How Scientists Are Brought Back into Science—The Error of Empiricism

Mieke Boon

Introduction

With the rise of A.I., expert-systems, machine-learning technology and big data analytics, we may start to wonder whether *humans as creative, critical, cognitive and intellectual beings* will become redundant for the generation and application of knowledge. And additionally, will the increasing success of machine-learning technology in finding patterns in data make *scientific knowledge* in the form of theories, models, laws, concepts, (descriptions of) mechanism and (descriptions of) phenomena superfluous?[1] Or can it be argued that human scientists and human-made scientific theories etc. remain relevant, even if machines were able to find data-models that adequately but *incomprehensibly* relate or structure data—for example, *data-models* that provide *empirically adequate* mapping relationships between data-input and data-output or determine *statistically sound* structures in data-sets.[2]

[1] In this chapter, 'theory' is taken in a broad sense, encompassing different kinds of scientific knowledge such as concepts, laws, models, etc. The more general term 'scientific knowledge' encompasses different kinds of specific *epistemic entities* such as theories, models, laws, concepts, (descriptions of) phenomena and mechanisms, etc., each of which can be used in performing different kinds of *epistemic tasks* (e.g., prediction, explanation, calculation, hypothesizing, etc.).

[2] On the terminology used in this chapter. In the semantic view of theories, *patterns in data* are also called *data-models* (see section "Empiricist epistemologies"), which are *mathematical representations of empirical data sets* (e.g., Suppe 1974; McAllister 2007). This chapter will adopt the term *data-model* in this very sense. In machine learning textbooks, data-models are also referred to as *mathematical functions*. Abu-Mostafa et al. (2012), for instance, speaks of the *unknown target function* f: X –> Y, where X is the input space (set of all possible inputs x), and Y is the output

M. Boon (✉)
Department of Philosophy, University of Twente, Enschede, The Netherlands
e-mail: m.boon@utwente.nl

M. Bertolaso, F. Sterpetti (eds.), *A Critical Reflection on Automated Science*, Human Perspectives in Health Sciences and Technology 1,
https://doi.org/10.1007/978-3-030-25001-0_4

Anderson (2008) indeed claims that the traditional scientific method as well as human-made scientific theories will become obsolete:

> This is a world where *massive amounts of data and applied mathematics replace every other tool* that might be brought to bear. *Out with every theory* of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. *With enough data, the numbers speak for themselves.* […]. The big target here isn't *advertising*, though. It's *science*. The *scientific method* is built around testable hypotheses. These *models*, for the most part, are *systems visualized in the minds of scientists*. The models are then *tested*, and experiments *confirm* or *falsify* theoretical models of how the world works. This is *the way science has worked for hundreds of years*. […]. *Scientists are trained* to recognize that *correlation is not causation*, that no conclusions should be drawn simply on the basis of correlation between X and Y (it could just be a coincidence). Instead, you *must understand the underlying mechanisms* that connect the two. Once you have a model, you can connect the data sets with confidence. *Data without a model is just noise*. […]. But faced with massive data, this approach to science — *hypothesize, model, test — is becoming obsolete*. (Anderson 2008, my emphasizes).

Essentially, Anderson suggests that the meticulous work done by scientific researchers aiming at scientific concepts, laws, models, and theories on the basis of empirical data, will become superfluous because learning machines are able to generate data-models that represent relationships and structures in the data. Each set of data will be fitted by a unique data-model, which implies that we can give up on generalization and unification endeavors. Intermediate scientific concepts, laws, models, and theories, which are desired by humans for obvious metaphysical beliefs, and which are also practically needed to deal with the limitations of their intellect, can be bypassed if relating, structuring and simplifying data—which basically is what science does according to Anderson's quote—can be done by machines.

If let's say, scientists such as Boyle, Gay-Lussac, and Hooke, had fed their experimental data to a machine (e.g., data consisting of the measured pressure, volume

---

space (e.g., y_1 is 'yes' for x_1; y_2 is 'no' for x_2; etc.). The *machine learning algorithm* aims to find a *mathematical function g* that 'best' fits the data, and that supposedly approximates the unknown target function *f*. Abu-Mostafa et al. call the function *g* generated by machine learning 'the final hypothesis.' Alpaydin's (2010), on the other hand, uses the notion of model and function interchangeably. An example (Alpaydin 2010, 9) is predicting the price of a car based on historical data (e.g., past transaction). Let *X* denote the car *attributes* (i.e., properties considered relevant to the price of a car) and *Y* be the price of the car (i.e., the outcome of a transaction). Surveying the past transactions, we can collect a *training data set* and the *machine learning program fits a function* to this data *to learn Y as a function of X*. An example is when the fitted function is of the form $y = w_1.x + w_0$. In this example, the *data-model* is a *linear equation* and w_1 and w_0 are the *parameters* (weight factors) of which the values are determined by the machine learning algorithm to best fit the training data. Alpaydin (2010, 35) calls this equation 'a single input linear model.' Hence, in this example, the *machine learning algorithm* to fit the training data includes only one property to predict the price of a car. Notably, the machine learning program involves a learning algorithm, *chosen by human programmers*, that confines the space in which a data-model can be found – in this example, the *learning algorithm* assumes the linear equation, while the *data-model* consists of the linear equation together with the fitted values of the parameters (w_0 and w_1).

and temperature in a closed vessel, or the weights and extensions of different springs, respectively), the machine would have generated a data-model to connect these data, which could then be used to make predictions at new physical conditions. The Boyle/Charles/Gay-Lussac laws for gasses and Hooke's law for elasticity would not have existed. Taking this a step further, Anderson's claim implies that scientific concepts such as 'the ideal gas law', 'the gas-constant' (R), and 'the elasticity coefficient' (k) would be unnecessary. We would not even need related scientific concepts, such as 'gas-molecules,' 'the number of Avogadro,' 'collisions of molecules,' and 'reversible processes.'[3]

This short expose aims to raise the question whether a future is conceivable in which nobody needs to understand science any longer—a future in which the production and uses of scientific concepts, laws, models, mechanisms, theories etc. can be replaced by machine learning algorithms that produce *epistemically opaque* data-models[4] and networks stored in machines that will do all kinds of epistemic tasks for us—which would imply indeed that humans no longer need to learn theories etc. nor how to apply scientific knowledge in solving problems. Conversely, are there reasons to believe that scientific researchers still have a role to play?

The structure of this article is as follows. Section "Machine-learning" presents a brief overview of machine-learning technologies and applications. The different kinds of ways in which computers and machine-learning technologies may replace human experts and scientists are discussed. A list of epistemic tasks is drawn up, about which it can be reasonably assumed that machine learning will outperform humans.

In section "Empiricist epistemologies", I aim to make plausible that the abilities of computers and machine-learning technologies largely correspond with ideas in the empiricist tradition about the character of knowledge and ways of (deductive or inductive) reasoning on the basis of knowledge—and vice versa, about how general knowledge can be derived (inductively and statistically) from observations and data.

I will revisit accounts of empiricism at the beginnings of the philosophy of science, including (neo)positivism, because authors such as Mach and Duhem have articulated the basic assumptions of empiricism in a clear and straightforward manner. My aim is to first explain why epistemological and normative accounts of

[3] Current machine learning practices show that machine learning algorithms are dependent in varying degrees on our theoretical and practical background knowledge. Therefore, another option regarding Anderson's assumptions is that the current state of knowledge *suffices* for this purpose. Yet, in the context of this article, it will be assumed that he means to say that machine learning technology will eventually develop to the extent that such knowledge will become superfluous in the construction of machine learning algorithms.

