempts to reduce clutter, that is, these cleaning activities are often the most time-consuming activities in scientific practice and require a lot of creativity and intuition. Because philosophy of science is, in my view, philosophy of science in practice, these activities deserve more attention. This article, therefore, proposes an attempt to complete Suppes's hierarchy of models by suggesting a methodology for designing and testing 'models of clutter' that account for the level of ceteris paribus conditions.

KEY WORDS: ceteris paribus conditions, clutter, hierarchy of models, model of data, philosophy of science in practice

1. Introduction

Because philosophy of science should indeed be about science, that is, the practice of science, a few likeminded philosophers founded in 2006 an organization, the Society for the Philosophy of Science in Practice, that aims at supporting this kind of philosophy. Philosophy of science-in-practice is philosophy that analyses science in the making, that is, the daily practice of scientific research and everything that such practice entails (e.g. processes of inquiry, institutional settings and social dynamics among investigators).[1] Within this approach, philosophers use empirical methods drawn from the historical or social sciences (such as archival research, ethnographies or interviews) to acquire insights into and evidence of scientists' research behaviour [see Boumans, Leonelli, 2013].

In studying practices of empirical research, particularly those of social science, I found and still find it striking how little theory plays a role in these practices. This is in sharp contrast to the presumptions of traditional philosophy of science that privileges theory. Many, if not most, of the practices of empirical research are very much *un*related to theory. I could even – perhaps better – say they are *un*connected, as I will explain below.

An explanation that these empirical research practices are theory-poor is provided by James Woodward's [1989] 'Data and Phenomena.' This article shows that the procedures that enable inferences from claims about data to claims about phenomena are actually procedures that aim at the reduction of errors, that is, the cleaning of data; and moreover it shows that theories are hardly of any help in designing these cleaning procedures. I prefer to call these procedures to reduce errors and noise cleaning activities because of the ambiguity of what the errors and noise are; in this sense they are very similar to clutter. In Woodward's article the prominent cleaning activity is filtering, but his account can also applied to other activities of cleaning, like polishing and bleaching.

According to Woodward, theories explain or predict (only) the relatively stable and general features of the world, called phenomena, and are not concerned with local and idiosyncratic conditions. Data, unlike phenomena, are local. The differences between data and phenomena can be characterized in different ways, as was shown in 'Saving the Phenomena' [Bogen and Woodward, 1988], but the leading characterization of this difference in 'Data and Phenomena' is "in terms of the notions of error applicable to each" [p. 394]:

[1] For reasons of completeness, I should mention that philosophy of science can also be seen as philosophy-of-science in practice, which is philosophy directly engaged with scientific research through interaction with scientists about philosophical problems or/and collaborations on joint questions [Kosolosky, 2012]. Although I see this latter approach as equally valuable, it will not be discussed in this article.

> The problem of detecting a phenomenon is the problem of detecting a signal in this sea of noise, of identifying a relatively stable and invariant pattern of some simplicity and generality with recurrent features – a pattern which is not just an artifact of the particular detection techniques we employ or the local environment in which we operate. Problems of experimental design, of controlling for bias or error, of selecting appropriate techniques for measurement and of data analysis are, in effect, problems of tuning, of learning how to separate signal and noise in a reliable way. [Woodward, 1989, pp. 396–397]

Woodward arrives at this characterization of the difference between data and phenomena by having looked very closely at the kind of activities that take place in practice:

> Underlying the distinction between data and phenomenon is the idea that the sophisticated investigator does not proceed by attempting to explain his data, which typically will reflect the presence of a great deal of noise. Rather, the sophisticated investigator first subjects his data to a great deal of analysis and processing, or alters his experimental design or detection technique, all in an effort to separate out the phenomenon of interest from extraneous background factors. [Woodward, 1989, p. 397]

Note that Woodward is referring to a "sophisticated investigator" and not to a sophisticated theory that would be instrumental in separating signal from noise. Practice is about the reduction of noise and this asks for expertise rather than theory.

