Big Data – The New Science of Complexity

Wolfgang Pietsch (pietsch@cvl-a.tum.de)
Munich Center for Technology in Society, Technische Universität München, Germany

Abstract

Data-intensive techniques, now widely referred to as 'big data', allow for novel ways to address complexity in science. I assess their impact on the scientific method. First, big-data science is distinguished from other scientific uses of information technologies, in particular from computer simulations. Then, I sketch the complex and contextual nature of the laws established by data-intensive methods and relate them to a specific concept of causality, thereby dispelling the popular myth that big data is only concerned with correlations. The modeling in data-intensive science is characterized as 'horizontal'—lacking the hierarchical, nested structure familiar from more conventional approaches. The significance of the transition from hierarchical to horizontal modeling is underlined by a concurrent paradigm shift in statistics from parametric to non-parametric methods.

1. Introduction

For some time, computer scientists have been speaking of a fourth paradigm in scientific research, in addition to—as they say—theory, experiment, and simulation. The classic statement is by Jim Gray, a Turing award winner and former employee of Microsoft Research. In one of his last talks before he went missing at sea in 2007, Gray declared: 'The world of science has changed, and there is no question about this. The new model is for the data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline. The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration.' (Gray 2007, xix) The talk was transcribed and resulted in a collected volume titled *The Fourth Paradigm* (Hey et al. 2009).

Big data is the latest buzzword in information technology. The term has been defined in various ways, most of them referring to the pure amount of information or to the technical challenges that big data poses in terms of the so-called 'three Vs' volume, velocity and variety (Laney 2001). From a methodological perspective, other characteristics are more relevant and will be taken as defining features throughout this essay: big-data science requires (i) *data representing all relevant configurations of the examined phenomenon*. For complex phenomena, this implies high-dimensional data, i.e. data sets involving many parameters, as well as a large number of observations covering a wide range of combinations of these parameters. Ideally, the data contains all necessary information to make predictions without referring to general hypotheses or rules. (ii) The second crucial characteristic concerns the *automation of the entire scientific process,* from data capture to processing to modeling. This allows sidestepping some of the limitations of the human cognitive apparatus but also leads to a loss in human understanding regarding the results of data-intensive science.

The current debate on big data is laden with philosophy-of-science concepts like explanation, modeling or causation. However, lack of conceptual clarity and rigor has led to considerable confusion regarding the real impact and methodological novelty—for example when debating controversial statements such as that big data allegedly involves a shift from causation to correlation (Schönfelder & Cukier, Ch. 4) or that it implies 'the end of theory' (Anderson 2008). Philosophy of science can provide the necessary conceptual tools to adequately assess such claims.[1]

The essay provides an overview of the main issues that are relevant in connection with big data from a philosophy-of-science perspective. Following Gray, I argue in Section 2 that big-

[1] To the best of my knowledge, the only place where data-intensive science has been discussed from a philosophy-of-science viewpoint is pioneering work by Sabina Leonelli including an edited volume on big data in biology (Leonelli 2012a) as well as several other papers on various issues, for example the significance of classification (Leonelli 2012b). Leonelli arrives at somewhat different conclusions, for example concerning the role of theory and hypotheses in data-intensive science. This is largely due to differences in aim. While Leonelli wants to account for the actual role that data-intensive methods play in biology, I try to sketch an idealized picture of data-intensive science as a research mode mostly abstracting from complications arising in scientific practice.

data science should be distinguished from computer simulations as two distinct ways how information technologies are employed in science. Broadly speaking, the latter can be understood as computationally examining the implications of a mathematical model, while the former searches for causal structure inherent in large data sets.

In Section 3, the scientific laws resulting from big-data techniques are given closer scrutiny.[2] Contrary to popular conception, big data is interested not only in correlations but essentially aims at causal knowledge. This will be argued for on the basis of an account of eliminative induction in the tradition of Francis Bacon and John Stuart Mill. Big-data laws show a number of characteristics that are familiar from methodological studies in the applied sciences. They are (i) complex specifying a large number of conditions under which a phenomenon will occur or, more or less equivalently, they only hold ceteris paribus. Relatedly, they are (ii) contextual, i.e. such laws often have only a small number of instantiations.