[4] The notion of *epistemic opaqueness* of a process has been introduced by Humphreys (2009, 618): "a process is epistemically opaque relative to a cognitive agent X at time t just in case X does not know at t all of the epistemically relevant elements of the process. A process is essentially epistemically opaque to X if and only if it is impossible, given the nature of X, for X to know all of the epistemically relevant elements of the process."

science developed in the (neo-)positivist and empiricist tradition, make it very hard to articulate our intuitive discomfort about the suggestion that machines could take over and make human scientists virtually superfluous. I aim to make plausible that on an empiricist epistemology the elimination of any human contribution to scientific knowledge is in fact already built in as a normative ideal. Attempts to ensure the superiority of science seem to assume that the objectivity of epistemic results and of methods testing these results consists of some kind of algorithmic reasoning, be it deductive or (statistical) inductive. If this is so, it should not come as a surprise that we are forced to believe that data-models produced by machine learning algorithms are just better.

Three well-known ideas developed in the empiricist tradition will be discussed to show that a strict empiricist epistemologies indeed support the claim that objective, although opaque, data-models produced in machine learning processes can replace and may even be preferable to human-made scientific knowledge: (1) *Hempel's model of scientific explanation*, which supports the idea that the supposed laws and correlations operating in D-N and I-S explanation schema's can be interpreted as data-models constructed to represent input-output relationships in larger sets of observed or measured data; (2) The rejection of a *distinction between data and phenomena*, which supports the idea that (descriptions of) phenomena can be reduced to statistically sound data-models generated in machine learning processes; and (3) The *semantic view of theories*, which supports the idea that scientific knowledge in the form of theories or models does not add much to empirically adequate and/or statistically sound data-models to represent data.

Hence, several ideas central to empiricist epistemologies supports the belief that ultimately scientific knowledge is no longer needed, and show that the empiricist tradition offers hardly any possibilities for a more positive appreciation of the epistemic and cognitive roles of human scientists.

In the last section (section "Knowledge in the age of machine-learning technologies"), it will be argued that empiricist epistemologies are flawed, or at least too limited to understand the crucial role of scientific knowledge (theories, models, etc.) and human scientist in epistemic practices such as the engineering and biomedical sciences. It will be argued that a better understanding of knowledge in the age of machine-learning technologies requires widening our philosophical scope in order to include epistemological issues of *using* knowledge for all kinds of practical purposes. To that aim, philosophical accounts of science must start from the side of *epistemic tasks* and *uses* (e.g., Boon 2017c) and address questions such as, *how* science produces knowledge that can be *used*, and *how is it possible* that knowledge can be used anyway—for instance, in discoveries, technological innovations, 'real-world' problem-solving, and in creating and controlling functionally relevant phenomena by means of technology (e.g., Boon 2017a). Finally, on the basis of this analysis, many roles of scientists and of comprehensive scientific knowledge can be pointed out, which is how the human is brought back in science.

m.boon@utwente.nl

Machine-Learning

Machine-Learning Technologies

Machine-learning algorithms are increasingly used in dealing with complex phenomena or systems, aiming to detect, predict or intervene with the complex physical phenomena or systems in developing technological application such as in biomedical and healthcare contexts. Examples are machine-learning applications in the prediction and prognosis of chronic diseases (Kourou et al. 2015; Dai et al. 2015); drug discovery (Lima et al. 2016); brain imaging (Lemm et al. 2011); and genetics and genomics (Libbrecht and Noble 2015).

Other well-known machine-learning applications aim at automated pattern recognition in ways that replace human experts. For instance, recognizing *visual images*, which require the eye of an expert but not in-depth theoretical knowledge. Examples are automatic face recognition (Odone et al. 2009; Olszewska 2016); automated visual classification of cancer (Esteva et al. 2017); vision technologies for biological classification (Tcheng et al. 2016); and, forensics (Mena 2011). Another application of machine-learning concerns pattern-recognition in the sense of discovering *patterns*, *correlations* and *causal relationships* in (big) data sets, especially in the social sciences. Originally, these kinds of data-sets were analyzed by means statistical programs such as SPSS. Examples of machine-learning technologies drawing on finding patterns and structures in order to make proper predictions about specific cases situations, are: financial risk management (van Liebergen 2017); fraud detection (Phua et al. 2010); and manufacturing (Wuest et al. 2016). In these kinds of applications, machine-learning technologies develop towards more advanced strategies of finding patterns in data, e.g., by coupling data from different sources, and strategies such as network-based stratification to detect correlations or even causal structures (e.g., Hofree et al. 2013) that would be impossible through more traditional statistical programs.

Notably, machine learning is different from computer simulations, which utilize scientific knowledge to build mathematical models (e.g., sets of differential equations) that can be run on a computer—scientists use these simulation models, for instance to view dynamic processes and to investigate how changes in parameter values affects these processes. The machine-learning process does not draw on scientific models that are constructed by means of theories, laws, mechanisms and so forth. No theory or mechanism or law needs to be fed to the machine-learning process. Instead, the learning problem of the machine is to find a data-model that presents a correct mapping relationship between input and output data of a training-set (Alpaydin 2010; Abu-Mostafa et al. 2012). For example, in ML systems concerning face recognition, the relevant task is *classification* in which the inputs, which are the images of human faces, are classified into the individuals to be recognized, which are the outputs.

What Machines Can Do

Given the currently known examples, computers and machine-learning technologies can do different types of things for different uses, thereby taking over intellectual capabilities and types of reasoning that were previously carried out by experts and scientific researchers. Here, I propose a provisional categorization of *epistemic tasks* that can be performed by both humans and machine learning technologies, with the aim of making clearer how capacities of computers relate to those of humans:

(a) 'Match': Machine-learning technologies have the ability to learn to 'match' a visual images or data-strings (the input data) with a specific image or data-string somewhere sitting in a large data-set (e.g., automated face-recognition, finger-print recognition, matching of DNA profiles). Accordingly, ML technology is able to somehow mimic the human ability to recognize relevant similarities between visual images, or structural similarities in graphical pictures. It is often still possible to check (e.g., by and expert) whether the technology performs at least as good as the expert, but the ML-technology will outperform humans in speed. If images or data-strings get more complex, machines may perform more reliable or at a higher statistical precision (i.e., pointing out how reliable the outcome is).

(b) 'Interpret': Machine-learning technologies have the ability to learn to 'interpret' visual images as belonging to a specific type, in accordance with categories defined by humans. Accordingly, ML technology is able to take over the human ability to recognize or interpret the image as of a specific type of object, to belong to a specific category, or to subsume it under a specific concept (e.g., "that is an oak," "that is a car of brand Z," "that is *Picea mariana* rather than *Picea glauca*"). In these applications experts may have played a role in supervising the machine-learning process (e.g. Tcheng et al. 2016). Here as well, it is often still possible to check (by an expert) whether the technology performs at least as good as the expert, but the ML-technology will outperform humans in speed.

(c) 'Diagnose': Similarly, machine-learning technologies have the ability to learn to 'diagnose' data-strings as probably belonging to a specific class within pre-set categories, which may be generated by humans, but also by means of machines. Hence, ML technology is able to infer from limited information about a specific target that "it probably belongs to a specific category and therefore will probably also have several additional properties" (e.g., as in personalized advertisement of buyers; financial risk assessment of customers; and, in medical diagnosis of patients).