Woodward characterizes practice by the efforts to remove clutter in order to arrive at the true facts about the phenomenon or object of investigation. The core idea of "separating signal and noise" implies a metaphysics that is very similar to the view supposed to be expressed by Michelangelo: "Every block of stone has a statue inside it and it is the task of the sculptor to discover it." This idea is for two reasons problematic: First, it assumes that eventually a sharp line can be drawn between signal and clutter. Second, as any sculptor would admit, chipping marble, which is a very hard stone, is not just a matter of elimination, but requires a lot of craftsmanship.

Craftsmanship and expertise is acquired by a lot of training and education. And due to the works of Michael Polanyi [1958; 1966], it is now generally acknowledged that an important part of expertise is tacit and personal knowledge – so knowledge that is not and cannot be made explicit in our theories. Which brings us to the following dilemma: How to investigate philosophically – so not sociologically, psychologically, or ethnographically – a research practice where knowledge to a large extent is tacit and intuitive? My proposal is that this can be done by studying the documents one will find at the sites of practice. These documents can include a variety of printed materials: almanacs, dictionaries, guides, handbooks, instructions, reports, teaching materials, tutorials, yearbooks and anything else one can find on desks and shelves at the research site or in the corners to where they have been thrown away out of frustration or because they became redundant. For a philosophy of science-in-practice, these documents have proven to be very informative sources to gain a deeper understanding of specific research practices, particularly of those where theories do play a minor role.

A very useful framework to study the practice of empirical research in more detail is Patrick Suppes's [1962] hierarchy of models. Suppes's account of a hierarchy of models was introduced for the first time in his 1960 article on the meaning and uses of models. His reason for introducing this idea of a hierarchy of models was the "radical" difference between the "logical type" of models used in theory and those used in experiment: "The maddeningly diverse and complex experience which constitutes an experiment is not the entity which is directly compared with a model of a theory" [p. 297]. In philosophy of science, to make a comparison between theory and experiment possible, usually "drastic assumptions of all sorts are made in reducing the experimental experience […] to a simple entity ready for comparison" [p. 297]. A plurality of models between these two levels could reduce the need for these drastic assumptions.

A more detailed discussion of the hierarchy of models appeared in his 'Models of Data' [1962]. He argued that this paper was written to overcome the "sins of philosophers of science […] to overly simplify the structure of science" [p. 260] by representing scientific theories as logical calculi and then to “go on to say that a theory is given empirical meaning by providing interpretations or coordinating definitions for some of the primitive or defined terms of the calculus” [p. 260]. Instead of this overly simplistic view of how theories are related to data, Suppes argued that “a whole hierarchy of models stands between the model of the basic theory and the complete experimental experience” [p. 260], see Table 1. A model at one level is given empirical meaning by a specifically defined connection with the model at a lower level.

Table 1. Hierarchy of theories, models, and problems. Source: Suppes 1962, p. 259

Theory of	Typical Problems
Linear response models	Estimation of Θ, goodness of fit to models of data
Models of experiment	Number of trials, choice of experimental parameters
Models of data	Homogeneity, stationarity, fit of experimental parameters
Experimental design	Left-right randomization, assignment of subjects
Ceteris paribus conditions	Noises, lighting, odors, phases of the moon

At the lowest level, Suppes placed “*ceteris paribus* conditions:” “Here is placed every intuitive consideration of experimental design that involves no formal statistics. Control of loud noises, bad odors, wrong times of day or season go here” [p. 258]. Although Suppes distinguishes explicitly a lowest level of real practice – long before the “practice turn” became more fashionable in philosophy of science – that is, the level where one has to deal with “noises, lighting, odors, phases of the moon” – he does not provide a model for this level, and therefore is not able to connect this level to the upper levels. A model is not provided because he assumed it is not feasible due to “the seemingly endless number of unstated *ceteris paribus* conditions” [p. 259]. In other words, this lowest layer of dealing with the ceteris paribus conditions cannot be covered by any model because of the infinite number of conditions one has to account for. As a result, it cannot be connected to the level of experimental design above it.