In Section 4, the modeling in big-data science is analyzed. It lacks a number of features that are typical for more conventional scientific modeling geared at an efficient reduction of data and an adequate structuring of knowledge. By contrast, big-data-models (i) have no pronounced hierarchical structure, which implies that (ii) they lack substantial explanatory power. (iii) They rely on few modeling assumptions and (iv) they are quite complex because little of the data is actually discarded. The big-data type of modeling will be referred to as *horizontal* modeling in contrast to the *hierarchical* modeling characteristic of more traditional scientific methodology.

In Section 5, the explanatory power of big-data science is examined in further detail. For this purpose, two kinds of explanations are distinguished: unificationist explanation referring to general laws or rules and causal explanation. As should be clear from the discussion above, big data allows for causal explanation but largely fails to provide unificationist explanation.

Section 6 provides further evidence for a qualitative change in the nature of modeling by describing a concurrent paradigm shift in statistics from parametric to non-parametric methods. In the latter, predictions are calculated on the basis of the original data and a suitable algorithm without the detour over a parametric model, which attempts to summarize the data in terms of a relatively simple equation involving a few parameters, e.g. a linear function or a Gaussian distribution.

Chapter 7 concludes by summarizing how big-data science provides new ways to deal with complexity. Owing to the automation of the entire scientific process, the epistemic boundary conditions of computers and sensor networks become relevant. The resulting computer or network epistemology turns out more apt to deal with complexity than our limited human capabilities.

---

[2] As Peter Norvig, research director at Google, writes: 'In complex, messy domains, particularly game-theoretic domains involving unpredictable agents such as human beings, there are no general theories that can be expressed in simple equations like $F = m\ a$ or $E = m\ c^2$. But if you have a dense distribution of data points, it may be appropriate to employ non-parametric density approximation models such as nearest-neighbors or kernel methods rather than parametric models such as low-dimensional linear regression.' (2009) Many of the ideas elaborated in this essay take much inspiration from scattered writings of Norvig.

2. Big-data science versus computer simulations

The impact of information technologies on the scientific method has mostly been discussed with respect to computer simulations, e.g. of the climate, the early universe or of financial markets (e.g. Winsberg 2010; Humphreys 2004; Gramelsberger 2010; Frigg & Rice 2009; Beisbart & Norton 2012). I will now argue that big data as characterized in the introduction allows for methodologically novel applications of information technologies in science. The argument draws on Jim Gray's distinction between computational science on the one hand, dealing with simulations, and informatics on the other hand, which is concerned with collecting and analyzing information and in his view constitutes the fourth paradigm (2007, xix-xx).

The notion of computer simulations as depicted in the philosophical literature subsumes a number of different practices. A useful distinction concerns computer simulations in a narrow and in a broad sense (e.g. Frigg & Rice 2009). The former refers to the use of computers to solve equations that are not accessible analytically, e.g. scattering cross sections of complex atoms or molecules. The latter refers to the computational implementation of analytically intractable models. Examples are computational molecular dynamics or agent-based modeling in the social sciences.

All computer simulations have in common that they rely on elaborate modeling assumptions that originate outside of the computer, for example in terms of dynamical equations or rules of evolution (Winsberg 2013). Since the computer serves as tool to explore consequences of these modeling assumptions, computer simulations are mainly of deductive nature. This fits well with Claus Beisbart's proposal that they should be understood as arguments (2012) and with how typical simulations are characterized by Eric Winsberg as a sequence of theory-model-treatment-solver-results (2010, 11). According to this author, a model is first derived from a general theory. Then, specific values are assigned to the parameters and boundary conditions in the treatment phase. Finally, model and treatment are translated into an algorithm to yield the desired results.

This brief characterization of computer simulations illustrates some of the differences in comparison with big-data science:

(i) The starting point is different. While computer simulations analyze a *mathematical model*, big-data science examines a collection of *data*. While the former are largely model-driven, the latter is mainly exploratory.

(ii) Relatedly, the nature of the inferences differs. While computer simulations mainly derive *deductive* consequences of a computational model, big-data science aims at an *inductive*, mostly causal analysis of a phenomenon as will be further elaborated in Section 3.