(d) 'Structure': Machine-learning technologies have the ability to learn to find patterns, correlations and causal relations in data, which is a task originally done by humans or by statistical programs. When data-sets get more complex (which can also be considered as 'richer'), the relationships will become more complex (which can also be considered as 'more refined'), which may then be accepted as empirically adequate but opaque structures in data. These structures, in turn, can be utilized in machines learning to ‘match,’ ‘interpret,’ or ‘diagnose.’

(e) ‘Discover’: Additionally, structures found in data by ML technologies may point out, or point at (physical or social) phenomena, very similar to how human researchers infer from observed occurrences, causal relationships or measured regularities to (physical or social) phenomena. Yet, it will require human researchers to draw the relationship between computer outcomes and the real world, because the pattern does not speak for itself.

(f) ‘Calculate’: Machine-learning technologies are enabled by computers (the machine). Automated calculation was the first example of computers outperforming humans in accuracy and speed. Humans can check the calculations, and assume that the algorithm by which the computer performs the calculation somehow maps the algorithm as we understand it (e.g., adding up instead of multiplying).

(g) ‘Simulate’: Similarly, computer programs running complex simulations of dynamic processes outperform humans in accuracy and speed, as well as in handling complexity. Here, as has been briefly explained above, the adequacy of the computer program is firstly checked by how the scientific model (on the basis of which the computer-program was build) was constructed.

(h) ‘Integrate’: The performance of machine-learning technology will multiply if the mentioned abilities are combined. Natural language translation is an example, but also biomedical applications, for instance, as expressed in expectations regarding *personalized* and *precision* medicine.

This overview shows that, while computers already performed better than humans with regard to *deductive* reasoning in calculation and simulation—which basically consists of performing repetitive tasks guided by logical rules— they now also start to get better than humans in *recognizing* patterns and structure in data or pictures, for matching, interpreting and diagnosing purposes. Additionally, machine learning technology may contribute to the *discovery* of new theoretical concepts or categories, but in this case, the crucial role of humans is still to recognize the discovered structure (pattern, correlation or causal relationship) as a *representation* of something that is traceable or existent in reality, i.e., as a (physical or social) phenomenon.

One of the major *applications* of ML technology is their uses in making correct *predictions*. Computers were already widely used in their ability of *deductive inference*, thereby making deductively correct predictions—i.e., the prediction is logically correct, but may be empirically inadequate due to errors in the underlying models or the computational procedures. Machine-learning technology adds predictions that are based on *inductive inference*, which means that the algorithms (i.e., the correct mapping relationship in a learning set) is applied in new situations to predict statistically expected outcomes.

This vast range of machine-learning applications may suggest that scientific researchers and scientific knowledge become superfluous as learning from large data-sets, algorithms and data-models will be developed at a degree of complexity and adequacy far beyond the capacity of the human intellect. Yet, in section "Knowledge in the age of machine-learning technologies", it will be argued that scientists and scientific researchers still play a crucial role.

Empiricist Epistemologies

Basic Assumptions of Empiricism

The first claim of this paper is that, *if* we accept some of the fundamental presuppositions of empiricism, it becomes very difficult to argue against the idea that machines will ultimately perform better than human scientists. Presuppositions central to empiricist strands can be divided into two kinds, one normative and one epistemological. The normative ideal is driven by the desire to prevent superstition and abuse of power through knowledge, by requiring knowledge to be verifiable in principle, and is one of the reasons why objectivity plays a central role in science. Linked to this is also the explicit aim of avoiding metaphysical claims in science. This then requires an epistemology that explains how objectivity can be achieved while avoiding metaphysical content. Yet, empiricist epistemologies are not necessarily normatively motivated, but can also be determined by purely epistemological convictions. In order to substantiate my first claim, I will first outline the basic assumptions of empiricism by reference to Mach's Positivism and Logical positivism.

Central to Mach's positivism is the idea that the subject matter of scientific theories is phenomenal regularities.[5] Theories characterize these regularities in terms of *theoretical terms*, which need to be grounded in observation. Accordingly, theoretical terms in our theories and laws have to be explicitly defined in terms of phenomena, and are nothing other than abbreviations for such phenomenal descriptions. Additionally, Mach maintained that one must reject any *a priori* (or metaphysical) elements (such as causality) in the constitution of knowledge of things.

Logical positivism agreed that the subject matter of scientific theories is phenomenal regularities and that theories characterize these regularities in terms of theoretical terms being conventions used to refer to phenomena, and indeed added to positivism that a scientific theory is to be axiomatized in mathematical logic that specifies the relationships holding between theoretical terms.

The preliminary point I aim to make based on this brief overview, is that if these basic assumptions of a strict empiricism were correct, the theoretical terms (also called scientific concepts), mathematical relationships between them (also called scientific laws) and theories (also called axiomatic systems) generated in science by the meticulous efforts of scientists, are in fact quite arbitrary intellectual instruments to fit the data, which, in principle, can be replaced by the data-models

[5] Frederick Suppe (1974, Chapter One) presents a comprehensive outline on the historical background to the so-called Received View, which develops from positivism to logical positivism (e.g., Carnap) and logical empiricism (e.g., Hempel).

generated and executed in machine-learning technologies. Additionally, since machines can handle much bigger data-sets, and because machines are not confined by the kinds of idealizations and simplifications humans need to make in order to fit data into comprehensive mathematical formalisms, we may expect that machines will handle the inherent irregularity and complexity of data-sets more effectively than the human intellect ever could.

Scientific Explanation

Also Duhem's philosophy of science stands in the tradition of positivism and conventionalism of the late 19th and early 20th century. In accordance with the basic assumption of this tradition, Duhem denies that theories of physics present (causal) explanations. Instead, an explanation is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as simply, as completely, and as accurately as possible a set of experimental laws. Experimental laws on this view, are simplified or idealized general descriptions of experimentally produced observable effects. Concerning the very nature of things, or the realities hidden under the phenomena described by experimental laws, a *theory tells us absolutely nothing*. On the contrary, from a purely logical point of view, there will always be a *multiplicity of different physical theories equally capable of representing a given set of experimental laws* (Duhem [1914] 1954; Craig 1998).[6]

Hempel's (1962, 1966) two models of explanation agree to the basic assumptions of empiricism as well. Although Hempel emphasizes that one of the primary objectives of the natural sciences is to *explain* the phenomena of the physical world, he defends that formal accounts of explanation should avoid the metaphysical notion of causality. Similar to Duhem, Hempel claims that: the explanation fits the phenomenon to be explained into a pattern of uniformities and shows that its occurrence was to be expected, given the specified laws and the pertinent particular circumstances. Explanations, therefore, may be conceived as deductive arguments whose conclusion is the explanandum sentence, E, and whose premise-set, the explanans, consists of general laws, L_1, L_2, ..., L_r, and of other statements C_1, C_2, ..., C_k, which make assertions about particular facts. Hempel calls explanatory accounts of this kind, explanations by deductive subsumption under general laws, or deductive-nomological (DN) explanations. The second model involves explanation of phenomena by reference to general laws that have a probabilistic-statistical form. In this case, the explanans does not logically imply the explanation, but involves inductive subsumption under statistical laws, called inductive-statistical (IS) explanation. In this case, the statistical laws make it only likely that the phenomenon was to be expected.

[6]A clarifying phrase "*to save the phenomena*" to capture the empiricist idea of how knowledge is obtained from data was originally introduced by Duhem (2015/1909) and later adopted by, among others, Van Fraassen (1977, 1980) and Bogen and Woodward (1988).