At the level of ceteris paribus conditions one aims at reducing clutter: to mute loud noises, to refresh the air from bad "odors", or to re-organize the schedule for observations. These attempts to reduce clutter, that is, these cleaning activities are often the most time-consuming activities in scientific practice and require a lot of creativity and intuition. This article proposes an attempt to complete Suppes's hierarchy of models by connecting the level of ceteris paribus conditions to the levels above by suggesting a model of clutter that accounts for this level of practice. Therefore, I will discuss the guides of metrology, in particular the *Guide to the Expression of Uncertainty in Measurement* [JCGM 100 2008].[2] This *Guide* does not provide any theory of uncertainty nor measurement, but is written to guide practitioners in dealing with clutter. As will be shown, this *Guide* proposes implicitly a model of clutter.

2. Guide to the expression of uncertainty in measurement

The "seemingly endless number of unstated *ceteris paribus* conditions" is not only a problem of experimental practices discussed by Suppes, but also for measurement practices. For this reason the *Guide to the Expression of Uncertainty in Measurement* [JCGM 100, 2008] was developed to deal with this kind of problems.[3] The *Guide* proposes to provide a measurement

---

[2] Metrology is the shared view on measurement of eight international metrological organizations: the International Bureau of Weights and Measures (BIPM), the International Electrotechnical Commission (IEC), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Laboratory Accreditation Cooperation (ILAC), the International Organization for Standardization (ISO), the International Union of Pure and Applied Chemistry (IUPAC), the International Union of Pure and Applied Physics (IUPAP), and the International Organization of Legal Metrology (OIML). Their shared view can be found in the publications of the Joint Committee for Guides in Metrology (JCGM).

[3] In 1977, "recognizing the lack of international consensus on the expression of uncertainty in measurement" [p. vi], the Comité International des Poids et Mesures (CIPM), requested the Bureau International des Poids et Mesures (BIPM) to address the problem in conjunction with the national standards laboratories and to make a recommendation. The BIPM installed in 1979 a Working Group on the Statement of Uncertainties that recommended to develop a detailed guide, which became the *Guide to the Expression of Uncer-*

result in terms of a description consisting of an estimate of the value of the "measurand" and a statement of the uncertainty of that estimate. The expression of the uncertainty of the result of a measurement is supposed to reflect the lack of exact knowledge of the value of the "measurand":

> The first step in making a measurement is to specify the measurand – the quantity to be measured; the measurand cannot be specified by a value but only by a description of a quantity. However, in principle, a measurand cannot be completely described without an infinite amount of information. Thus, to the extent that it leaves room for interpretation, incomplete definition of the measurand introduces into the uncertainty of the result of a measurement a component of uncertainty that may or may not be significant relative to the accuracy required of the measurement. [JCGM 100, 2008, p. 49]

A description in terms of "errors" would be misleading about the epistemological results obtained in a measurement practice, because it implies knowledge of the endless number of unstated conditions. Therefore, error is considered to be an idealized concept.

The *Guide* emphasizes that uncertainty means "doubt about the validity of the result of a measurement" [p. 2]. To evaluate this doubt, uncertainty is formally defined as a "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand" [p. 2]. But uncertainty of measurement comprises many components. Some of these components may be evaluated from the statistical distribution of the results of a series of measurements and can be characterized by experimental standard deviations. The other components, which also can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information. As a result, uncertainty can be evaluated in two different ways, called Type A evaluation and Type B evaluation:

> Type A evaluation (of uncertainty): "method of evaluation of uncertainty by the statistical analysis of series of observations" [p. 3].

---

tainties in Measurement published in 1980. The 2008 *Guide* used in this paper is the 1980 *Guide* "with minor corrections."

Type B evaluation (of uncertainty): "method of evaluation of uncertainty by means other than the statistical analysis of series of observations" [p. 3].