(iii) A crucial distinction concerns the *automation of the scientific process*. While for big-data science, the entire process from data collection to modeling to prediction is automated, computer simulations only assist with inferences from a mathematical model, i.e. with a minor part of the scientific process.

(iv) Consequently, the loss in *explanatory power* is different. In computer simulations, the modeling assumptions can be explicitly stated and remain a firm explanatory ground for the model predictions. Due to the automation of the modeling process, such a basis is lacking in big-data science. The issue will be addressed in further detail in Section 5.

(v) Computer simulations are sometimes described as *experiments with a model* (e.g. Rohrlich 1991, 507), in the sense that the model properties are analyzed by changing parameters and initial conditions while observing the impact on the model outcome. By contrast, little of this experimental character appears in big-data modeling, because all relevant configurations of the system are already given in terms of data. This correlates well with the fact that computer simulations are usually carried out when data is scarce or not accessible, while big-data science relies on an abundance of data.

(vi) Last not least, the theoretical differences are mirrored in the *computational hardware* employed in the different practices. While computer simulations mainly rely on the computational power and storage capacity of a single computer or computer cluster, big-data science draws on a wider variety of information technologies including sensor devices enabling automated data capture, such as gene sequencing machines, and network infrastructure that links these sensors with servers and computers, thus allowing the efficient and automated buildup and processing of large high-dimensional data sets.

3. The contextualization of science: Big-data laws and causality

In a much-cited and influential article, journalist Chris Anderson, at the time editor in chief of the technology and lifestyle magazine Wired, wrote some controversial remarks how big data affects science: 'Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.' (2008) The ideas condensed in this brief statement have been readily picked up by both mainstream media and academia. Occasional critical reactions have mostly focused on Anderson's thesis concerning 'the end of theory'.[3] I will now explain why Anderson's claim about the significance of causation is mistaken, the changes in modeling will be discussed in Section 4.

Nancy Cartwright once highlighted as central feature of causation that causal knowledge can ground effective strategies (1983, Ch. 1). A mere correlation cannot tell how to effectively intervene in the world, e.g. the birthrate cannot be changed by increasing the population of storks, even though studies consistently show a significant correlation between both quantities. By contrast, headaches can be cured by taking acetylsalicylic acid because there is a direct causal connection. Thus, if big data is about making predictions regarding interventions, e.g. making people vote for a specific candidate or click on a certain ad, then it must aim for causal knowledge and cannot be satisfied only with correlations. Note that reliable predictions, even if they cannot justify interventions, always require causal

[3] It is quite revealing that Anderson misquotes Google research director Peter Norvig with the statement: 'All models are wrong, and increasingly you can succeed without them.' (2008) In a reply on his web page, Norvig clarifies: 'That's a silly statement, I didn't say it, and I disagree with it.' (2009) Certainly, there will always be modeling assumptions in any scientific endeavor. Norvig's actual point had concerned changes in the nature of modeling resulting from big data (cp. Section 4).

knowledge, at least regarding the existence of a common cause. For example the weather can be predicted on the basis of a barometer reading because air pressure is a common cause of both, while of course the weather cannot be changed by tinkering with the barometer needle.

Notwithstanding this simple argument, the phrase 'correlation supersedes causation' is ubiquitous in the debate on big data.[4] Even academics endorse the idea like psychologist Gary Marcus of New York University[5] or Viktor Mayer-Schönberger of the Oxford Internet Institute, who makes it a central thesis of his recent book on big data (2013) coauthored with Kenneth Cukier of the British weekly *The Economist*. The confusion essentially results from a conflation of causation with mechanistic explanation. However, recent technical work on causality (e.g. Pearl 2000; Spirtes et al. 2000) as well as conceptual analysis (Woodward 2003) show that knowledge about prediction and control can be derived without a deeper understanding of any underlying mechanism. In this article, I will argue for the same point on the basis of a difference-making account of causality.

In the following, I will first outline the basic structure of typical big-data problems, I will then present the mentioned difference-making account and will proceed to show how it plays a role in some of the classic big-data algorithms.