Similar to Duhem, Hempel's notion of explanation entails that an explanation only tells that, based on our empirical knowledge of the world so far, the phenomenon was to be expected—the phenomenon 'fits to', or 'can be subsumed under,' the regularities, patterns and correlations that have been found in observations and experimentally produced data.

Again, if, as empiricist epistemologies suggest, this is what science ultimately has to offer—that indeed, theories, models, laws and scientific concepts can be traced back to data, and are just helpful instruments that do not add anything to our knowledge about the world—then, it is to be expected that eventually machines will outperform human scientists. For, as especially Duhem's position suggests, there is no good reason to belief that the regularities, patterns and correlations in data found by humans would be better than the empirically adequate but opaque data-models found by a machines—and additionally, if empiricists are correct, there is no reason to doubt that machines will be capable to accurately fit a particular phenomenon into data-models stored in machines such as to predict that given a certain input a specific output is to be expected (with a specified probability).

There is a large literature on explanation that argues against Hempel's account of explanation, claiming that, although Hempel's theory may succeed in avoiding the (metaphysical) concept of causality, it is insufficient to account for the proper meaning of explanation. Well-known counter-examples, which meet Hempel's criteria of DN or IS explanations but are considered improper explanations, are: the barometer explaining the storm (which illustrates the problem of *common cause*); the length of the shadow of the flagpole explaining the length of the flagpole (which illustrates the problem of *symmetry*); and, taking the birth-control pill explaining why male do not get pregnant (which illustrates the problem of *explanatory relevance*). Conversely, counter examples that do not meet Hempel's criteria, but are considered proper explanations have been given, such as: the mayor's untreated syphilis explains why he got paresis (which illustrates the problem of *low probabilities*).

The briefly listed arguments against Hempel's logical empiricist account of explanation concern the meaning of explanation, assessed by what is commonly (and rather intuitively) taken as proper and improper (scientific) explanations. The listed arguments boil down to the idea that an explanation ought to be an answer to a *why* question, and therefore should refer to a relevant (physical) *cause*. But because reference to hidden causes is based on empirically untestable and thus metaphysical convictions, this is indeed what (strict) empiricism aims to avoid.

In the context of this article, the issue is *whether the opaque data-model generated by machine-learning technologies count as explanations* for the relationships found between input and output. As has been argued above, Duhem rejects (causal) explanations entirely, and may therefore agree that the possibility of empirically grounded algorithms produced by machines from which new conclusions can be derived, proves this even better. So, his point entails that we need no explanations anyway. Yet, contrary to Duhem, many of us will hold that we need explanations, and that an opaque data-model together with specific conditions producing an outcome—which basically is the logical or mathematical structure of an explanations on Hempel's account—is not a proper explanation for that outcome. But then the issue is, what 'being a (scientific) explanation' actually adds, and conversely, what is it that apparently is not provided by the opaque data-model. Does our resistance to the idea that an explanation in terms of an opaque data-model is not any better than an explanation in terms of theories and laws merely rely on deep 'scientific realist' intuitions, according to which—paraphrasing Van Fraassen (1980)—an explanation gives us a literally true story of what the world 'behind' the observable phenomena is like? Differently phrased, on a scientific realist view, opaque data-models do not provide explanations because genuine explanations *describe* or *represent* the unobservable (physical) causes (or mechanisms, processes, phenomena, or structures otherwise) that bring about the observed (physical) phenomena. In the last section, I will return to this issue, namely whether it is merely our metaphysical disposition, or whether genuine explanations are more than data-models that fit the data.

Data and Phenomena

The issue whether we eventually will need human-made explanatory laws and theories, rather than opaque data-models that merely fit the data, is at the heart of the question about *explanation* discussed in the previous section. Here, it will be laid out that the presuppositions of strict empiricism also challenge the distinction between data and phenomena as proposed by Bogen and Woodward (1988), because strict empiricism agrees to the idea that phenomena are nothing more than statistically justified mathematical structures in data.

Bogen and Woodward (1988) contest that there is a direct relationship between theories and data as assumed in strict empiricism. Instead, according to B&W, the notion of phenomena is crucial for understanding the relationship between data and theories. Therefore, different from the empiricist tradition, in particular Van Fraassen (1977) who builds on Duhem, a conceptual distinction is needed between data and phenomena. Loosely speaking, scientists infer to phenomena based on data, because data are idiosyncratic to particular experimental contexts and typically cannot occur outside them, whereas phenomena are objective, stable features of the world. Phenomena, therefore can occur outside of the experimental context, and are detectable by means of a variety of different procedures, which may yield quite different kinds of data, whereas data reflect the influence of many other causal factors, including factors that have nothing to do with the phenomenon of interest and instead are due to the measurement apparatus and experimental design (B&W 1988; Woodward 2011).

B&W's (1988) position, including some of the clarifications by Woodward (2011) and Bogen (2011), can be summarized as follows: (1) Phenomena are distinct from data, where data is what is directly observed or produced by measurement and experiment; (2) Often phenomena are unobservable, or at least, not observable in a straightforward manner; (3) B&W think of phenomena as being in the world, not just the way we talk about or conceptualize the natural order—i.e., phenomena exist independently of us, but beyond that B&W are ontologically noncommittal; (4) B&W don't want phenomena to be some kind of low level theories; (5) Phenomena are inferred from data; (6) Data produced by measurement and experiment serve as evidence for the existence or features of phenomena; and, (7) Theories aim at providing explanations of phenomena, whereas it is difficult to provide explanations of data from theory (even in conjunction with theories of instruments, non-trivial auxiliaries, etc.).

Bogen and Woodward's (1988) notion of phenomena has been criticized by several authors. McAllister (1997, 2011) assumes that B&W describe phenomena both as investigator-independent constituents of the world, and as corresponding to patterns in data-sets. He criticizes this view by arguing that there are always infinitely many patterns in any data-set, and so the choice of one as being a phenomenon is subjectively stipulated by the investigator, which make phenomena investigator-dependent. Also Glymour's (2002) criticizes on the point that B&W leave open the question of how scientists discern or discover phenomena in the first place. Are phenomena merely summaries of data? Or is there something more to phenomena than just patterns or statistical features. Glymour argues there is not. According to him scientists infer from data to patterns by means of statistical analysis, which does not add anything new to the data. This implies that phenomena coincide with patterns in data, and that no subjective grounds are involved. Accordingly, Glymour concludes that Bogen and Woodward are mistaken in thinking that a distinction between phenomena and data is necessary, while McAllister (1997) is mistaken in thinking that the choice about 'which patterns to recognize as phenomena' can only be made by the investigator on subjective grounds.[7]

Within a machine-learning context, we may start to wonder what B&W actually have in mind when distinguishing between data and phenomena. They take Nagel's example of the melting point of lead to explain this:

> Despite what Nagel's remarks seem to suggest, one does not determine the melting point of lead by observing the result of a single thermometer reading. To determine the melting point one must make a series of measurements. […]. Note first that Nagel appears to think that the sentence 'lead melts at 327 degrees C' reports what is observed. But what we observe are the various particular thermometer readings—the scatter of individual data-points. […]. So while the *true melting point is certainly inferred or estimated from observed data*, on the basis of a theory of *statistical inference and various other assumptions*, the sentence 'lead melts at 327.5 + 0.1 degrees C'—the form that a report of an experimental determination of the melting point of lead might take—*does not literally describe what is perceived or observed*. (Bogen and Woodward 1988, 308–309, my italics).