The dispersion is expressed as variance. The estimated variance characterizing an uncertainty component obtained from a Type A evaluation is calculated from a series of repeated observations and hence is the statistically estimated variance. For an uncertainty component obtained from a Type B evaluation, the estimated variance is evaluated using other "available knowledge" [pp. 6–7].

As the *Guide* explains [p. 64], if a measurement laboratory has limitless time and resources, it can conduct an exhaustive statistical investigation of every conceivable cause of uncertainty, for example, by using many different makes and kinds of instruments, different methods of measurement, different applications of the method, and different approximations in its theoretical models of measurement. The uncertainties associated with all of these causes can then be evaluated by the statistical analysis of these series of observations and the uncertainty of each cause will be characterized by a statistically evaluated standard deviation. In other words, all of the uncertainty components will be obtained from Type A evaluations. Since such an investigation is not feasible in practice, many uncertainty components must be evaluated by using a mathematical model of the measurement and "the law of propagation of uncertainty" ([p. 7].

The model of the measurement is a functional relationship f between the measurand Y and its influencing quantities, called input quantities X_1, X_2, …, X_N:

$$Y = f(X_1, X_2, \ldots, X_N)$$

As the *Guide* [p. 9] notes, if data indicate that f does not model the measurand to the degree imposed by the required accuracy of the measurement result, additional input quantities must be included in f to eliminate the inadequacy. This may require introducing an input quantity to reflect the incomplete knowledge of a phenomenon that affects the measurand.

For an estimate of an input quantity X_i, denoted by x_i, that has not been obtained from repeated observations, the associated estimated variance is evaluated by “scientific judgment” based on all of the available information on the possible variability of X_i. The pool of information may include

– previous measurement data;

– experience with or general knowledge of the behaviour and properties of relevant materials and instruments;

– manufacturer’s specifications;

– data provided in calibration and other certificates;

– uncertainties assigned to reference data taken from handbooks. [p. 11]

> The proper use of the pool of available information for a Type B evaluation of standard uncertainty calls for insight based on experience and general knowledge, and is a skill that can be learned with practice. [JCGM 100, 2008, p. 12]

The law of propagation of uncertainties combines the uncertainties of the input quantities. In the most simple case, that is, the case of uncorrelated influences, the law is presented as follows:

$$u_c^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

Although the *Guide* only discusses a model of measurement, this account can nevertheless easily be generalized to other empirical research practices, like experimentation. It shows how to model the level of ceteris paribus conditions, namely, by expressing an “endless number of unstated ceteris paribus conditions” in terms of uncertainties. Because the mathematical model may be assumed to be incomplete with respect to all possible influences, the evaluation of uncertainty is a necessary part of the description to account for that part on which one is ignorant.

From the *Guide* one could therefore arrive at a revised hierarchy of models that includes a model of the level of ceteris paribus conditions, see Table 2.

Table 2. Hierarchy of models

Model of measurement or experiment	Specification of f and input quantities X_i
Models of data	Statistical distributions and subjective probability distributions
Ceteris paribus conditions	Uncertainty components obtained from Type A and Type B evaluations, law of propagation of uncertainties

3. The use of the senses

The *Guide* not only emphasizes that for a Type B evaluation experience and scientific judgement is needed, but also Type A evaluations "require the application of some judgement" [p. 61]:

> The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value. [JCGM 100, 2008, p. 8]

The Guide, however, does not discuss this judgement further, and leaves it black-boxed as an "art of measurement." In a similar vein, Suppes referred to "intuitive consideration" that, according to him, needs no further exploration.[4]

The problem of judgement is that it is personal and subjective, which easily leads to the suspicion that it is biased. Moreover, looking at current debates about expert judgement, particularly in medicine, "subjective" and "biased" are often considered to be synonyms. To clean personal judgements from bias, procedures have been developed to make these judgements less subjective. These procedures attempt to address the senses in a more direct way, so as to avoid any interference by the mind as much as

[4] The term "art" is often used to black-box those aspects of research where intuition and personal judgement play a role, with the implicit suggestion that is not part of *science*, and hence does not need further exploration.

possible. The mind is considered to contain various kinds of prejudices that bias our "views." It is assumed that the senses are more unbiased than the mind. Although this is supposed to apply for all senses, that is, sight, hearing, smell, taste, and touch, I will only discuss here vision.