3a Big-data problems

Typical big-data problems concern classification or regression of an output variable y with respect to a large number of input parameters x, also called predictor variables or covariates, on the basis of large training sets. The main differences compared with conventional problems in statistics consist in the high-dimensionality of the input variable and the amount of data available about various configurations or states of the system. For example, an internet store wants to know how likely someone buys a certain product depending on surf history, various cookies and a user profile as well as based on data of other users who have either bought or failed to buy the product. A medical researcher examines which combinations of genetic and environmental factors are responsible for a certain disease. A political adviser is interested how likely a specific individual is going to vote for a certain candidate based on a profile combining for example voting history, political opinions, general demographics, or consumer data.

In a *classification* problem, the output variable has a finite number of possible values. In a *regression* problem, the output variable is continuous. In order to establish an adequate and reliable model, extensive training and test data is needed. Each instance in the training and test sets gives a value for the output variable dependent on at least some of the input parameters. The training data is used to build the model, e.g. determine relevant parameters, the test data to validate and verify the model. Using part of the data to determine the accuracy of a model is commonly referred to as cross-validation.[6]

---

[4] Compare for example the recent compilation on http://www.forbes.com/sites/gilpress/2013/04/19/big-data-news-roundup-correlation-vs-causation/ accessed 15.6.2013

[5] http://www.newyorker.com/online/blogs/elements/2013/04/steamrolled-by-big-data.html accessed 31.7.2013

[6] An excellent introductory textbook is Russell & Norvig (2009).

In this essay, we cannot delve into the technical details of the various algorithms employed in big-data science, such as support vector machines, forests or neural networks. I will now however introduce two simple algorithms, classificatory trees and naïve Bayes, in order to establish how under certain premises these can derive causal knowledge.

Classificatory trees (Russell & Norvig 2010, Ch. 18.3.3) are used to determine whether a certain instance belongs to a particular group A depending on a large number of parameters C1, …, CN. With help of the training data, the tree is set up recursively. First, the parameter CX is determined that contains the largest amount of information with respect to the classification of the training data, as formally measured in terms of Shannon entropy. If CX classifies all instances correctly, the procedure is terminated. Otherwise, two subproblems remain, namely classifying when CX is present and when it is absent. This step is repeated until either all instances are classified correctly or no potential classifiers are left. If the algorithm is successful, the resulting tree structure gives a Boolean expression of necessary and sufficient conditions for A, which can be interpreted as a complex scientific law: e.g. if $(C3C2 \vee C4\neg C2)C1 \vee C6C5\neg C1$, then A.

Another simple big-data algorithm is naïve-Bayes classification, which is widely used for example in the identification of spam emails. The problem structure is the same as in the case of classificatory trees. A number of parameters C1, …, CN, representing for example certain words or sequences of words appearing in emails, is used to determine the probability that a specific instance is A or not, e.g. that an email is spam or not. Using Bayes' Theorem:

$$P(A|C1, \ldots, CN) = P(A)\, \Pi_{i=1,\ldots,n} P(Ci|A) / P(C1, \ldots, CN)$$

The 'naïve' part of the algorithm is that the parameters Ci are assumed to be independent given A, i.e. $P(C1, \ldots, CN|A) = \Pi_{i=1,\ldots,n} P(Ci|A)$, which of course may not be the case. As with classificatory trees, a training set is used to develop the model. It provides representative frequencies for joint occurrences of A and the different Ci and thereby the probabilities P(Ci|A), P(A), and P(C1, …, CN). On this basis, new instances can be classified given certain values Ci. Again, test instances can be set aside to cross-validate the model.

3b An appropriate account of causality for big-data problems

The framing of big-data problems as a mapping of boundary conditions to an outcome variable fits well with eliminative induction[7]—a scientific method whose history reaches back at least to the methodological writings of medieval thinkers like Robert Grosseteste and William of Ockham. The most elaborate frameworks are Francis Bacon's *method of exclusion* (1620/1994, Bk. 2), which arguably was considered the methodological foundation for modern science until the end of the 19th century, and John Stuart Mill's *methods of elimination* (1886, Bk. III, Ch. VIII). In the 20th century, eliminative induction has received little attention presumably due to prevailing anti-inductivist views. In the following, I can only highlight a few features that are crucial in the context of big-data science. For a more comprehensive overview of the method, compare Pietsch (2013).