[7] McAllister (2007) presents an in-depth technical discussion of how to find patterns in data (i.e., data-models). He argues that "the assumption that an empirical data set provides evidence for just one phenomenon is mistaken. It frequently occurs that data sets provide evidence for multiple phenomena, in the form of multiple patterns that are exhibited in the data with differing noise levels" (Ibidem, 886). McAllister's (2007, 885) also critically investigates how researchers in various disciplines, including philosophy of science, have proposed quantitative techniques for determining which data-model is the best, where 'the best' is usually interpreted as 'the closest to the truth,' 'the most likely to be true,' or 'the best-supported by the data.' According to McAllister, these "[data-]model selection techniques play an influential role not only in research practice, but also in philosophical thinking about science. *They seem to promise a way of interpreting empirical data that does not rely on judgment or subjectivity*" (Ibidem, 885, my emphasis), which he disputes.

In this example 'the true value of the temperature at which lead melts' is considered to be the *phenomenon*, which, according to B&W is determined by statistical analysis of a set of data taken in measurements. Based on this example, one may be inclined to conclude that Glymour (2002) is correct in claiming that phenomena do not add anything to data.

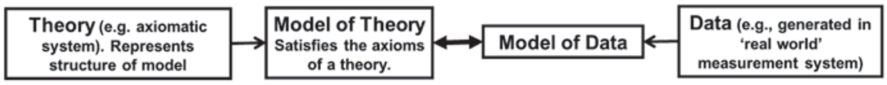
In discussing this issue a bit further, I will use the notion 'physical phenomena' rather than just 'phenomena' to stress that phenomena in the sense of B&W are considered independently existing (physical) things (objects, properties or processes). Additionally, I will use the notion 'conceptions of phenomena,' to account for the fact, rightly pointed out by B&W, that phenomena are usually not observable in a straightforward manner, but need to be discovered and established. Hence, the notion of phenomena is connected to the notion of scientific concept, because a *scientific concept* can be considered *a conception of a physical phenomenon*, which, once the *meaning of the concept* is sufficiently established, becomes a *definition* in a dictionary or textbook. This definition gets the character of a (literal) *description* of the phenomenon (Boon 2012a).

The pressing question is whether *the formation of concepts of phenomena* (including establishing their definitions) will be still required once machine-learning technologies are able to find statistically justified patterns in data in the sense suggested by Glymour. More generally phrased, will *data-models* generated by statistical analysis of data make all other scientific knowledge superfluous, and will machine learning technology be able to generate these data-models?[8]

The Semantic View of Theories

Acceptance of scientific knowledge in empiricist epistemologies involve two important rules: knowledge must be *objective*, and it must be *testable*. Ideally, therefore, knowledge and the way in which it is *tested* must be *independent of specificities of human cognition*, and the measured data used for testing it must be independent of the knowledge to be tested. The so-called semantic view of theories, which in one or another version is held by authors such Suppes (1960), Van Fraassen (1980), Giere (1988, 2010), Suppe (1989), and Da Costa and French (2003), accounts for

[8] Affirmative answers to these questions can be taken as an interpretation of Anderson's position. Notably, even machine learning scientists and textbooks promote that knowledge of any sort related to the application (e.g., knowledge of concepts, of relevant and irrelevant aspects, and of more abstract rules such as symmetries and invariances) must be incorporated into the learning network structure whenever possible (Alpaydin 2010, 261). Abu-Mostafa (1995) calls this knowledge *hints*, which are the properties of the target function that are known to us independently of the training examples – i.e., hints are auxiliary information that can be used to guide the machine's learning process. The use of hints is tantamount to combining rules and data in the learning network structure – hints are needed, according to Abu Mostafa, to pre-structure data-sets because without them it is more difficult to train the machine. In *image recognition*, for instance, there are *invariance hints*: the identity of an object does not change when it is rotated, translated, or scaled.



Schema 1 Semantical relationships between a theory and measured data according to the semantic view. *Theory acceptance* when (partial) isomorphism between model of theory and model of data

testing theories. It aims to account for the structure of theories and the *independent* relationship between theories and measurements, that is, between the outcomes predicted by the theory, and the outcomes of a measurement, by reducing the relationships between abstract theories, models and measured data to semantic relationships between abstract logical-, mathematical-structures and data-structures (see Schema 1).

Loosely speaking, the semantic view posits that a theory is a (usually deductively closed) set of sentences in a formal language, such as an abstract calculus, an axiomatic system, or a set of general laws (such as Newton's equations of motion), which enables to deduce logical consequences about particular types of physical systems (such as the model of a pendulum). The resulting model is a structure which is an *interpretation* (or *realization*) of the theory. Conversely, the theory *represents* the structure of the model. On this view, testing the *adequacy* of a theory only requires isomorphism (or similarity) between the model of the theory for a particular kind of system, and the measurement results called a model of the data. In brief, the semantic view explains how a theory is compared with measurements.[9] On Van Fraassen's (1980) version, *testing* whether a theory is *empirically adequate* means to assess (partial) isomorphism of a (mathematical) structures predicted by the theory (the models of the theory) and the structure in a set of measured data (the models of the data).

Obviously, the focus of the empiricist epistemology expressed in the semantic view is on the theory and how to test it. The question is not, for instance, whether the data-model is adequate. Conversely, in machine learning, the focus is on the data-model and how to test whether it is adequate. Therefore, from a machine learning perspective, someone may now ask 'why bother about the theory?' If machine learning technology can generate adequate data-models based on data, we do not need the theory any longer. Assume that a machine-learning technology has produced a data-model (although opaque and incomprehensive) that fits the data (see right part of Schema 1), and assume that the model of the theory is (partially) isomorphic with the data-model (see middle part of Schema 1), why would we need the left part of this schema anyway? Since, in empiricist epistemologies, the data and the data-model are taken as the solid ground of knowledge, the theory seems to be an unnecessary surplus. Hence, the *semantic view of theories* supports the idea that

[9] Notably, 'phenomena' in the sense of Bogen and Woodward (1988) do not occur in this view. Rather than *phenomena*, as B&W claim, the *model of data* mediates between the *measured data* and the *model of the theory*, which is a specific instantiation (interpretation, concretization) of the *theory* (see Schema 1).

m.boon@utwente.nl

scientific knowledge in the form of theories or models does not add much to empirically adequate and/or statistically sound data-models to represent data. Accordingly, it supports the belief that ultimately scientific knowledge is no longer needed. Again, the empiricist tradition offers hardly any possibilities for a more positive appreciation of scientific theories and the epistemic and cognitive roles of human scientists.

Knowledge in the Age of Machine-Learning Technologies

Empiricist Epistemologies: Theories Add Absolutely Nothing to Data-Models

In the previous section, it has been defended that a consequence of presuppositions and requirements of (anti-realist) empiricist epistemologies is that *explanations*, *phenomena*, and *theories* generated in science can (in principle, although maybe not yet in practice) be represented by, reduced to, or replaced with *data-models* generated by machine learning technologies. Empiricist epistemologies require that data-models adequately fit the data, but there are no specific epistemological reasons to require that data-models are intelligible for humans—that is, the fact that data-models generated by machines usually are opaque and incomprehensive for humans is not a problem in regard of the epistemic value of data-models. Additionally, referring to Duhem, and in his footsteps Van Fraassen, *theories tell us absolutely nothing about hidden realities*—rather, different theories may be equally capable of representing a given set of experimental laws. When taking *experimental laws*, in Duhem's wordings, to be *data-models*, this implies that no additional epistemic value is gained by theories over data-models, especially when data-models accurately represent large data-sets achieved by machine learning technologies. Hence, taking the semantic view of theories as a proper advancement of Duhem's ideas implies that the *epistemic* value of theories is to adequately represent *data-models*, where 'represent' means 'structural similarity,' i.e. being (partially) isomorphic. In turn, data-models *represent* the measured data. If we assume that representational relationships in science are transitive, this implies that from an epistemological point of view empirically adequate theories do not add anything to empirically adequate data-models[10]—as empirically adequate data-models already allow for adequate predictions of 'real-world' data, theories and models become unnecessary (see Schema 1). As a consequence, the claim that machine learning technologies will render human scientists and scientific knowledge superfluous accords with beliefs about the epistemic value of theories in anti-realist empiricist epistemologies.