The procedures that address vision are actually procedures of visualization, that is, procedures that make clutter visible, such that vision can be used in judging to remove the clutter. In practices where one can intervene physically, this is often done by colouring the object of study in such a way that the clutter becomes visible because it gets a different colour than the object to which it is attached. For example, erythrosine is a colouring agent that discloses dental plaque by colouring the plaque red.

In social science, one cannot always intervene with the phenomenon under investigation. As a consequence the visualization has to be done in another way, namely, by first visually displaying the observations before the observational errors can be erased.[5] As an example of such a cleaning procedure that is based on visualization and vision, I will discuss briefly the method of graduation in actuarial science.

Graduation is defined as "the process of securing from an irregular series of observed values of a continuous variable a smooth regular series of values consistent in a general way with the observed series of values" [Miller, 1946, p. 4].[6] Graduation is based on the view that there is an underlying law that produces a smooth, regular and continuous sequence of values, but that all kinds of disturbances have turned this sequence into an irregular one. The irregularities represents deviations from the true values, and thus the revised, graduated, sequence should be taken as a representation of the underlying law. However, the only knowledge about these laws are the observations:

---

[5] In Boumans 2016, I discuss extensively the Method of Graphs as a method of visualization and vision to correct for errors in statistics.

[6] Almost since its foundation the Actuarial Society of America has a tradition of promoting textbook publications on graduation. Miller's *Elements of Graduation* served as main education reference for a number of years following 1950 till 1985. Boumans 2015 discusses gradation more in detail as one of the "calculi of observations."

> Since no law of mortality, in the sense of a physical law, is known to us, nor is one likely to be discovered, we have no way of knowing *a priori* what the basic pattern of mortality is. We must therefore rely on the information supplied by observations of the rates of mortality actually being experienced. [Miller, 1946, pp. 1–2]

But graduation is not only characterized by smoothness. The other "essential quality," according to Miller [1946, p. 5] is "fit, or consistency with the observed data." These two different qualities are, however, "basically inconsistent." Improving one is at the cost of the other. Therefore, "any graduated series must of necessity follow a middle course between optimum fit and optimum smoothness; it must represent the result of a compromise between the two" [p. 5]. There exists, however, no standard for this "compromise," and therefore it must be left to the judgment of the graduator: "a graduation method must allow the graduator some latitude in choosing the relative emphasis to place on smoothness and fit in the graduated series" [p. 5]. But for the analysis of the observations there are too many possible methods of graduation available to choose from; the choice of the most appropriate graduation method is undermined by the observations. As Miller (1946) emphasized, "graduation does not have a single solution" [p. 7]. It depends upon the choice of the method, upon a choice on how much fit and how much smoothness there should be, the field of application, but also "upon the skill and experience of the graduator" [p. 7].

In actuarial science fit is defined as $F = \sum_i (\hat{x}_i - y_i)^2$ and smoothness as $S = \sum_i \left(\Delta^z \hat{x}_i\right)^2$, where the y_is constitute the time series and the $\hat{x}_i$s the graduated values of them. The evaluation of the F is a Type A evaluation because it is merely based on a statistical analysis. But smoothness is a Type B evaluation. This evaluation is based on judgements using visualization and vision. Smoothness is a feature of a visual display. Moreover, whatever method of graduation one designs, one can only evaluate this method for its aimed level of smoothness by looking at a graph that shows the graduated series. How well a specific graduation formula smooths a certain time series can only be evaluated by graphical display.