[7] not to be confused with a looser use of the same term in the sense of eliminating hypotheses until only the correct one remains

In eliminative induction, a phenomenon A is examined under the systematic variation of potentially relevant boundary conditions C1, …, CN with the aim of establishing *causal relevance* or *irrelevance* of these conditions, relative to a certain context B determined by further boundary conditions. Obviously, the framing corresponds exactly to that of big-data problems from the previous section. Eliminative induction provides the basic logic for exploratory experimentation when theoretical knowledge of a phenomenon is scarce.

The best known and arguably most effective method is the so-called *method of difference* that establishes causal relevance of a boundary condition CX by comparing two instances which differ only in CX and agree in all other circumstances C. If in one instance, both CX and A are present and in the other both CX and A are absent, then CX is causally relevant to A.[8] There is a twin method to the method of difference, called the *strict method of agreement*, which establishes causal irrelevance, if the change in CX has no influence on A. Eliminative induction can deal with functional dependencies and an extension of the approach to statistical relationships is straightforward.

Thus, causal relevance is a three-place relation: a boundary condition C is relevant to a phenomenon A with respect to a certain background B of further conditions that remain constant if causally relevant or are allowed to vary if causally irrelevant. The restriction to a context B is necessary because there is no guarantee that in a different context B*, the causal relation between C and A will continue to hold. Causal laws established by eliminative induction thus have a distinct *ceteris-paribus* character.

Eliminative induction can identify causal laws in terms of necessary and sufficient conditions for a phenomenon, relative to a certain context. More specifically, a cause established by this method can be formulated in terms of an *INUS*-condition (Mackie 1965): i.e. an *I*nsufficient, but *N*on-redundant part of an *U*nnecessary but *S*ufficient condition. Extensive information of all potentially relevant boundary conditions in as many different situations as possible is necessary to establish reliable causal knowledge of complex phenomena by means of eliminative induction. Exactly this kind of information is provided by big data.

Furthermore, it is easy to see that the logic of eliminative induction is implicit in both big-data algorithms of Section 3a. In the case of classificatory trees, the resulting structure clearly fits the logic of necessary and sufficient conditions. Under certain premises, which cannot be discussed here due to lack of space, the algorithm will identify true causes. For example, if there is a single necessary and sufficient condition CX among the C, the algorithm will return CX as cause.

Similarly, in the naïve-Bayes approach the presence of a sufficient condition, e.g. CX∧CY (i.e. P(A | CX,CY) = 1), directly implies P(A | C1,…,CN) = 1. Conversely, the absence of a necessary condition CX (i.e. P(CX|A) = 1) implies P(A| C1,…,¬CX,…,CN) = 0. Under certain additional premises and if the true causes are among the C, the naïve-Bayes approach results in the correct classification. While both algorithms are less rigorous than eliminative induction, they stand a good chance to identify the correct causes.

[8] Details under which premises such inferences are actually valid can be found in Pietsch (2013).

In general, including more parameters C will increase the probability that the actual cause of A might be among them, while admittedly also increasing the probability for spurious correlations, i.e. that boundary conditions accidentally produce the right classification. However, more data in terms of instances of different configurations can reduce the probability for such spurious correlations. Thus, more data in terms of parameters and instances will generally increase the chance that correct causal relations are identified by big-data algorithms.

3c Big-data laws

Mostly, the causal laws established in big-data science are complex relationships that (i) involve a *large number of parameters* and (ii) do not adhere to *simple functional dependencies*. Relatedly, these laws (iii) often hold only *in a few instances*, i.e. their applicability is very limited. Furthermore, there is no reason why (iv) a *hierarchy of increasing universality* should exist into which these laws can be systematically integrated.[9]

Certainly, generalized laws can always formulated, but for the price that these exhibit a large number of exceptions which renders them fairly useless beyond a certain heuristic value. As a tentative example, the statement that smoking causes lung cancer can provide useful heuristics but is obviously false as a generalized law, since various boundary conditions exist, under which the statement is wrong. Big-data science provides a methodology for the complex sciences which fail to show the neat, hierarchical structure of laws as familiar for example from physics.