[10]This claim only holds for anti-realist interpretations (as in Duhem and Van Fraassen) of the semantic view. Yet, the semantic view of theories also allows for realist interpretations of theories (e.g., Suppe 1989).

Empiricist epistemologies, therefore, support Anderson's (2008) claims cited in the introduction. Having this said, several arguments can be put forward against this conclusion.

Scientific Realism in Defense of Science

Anderson's (2008) claims can be more easily countered from scientific realist than from anti-realist viewpoints. Scientific realist positions are supported, at least in part, by the no-miracle argument: the successes of scientific theories would be a miracle unless we assume that theories truthfully describe or represent hidden realities behind the phenomena, which is why scientific realism is the best explanation for the successes of science. As a consequence, data-models, whether produced by human scientists or by machines, are epistemologically inferior to theories. Accordingly, contrary to the conclusion inferred from anti-realist empiricist epistemologies, scientific realists will believe that the successfulness of (approximately) true scientific theories cannot be superseded by data-models.

Additionally, scientific realists may argue that scientific theories have an *intrinsic* value, which has nothing to do with their epistemic or practical usefulness anyway. Many theories are not useful, at least not to begin with. One may even defend that the aim of science is not *useful* theories, but *true* theories. Science may be of epistemic and practical value to all kinds of applications such as in engineering and medicine, but this is a by-product of science, not its intended aim (also see Boon 2011, 2017c). Rather, science has an intrinsic *cultural* value in telling us what the world is like, which is a task that cannot be replaced by machine learning technologies whatsoever since incomprehensive, opaque data-models do not tell us anything meaningful about the world. Therefore, 'real science' and machine learning technologies operate in very different domains and must not be regarded as competing.

The Pragmatic Value of Scientific Knowledge in Epistemic Tasks

Empiricist epistemologies do not deny the pragmatic value of science and agree indeed that *pragmatic criteria* play a role in the *acceptance* of theories, but only deny that pragmatic criteria add to the epistemic value of theories (e.g., Van Fraassen 1980). In section "Empiricist epistemologies", it has been argued that machine-made data-models may become capable to perform better in regard of *epistemic criteria* (esp. empirical adequacy regarding the data) as compared to human-made scientific knowledge. In addition, it has been suggested that the generation and use of data-models for all kinds of *epistemic tasks* can be carried out by machine learning technology, which will in many cases perform better than human scientists who aim to generate and use scientific knowledge for similar tasks (see overview in section "Machine-learning"). It has also been argued that machine-made data-models usually are incomprehensible, opaque, and even inaccessibly 'sitting' in the machine, to the effect that they cannot be used by human epistemic agents. Therefore, machines do not produce scientific knowledge as in 'traditional' scientific practices— i.e., epistemic entities such as theories, laws, models and concepts that can be obtained from the machine and utilized in, say, self-chosen epistemic tasks by humans. Even if it were possible to obtain the data-models from the machine, they would be useless for epistemic uses by humans as these data-models do not meet relevant pragmatic criteria to enable such uses. The other way around, in order to be useful for humans in performing epistemic tasks, scientific knowledge must also meet pragmatic criteria.

The crux of pragmatic criteria such as *consistency*, *coherency*, *simplicity*, *explanatory power*, *scope*, *relevance* and *intelligibility* in generating and accepting scientific knowledge is to render scientific knowledge manageable for humans in performing epistemic tasks. I will leave unanswered whether machine learning technologies cannot offer this in principle. But if they cannot, a future without science would require machines to take over every epistemic task, which seems unlikely already regarding our daily interactions with the world.

Preparing the Data

Much needs to be in place before the machine-learning can even begin. Data-sets need to be generated, prepared and gathered, which requires epistemic activities by humans, such as designing experimental set-ups and measurement equipment (e.g., as in the experiments of Boyle, etc. in section "Introduction"). These epistemic tasks require scientific and background knowledge. As stated above, knowledge must meet specific pragmatic criteria to be manageable when performing epistemic tasks. For example, knowledge must be such that epistemic agents can see which real-world target-systems the knowledge is applicable to—for example, in order to recognize or explain specific phenomena in the data-generating experimental set-up. Conversely, it requires of scientists to have the cognitive ability to *think*, *theorize*, *conceptualize*, *explain*, *mathematize*, and *interpret* by means of scientific knowledge when performing epistemic tasks, not only when setting up the data-generating instrumentation and seeing to its proper functioning, but also in assessing and interpreting the data, drawing relationships between data from different sources, and for making the distinction between 'real' phenomena and artifacts. These crucial cognitive abilities go well beyond what empiricist epistemologies can explain, require, or allow in view of the requirements of objectivity.

The necessity to prepare data that are *about something in the real world* also implies that *phenomena* are crucial in scientific practices, even when only aiming at the generation of data for machine-learning processes. Harking back to the discussion above, the way in which Bogen and Woodward (1988) think about phenomena forces them to accept that *phenomena* coincide with *data-models*. However, this notion of physical phenomena is far too narrow regarding the uses of this notion in scientific practices, even if only practices aimed at measuring data. The description of a physical phenomenon such as 'lead melts at 327.5 + 0.1 degrees C' is not grasped by the number (i.e., the value) in this proposition. In contrast to the data-model that is statistically derived from the measurements in the way suggested by B&W, the described phenomenon can be analyzed in terms of a set of interrelated *heterogeneous* aspects, such as: the *observation* that substances (including lead) can melt; an *understanding* of the concept 'temperature' (also see Chang 2004); the *observed* regularity that a substance (including lead) always melts at approximately the same temperature; an *understanding* of the concept 'melting-point'; the *conception* that 'having a melting-point' is a specific characteristic of substances; the *regulative principle* that at the same (experimental) conditions the same effects will happen (Boon 2012b); the *assumption* that the temperature at which a substance melts (the melting-point) is an exact number; the *assumption* that the temperature can be measured at a pretty high accuracy; the *decision* or *assumption* that the observed fluctuations in the observed melting temperatures are due to (partially) unknown causes of the experimental set-up and measurement tools (e.g., Mayo 1996); and finally, an *understanding* of the workings of the measurement tools. In short, the full conception of the physical phenomenon consists of a collection of heterogeneous, mutually related but heterogeneous aspects, which are generated in a number of cognitive actions by human scientists, rather than being a statistically derived number only.

This elaboration of B&W's example shows that skills, knowledge, and understanding of scientists are required to establish both the physical phenomenon—which involves the experimental and measurement set-up, and also their proper operations to get a stable and reproducible measurement of the temperature at which lead melts—, as well as the *conception of the phenomenon*, even if the phenomenon under study is as simple as 'lead melts at 327.5 + 0.1 degrees C.' Additionally, this brief analysis shows that the formation of the concept of a phenomenon and physically establishing it in an experimental set-up, go hand-in-hand, and necessarily involve all kinds of basic assumptions that cannot be empirically tested (Chang 2004; Boon 2012a, b, 2015).