4. Conclusions

Last year, when Krzysztof Nowak-Posadzy approached me to inquire whether I would be interested in contributing to a special issue on economic methodology, I became motivated to contribute when he suggested that I could write an article about the directions economic methodology according to me should pursue. In my view a lot of work in economic methodology still follows in the tracks of the more traditional philosophy of science with its main focus on theory. This article is written to invite economic methodologists to turn their interest to empirical practices, particularly to those practices where theories do not play so much of a role. This also implies a suggestion not to study those experimental or testing practices where the investigations are aiming at evaluating theories, because then theories would still remain at the core of the philosophical investigations.

Having studied these theory-weak practices, I found that expert judgements are essential. The consequence of this finding is that studying these research practices implies that expert judgements should also be investigated. There is of course a rich literature in psychology, cognitive science, and philosophy on judgements, but actually there is not much in philosophy of science, let alone in economic methodology.[7]

To study expert judgements from an economic methodology perspective means that those practices should be studied where these personal judgements are most explicitly and visibly needed. In my view these are the practices where the most personal of all epistemological sources are needed, that is, those practices where the senses are employed to make epistemological judgements.

[7] There are of course exceptions to this rule, take for example Martini 2014, or his other publications on experts and expertise.

In this paper I have focused on vision and the practice of visualization, with no other reason than that vision is the sense that is most often employed in research. For a similar reason, Annamaria Carusi [2012] has advocated that more attention should be paid in philosophy of science to the epistemological role of visualizations: "it is necessary to understand how vision works embedded in epistemic contexts, as playing a crucial role in the formation of evidence for claims" [p. 107]. A similar position is expressed in Nicola Mößner [2015].

References

Bogen J., Woodward J., (1988), "Saving the phenomenal", *The Philosophical Review* 97 (3), pp. 303–352.

Boumans M., (2015), *Science Outside the Laboratory. Measurement in Field Science and Economics,* New York, Oxford University Press.

Boumans M., (2016), "Graph-based inductive reasoning", *Studies in History and Philosophy of Science* 59, pp. 1–10.

Boumans M., Leonelli S., (2013), "Introduction: on the philosophy of science in practice", *Journal for General Philosophy of Science*, 44 (2), pp. 259–261.

Carusi A., (2012), "Making the visual visible in philosophy of science" *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 6 (1), pp. 106––114.

Joint Committee for Guides in Metrology (JCGM) 100, (2008), *Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement*. JCGM.

Joint Committee for Guides in Metrology (JCGM) 104, (2009), *Evaluation of Measurement Data: An Introduction to the "Guide to the Expression of Uncertainty in Measurement" and Related Documents*. JCGM.

Joint Committee for Guides in Metrology (JCGM), (2012), *International Vocabulary of Metrology: basic and general Concepts and Associated Terms*, 3rd edition, JCGM.

Kosolosky L., (2012), "Philosophy-of-science in practice vs. philosophy of science-in-practice", *Newsletter SPSP*, Winter, pp. 9–10.

Martini C., (2014), "The role of experts in the methodology of economics", *Journal of Economic Methodology* 21 (1), pp. 77–91.

Miller M.D., (1946), *Elements of Graduation*, The Actuarial Society of America and American Institute of Actuaries.

Mößner N., (2015), "Visual information and scientific understanding", *Axiomathes*, 25, pp. 167–179.

Polanyi M., 1958, *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago, University of Chicago Press.

Polanyi M., (1966), *The Tacit Dimension,* London, Routledge.
Suppes P., (1960), "Comparison of the meaning and uses of models in mathematics and the empirical sciences", *Synthese* 12 (2/3), Proceedings of the Colloquium: The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, pp. 287–301.
Suppes P., 1962, „Models of data", [in:] *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, eds. E. Nagel, P. Suppes, A. Tarski, Amsterdam, Stanford University Press, pp. 252–261.
Woodward J., (1989), "Data and phenomena", *Synthese* 79, pp. 393–472.

Marcel J. Boumans
Faculty of Law, Economics and Governance,
Utrecht University School of Economics (U.S.E.),
Kriekenpitplein 21–22,
3584 EC Utrecht, Netherlands,
Phone: +31(0)302536287,
e-mail: m.j.boumans@uu.nl