Eliminative induction is well-suited to identify big-data laws with the mentioned characteristics. The method can deal with large numbers of parameters and with arbitrarily complex relationships that cannot be condensed into simple equations.[10] It can account for context-dependency since it establishes causal relevance only with respect to a certain background of further boundary conditions. Furthermore, eliminative induction can identify causal relations that hold only in a few instances or even singular causation as follows from the method of difference, in which only two instances are compared. No regularity has to be taken into account to determine a cause.[11]

Before the advent of big data, the causal structure of complex phenomena was extremely difficult to analyze as it was almost impossible to efficiently collect and handle high-dimensional data. Mostly, scientists worked with dubious simplifications, e.g. that all but a few main influences on a phenomenon could be neglected and that these main influences adhered to simple functional relationships. But these assumptions, which are for example implicit in the structural equation modeling that is ubiquitous in the social sciences, were

[9] These characteristics fit well current discussions about the role of complexity in science, see in particular Mitchell (2008).
[10] Compare the discussion about functional dependence in Pietsch (2013).
[11] Note that eliminative induction does *not* lead to a regularity view of causality. In fact, the notions of causality resulting from enumerative and from eliminative induction are entirely distinct. While the classic regularity view of causality is based on enumerative induction and focuses on the number of instances, eliminative induction emphasizes the variation of boundary conditions.

chiefly motivated not by empirical considerations but merely by the need to make the data fit the available scientific toolbox.

By taking into account an ever larger number of boundary conditions, big data makes possible an increasing *contextualization of science*. Data-intensive science based on eliminative induction does not examine phenomena at an ensemble level, i.e. what collections of instances have in common, but can address the individual case. Examples of this development are the promise of personalized medicine that many health professionals see in big data, the microtargeting approaches of big-data political campaigns, personalized web search or individualized online advertising.

3d Data threshold

For many phenomena there exists a relatively sudden change when data-driven approaches become effective (Halevy et al. 2009)—a transition point that could be called a *data threshold*. Halevy et al give a plausible explanation for its existence: 'For many tasks, once we have a billion or so examples, we essentially have a closed set that represents (or at least approximates) what we need, without generative rules.' (2009, 9) The notion of causality sketched in Section 3b accounts quite well for this data threshold, since all causal relations can be derived with the method of difference and the strict method of agreement once all relevant configurations of the examined phenomenon are represented in the data (cp. characteristic i from the introduction). Under such circumstances, no abstract or general laws are necessary to make predictions about the phenomenon, which enables the horizontal modeling to be described in the next section.

4. Horizontal modeling

4a Computer epistemology

As stated in the introduction, a central feature of big data concerns the automation of the entire scientific process from data collection to data processing and model building to making novel predictions. Data-intensive science is the first genuine machine science in which all essential steps can be automated.

The epistemic boundary conditions of data-intensive science differ substantially from those under which the human cognitive apparatus models phenomena, in particular in terms of storage capacity and computational power. Most importantly, while humans have to be very efficient in determining which data to keep and which to forget or not even perceive in the first place, computers can often store and handle all the data they are collecting.

As argued in Section 3, big-data modeling is particularly suited for the causal analysis of complex phenomena, when large amounts of data have to be taken into account, with which human memory cannot deal anymore. On the other hand, the models used in big-data science are often much simpler than the elaborate theoretical structures developed by humans mainly for the purpose of data reduction. If there is enough data to adequately represent a system, there is no need for complex, hierarchical models anymore—at least with respect to

predictions. As Halevy et al write: 'invariably, simple models and a lot of data trump more elaborate models based on less data.' (2009, 9) Before providing an overview of the characteristics that distinguish big-data modeling from more conventional scientific modeling let us first take a look at two examples that can illustrate the differences.