Expanding on this analysis, it can be argued that empiricist epistemologies are flawed in believing that the *theory-ladenness* of data is fundamentally problematic as it threatens the objectivity of science. More specifically, Bogen and Woodward are mistaken to hold that phenomena should not be some kind of low-level theories (claim 4). To the contrary, *theory-emptiness* of data fed to machine-learning processes would really be a problem. In actual scientific practices, the production of data representing supposed physical phenomena usually develops in a process of *triangulation* together with the development of the experimental set-up and measurement techniques and with the construction and application of scientific knowledge of all kinds. The data, phenomena and theory, as well as our interpretations of measurements and understanding of the working of instruments and experimental set-ups are intrinsically conceptually entangled (e.g., Chang 2004; Feest 2010; Van Fraassen 2008, 2012; Boon 2012a, 2017a; Van Fraassen 2008, 2012).

Epistemic Tasks in Engineering and Biomedical Sciences

In machine-learning-technologies, descriptions of (physical or social) phenomena are reduced to (and represented by) data-models, which is considered unproblematic within empiricist epistemologies. As sketched above, the data or data-model representing the phenomenon entail hardly any information relevant to epistemic tasks in dealing with phenomena, for instance, in aiming to *interact* with the targeted phenomenon in one or another way.

Yet, these kinds of epistemic tasks are at the center of the so-called applied sciences such as the engineering and biomedical sciences. These research practices aim at scientific knowledge about targeted (bio)physical phenomena, and about technological instruments that can possibly produce or control them—for the sake of the targeted phenomenon, not first theories, which are considered to be the focus of basic sciences.[11] These practices function in the way sketched above, which is to say that experimentally producing and investigating targeted phenomena (e.g., a phenomenon we aim to produce for a specific technological or medical function) is *entangled* with the generation of scientific knowledge and the development of technological instruments and measurement apparatus relevant to the phenomenon. Every tiny step in these intricate research processes involves epistemic tasks—e.g., to explain, interpret, invent, idealize, simplify, hypothesize, model, mathematize, design, and calculate—for which all kinds of practical and scientific knowledge is crucial and needs to be developed in the research process. Therefore, scientific knowledge needs to be *comprehensible* to the extent that it allows for these epistemic

[11] In other work, I have explained from a range of different philosophical issues, the crucial role of phenomena in the 'applied' research practices and what this means for our philosophical understanding both of scientific knowledge and of the aim of science (Boon 2011, 2012a, b, 2015, 2017a, c, forthcoming). The idea that these application-oriented scientific research practices aim at scientific knowledge in view of *epistemic tasks* aimed at learning how to do things with (often unobservable, and even not yet existing) physical phenomena has led to the notion of scientific knowledge as *epistemic tool* (Boon and Knuuttila 2009; Knuuttila and Boon 2011; Boon 2015; Boon 2017b,c; also see Nersessian 2009; Feest 2010; Andersen 2012). The original idea of scientific knowledge (or, originally more narrowly stated, 'scientific models') as epistemic tools, proposes to view scientific knowledge—such as descriptions, concepts, and models of physical phenomena—firstly as *representations of scientists' conceptions of aspects of reality*, rather than representations in the sense of a two-way relationship between knowledge and reality (as in anti-realist empiricist epistemologies as well as in scientific realism). The point of this (anti-realist) view is that someone can represent her conception (comprehension, understanding, interpretation) of aspects of reality by means of *representational means* such as text, analogies, pictures, graphs, diagrams, mathematical formula, and also 3D material entities. Notably therefore, scientists' conceptions of observable as well as unobservable phenomena arrived at by intricate reasoning processes (creative, inductive, deductive, hypothetical, mathematical, analogical, etc.), which employ all kinds of available epistemic resources, can be *represented*. By representing, scientists' conceptions become *epistemic constructs* that are public and transferable. Knuuttila and I have called these constructs *epistemic tools*, that is, conceptually meaningful tools that guide and enable humans in performing all kinds of different *epistemic tasks*.

tasks. Especially regarding these kinds of practically oriented scientific research practices in which human scientists aim at comprehensible scientific knowledge as well as epistemic and practical resources such as measurement instruments, technological procedures, and methods, it is inconceivable that machine-learning-technologies will make science and scientists superfluous.

The Error of Empiricism

Empiricist epistemologies insufficiently account for the types of epistemic tasks that are crucial for the development and use of epistemic and practical tools, which in turn are used in the development of, for instance, medical technologies. This shortcoming already applies to the generation of data that can be fed to machines that generate data-models for specific purposes. Empiricist epistemologies therefore miss out on crucial aspects of the uses and generation of scientific knowledge (theories, models, etc.) in intricate scientific processes taking place in application-oriented research practices like the engineering and biomedical sciences, and thus give room to beliefs such as defended by Anderson (2008). Rethinking the philosophical presuppositions of empiricist epistemologies that seem to force us to the view that science will become superfluous in the age of machine learning can help in gaining insights that bring the scientist back into science.

Acknowledgements This work is financed by an Aspasia grant (2012–2017 nr. 409.40216) of the Dutch National Science Foundation (NWO) for the project "Philosophy of Science for the Engineering Sciences." I want to acknowledge Koray Karaca for critical suggestions and fruitful discussions; I am indebted to him for introducing me to the machine learning literature, which I have gratefully used in writing the introduction to this theme. I also wish to acknowledge Marta Bertolaso and her team for organizing the expert meeting (March 5–6, 2018 in Rome) on the theme "Will Science remain Human? Frontiers of the Incorporation of Technological Innovations in the Biomedical Sciences," and participants of this meeting for their fruitful comments, especially by Melissa Moschella.

References

Abu-Mostafa, Y. 1995. Hints. *Neural Computation* 7: 639–671. https://doi.org/10.1162/neco.1995.7.4.639.

Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.- T. (2012). *Learning from data*. AMLbook.com. ISBN 10:1-60049-006-9, ISBN 13:978-1-60049-006-4

Alpaydin, E. 2010. *Introduction to Machine Learning*. Cambridge: The MIT Press.

Andersen, H. 2012. Concepts and Conceptual Change. In *Kuhn's The Structure of Scientific Revolutions Revisited*, ed. V. Kindi and T. Arabatzis, 179–204. Routledge.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine,* June 23. Retrieved from: https://www.wired.com/2008/06/pb-theory/

Bogen, J. 2011. 'Saving the Phenomena' and Saving the Phenomena. *Synthese 182* (1): 7–22. https://doi.org/10.1007/s11229-009-9619-4.

Bogen, J., and J. Woodward. 1988. Saving the Phenomena. *The Philosophical Review 97* (3): 303–352. https://doi.org/10.2307/2185445.

Boon, M. 2011. In Defense of Engineering Sciences: On the Epistemological Relations Between Science and Technology. *Techné: Research in Philosophy and Technology 15* (1): 49–71. Retrieved from http://doc.utwente.nl/79760/.

———. 2012a. Scientific Concepts in the Engineering Sciences: Epistemic Tools for Creating and Intervening with Phenomena. In *Scientific Concepts and Investigative Practice*, ed. U. Feest and F. Steinle, 219–243. Berlin: De Gruyter.

———. 2012b. Understanding Scientific Practices: The Role of Robustness Notions. In *Characterizing the Robustness of Science After the Practical Turn of the Philosophy of Science*, ed. L. Soler, E. Trizio, Th. Nickles, and W. Wimsatt, 289–315. Dordrecht: Springer: Boston Studies in the Philosophy of Science.

———. 2015. Contingency and Inevitability in Science – Instruments, Interfaces and the Independent World. In *Science as It Could Have Been: Discussing the Contingent/Inevitable Aspects of Scientific Practices*, ed. L. Soler, E. Trizio, and A. Pickering, 151–174. Pittsburgh: University of Pittsburgh Press.

———. 2017a. Measurements in the Engineering Sciences: An Epistemology of Producing Knowledge of Physical Phenomena. In *Reasoning in Measurement*, ed. N. Mößner and A. Nordmann, 203–219. London/New York: Routledge.

———. 2017b. Philosophy of Science in Practice: A Proposal for Epistemological Constructivism. In *Logic, Methodology and Philosophy of Science – Proceedings of the 15th International Congress (CLMPS 2015)*, ed. H. Leitgeb, I. Niiniluoto, P. Seppälä, and E. Sober, 289–310. College Publications.

———. 2017c. An Engineering Paradigm in the Biomedical Sciences: Knowledge as Epistemic Tool. *Progress in Biophysics and Molecular Biology* 129: 25–39. https://doi.org/10.1016/j.pbiomolbio.2017.04.001.

———. forthcoming. Scientific methodology in the engineering sciences. Chapter 4. In *The Routledge Handbook of Philosophy of Engineering*, ed. D. Michelfelder and N. Doorn. Routledge. Publication scheduled for 2019.

Boon, M., and T. Knuuttila. 2009. Models as Epistemic Tools in Engineering Sciences: A Pragmatic Approach. In *Philosophy of Technology and Engineering Sciences. Handbook of the Philosophy of Science*, ed. A. Meijers, vol. 9, 687–720. Elsevier/North-Holland.

Chang, H. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.

Craig, E. 1998. Duhem, Piere Maurice Marie. In *Routledge Encyclopedia of Philosophy: Descartes to gender and science.*, vol. 3, 142–145. London/New York: Routledge.

Da Costa, N.C.A., and S. French. 2003. *Science and Partial Truth. A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

Dai, W., et al. 2015. Prediction of Hospitalization Due to Heart Diseases by Supervised Learning Methods. *International Journal of Medical Informatics* 84: 189–197.

Duhem, P. 1954/[1914]. *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

———. 2015/[1908]. *To Save the Phenomena: An Essay on the Idea of Physical Theory from Plato to Galileo*. Trans. E. Dolland and C. Maschler. Chicago: University of Chicago Press

Esteva, A., B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, and S. Thrun. 2017. Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. *Nature 542*: 115. https://doi.org/10.1038/nature21056.

Feest, U. 2010. Concepts as Tools in the Experimental Generation of Knowledge in Cognitive Neuropsychology. *Spontaneous Generations 4* (1): 173–190.

Giere, R.N. 1988. *Explaining Science*. Chicago/London: The University of Chicago Press.

———. 2010. An Agent-Based Conception of Models and Scientific Representation. *Synthese* 172 (2): 269–281. https://doi.org/10.1007/s11229-009-9506-z.

Glymour, B. 2002. Data and Phenomena: A Distinctions Reconsidered. *Erkenntnis 52*: 29–37.

Hempel, C.G. 1962. Explanation in Science and Philosophy. In *Frontiers of Science and Philosophy*, ed. R.G. Colodny, 9–19. Pittsburgh: University of Pittsurgh Press.
———. 1966. *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall.
Hofree, M., J.P. Shen, H. Carter, A. Gross, and T. Ideker. 2013. Network-based Stratification of Tumor Mutations. *Nature Methods 10*: 1108. https://doi.org/10.1038/nmeth.2651. https://www.nature.com/articles/nmeth.2651#supplementary-information.
Humphreys, P. 2009. The Philosophical Novelty of Computer Simulation Methods. *Synthese* 169 (3): 615–626. https://doi.org/10.1007/s11229-008-9435-2.
Knuuttila, T., and M. Boon. 2011. How Do Models Give Us Knowledge? The Case of Carnot's Ideal Heat Engine. *European Journal for Philosophy of Science 1* (3): 309–334. https://doi.org/10.1007/s13194-011-0029-3.
Kourou, K., et al. 2015. Machine Learning Applications in Cancer Prognosis and Prediction. *Computational and Structural Biotechnichnology Journal* 13: 8–17.
Lemm, S., et al. 2011. Introduction to Machine Learning for Brain Imaging. *NeuroImage* 56: 387–399.
Libbrecht, M.W., and W.S. Noble. 2015. Machine Learning Applications in Genetics and Genomics. *Nature Reviews Genetics* 16: 321–332.
Lima, A.N., et al. 2016. Use of Machine Learning Approaches for Novel Drug Discovery. *Expert Opinion on Drug Discovery* 11: 225–239.
Mayo, D.G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
McAllister, J.W. 1997. Phenomena and Patterns in Data Sets. *Erkenntnis 47* (2): 217–228. https://doi.org/10.1023/A:1005387021520.
———. 2007. Model Selection and the Multiplicity of Patterns in Empirical Data. *Philosophy of Science 74* (5): 884–894. https://doi.org/10.1086/525630.
———. 2011. What do Patterns in Empirical Data Tell Us About the Structure of the World? *Synthese 182* (1): 73–87. https://doi.org/10.1007/s11229-009-9613-x.
Mena, J. 2011. *Machine Learning Forensics for Law Enforcement, Security, and Intelligence*. Boca Raton: CRC Press.
Nersessian, N.J. 2009. *Creating Scientific Concepts*. Cambridge, MA: MIT Press.
Odone, F., M. Pontil, and A. Verri. 2009. Machine Learning Techniques for Biometrics. In *Handbook of Remote Biometrics. Advances in Pattern Recognition, ch. 10*, ed. M. Tistarelli, S.Z. Li, and R. Chellappa. London: Springer.
Olszewska, J.I. 2016. Automated face Recognition: Challenges and Solutions. In *Pattern Recognition – Analysis and Applications*, ed. S. Ramakrishnan, 59–79. InTechOpen. https://doi.org/10.5772/62619.
Phua, C., et al. (2010). A comprehensive survey of data mining-based fraud detection research. https://arxiv.org/abs/1009.6119
Suppe, F. (1974). *The Structure of Scientific Theories* (1979 second printing ed.). Urbana: University of Illinois Press.
———. 1989. *The Semantic Conception of Theories and Scientific Realism*. Urbana/Chicago: University of Illinois Press.
Suppes, P. 1960. A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. *Synthese* 12: 287–301.
Tcheng, D.K., A.K. Nayak, C.C. Fowlkes, and S.W. Punyasena. 2016. Visual Recognition Software for Binary Classification and Its Application to Spruce Pollen Identification. *PLoS ONE 11* (2): e0148879. https://doi.org/10.1371/journal.pone.0148879.
Van Fraassen, B.C. 1977. The Pragmatics of Explanation. *American Philosophical Quarterly 14*: 143–150.
———. 1980. *The Scientific Image*. Oxford: Clarendon Press.
———. 2008. *Scientific Representation*. Oxford: Oxford University Press.
———. 2012. Modeling and Measurement: The Criterion of Empirical Grounding. *Philosophy of Science* 79 (5): 773–784.

Van Liebergen, B. 2017. Machine Learning: Revolution in Risk Management and Compliance? *The Capco Institute Journal of Financial Transformation* 45: 60–67.
Woodward, J.F. 2011. Data and Phenomena: A Restatement and Defense. *Synthese 182* (1): 165–179. https://doi.org/10.1007/s11229-009-9618-5.
Wuest, T., et al. 2016. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Production & Manufacturing Research* 4 (1): 23–45.