4b Two examples of big-data modeling

Machine translation belongs to the standard repertoire of big-data success stories. It illustrates particularly well the shift from complex models with relatively scarce data to simple models with a lot of data. Although somewhat of an oversimplification, two different approaches can be distinguished (Halevy et al. 2009). The *rule-based* approach models the complex hierarchy of grammatical rules of both languages and translates sentences by using a conventional dictionary. The *data-driven* or statistical approach largely neglects the grammatical structure and works instead with huge corpora of texts in combination with Bayesian inferential statistics. Usually, there will be monolingual corpora, e.g. in English and a foreign language, and bilingual corpora containing sample translations, all of them representative of current speech practice. The frequencies of words and word sequences in these corpora can be used to calculate the most probable translation of a foreign word sequence f into English e using Bayes' rule[12]: argmax_e P(e) P(f|e) / P(f).

The data-driven approach has been strikingly successful. Apparently, probability distributions of words and word sequences yield reasonable results for many tasks such as spellchecking or translation, while grammatical knowledge is largely dispensable. Two quotes from practitioners well illustrate this remarkable situation. Peter Norvig, who for a long time headed Google's machine translation group, once stated that they have been able 'to build models for languages that nobody on the team speaks'[13]. Frederick Jelinek, a pioneering and by now legendary figure in the field, is often quoted with saying that 'every time I fire a linguist, the performance of the speech recognizer goes up'[14].

It is not immediately obvious how machine translation fits the notion of causal modeling from Section 3, since it does not involve the physical necessity of empirical causal laws. However, the logic of necessary and sufficient conditions works just as well for the 'conventional necessity' of translation rules. The context of a word, i.e. all other words at various distances, eventually determines a sufficient condition for the best translation, resulting in an almost infinite number of complex and highly context-specific translation rules. It may well be the case that simple and general rules entirely fail to exist, i.e. that general rules have so many and so diverse exceptions that these cannot be listed. Under such circumstances, the data-driven horizontal modeling may be the only option available.

The second example comes from the social sciences regarding the use of data-intensive methods in American elections, in particular Barack Obama's 2008 and 2012 bids for presidential office. Political campaigning is a typical big-data problem as depicted in Section

[12] Cp. 'The Unreasonable Effectiveness of Data', talk given by Peter Norvig at UBC, 23.9.2010. http://www.youtube.com/watch?v=yvDCzhbjYWs at 38:00.
[13] Ibd. 43:45.
[14] http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/speechreco/team/, accessed 1.8.2013

3a. Voters are characterized in terms of hundreds or thousands of features x ranging from demographic data like age, race or gender to political opinions gathered in surveys to consumer data provided for example by credit card companies. Campaign managers are interested in causal relationships between these predictors and outcome variables like commitment to vote or allegiance to a certain candidate. The approach has been aptly called *microtargeting* (Issenberg 2012).

In the United States, abundant training data exists because citizens are often willing to volunteer information about their voting habits. The resulting models are developed algorithmically with little input of political expert knowledge. They are used to determine the probabilities that certain persons can be convinced to vote for a specific candidate and which means are most appropriate in terms of political message and medium, e.g. contact by mail, telephone or a personal visit. While previously, political campaigns addressed larger groups of people characterized by just a few parameters such as middle-class Caucasian male, microtargeting focuses on individual voters characterized by hundreds or thousands of variables. This allows correcting many implicit assumptions about the relevance of traditional variables like race, gender or class, essentially redrawing the conceptual boundaries between groups on an empirical basis. Indeed, big-data science is especially suited for the categorization and classification of phenomena in view of a specific purpose.

Many applications of big data in the social sciences have a structure that is completely analogous to microtargeting. Individuals are characterized in terms of a large number of parameters with a specific aim in mind, e.g. to find an appropriate search result or to make someone click a certain link or buy a certain product.

4c Characteristics of horizontal modeling

We are now in a position to list features that distinguish big-data modeling from more conventional scientific modeling. Due to the limited capacity for data storage and processing, the latter is geared at an efficient data reduction and an adequate structuring of knowledge resulting in a hierarchy of laws of increasing generality. Big-data modeling has a different nature due to the ability to handle enormous amounts of data. The hierarchical structuring becomes dispensable for prediction and manipulation, hence the term *horizontal modeling* for the big-data approach:

i) Predictions in horizontal modeling are made *directly from the data without taking recourse to a simplifying model*. Consequently, the laws can be very complex involving a large number of parameters and be highly context-specific with often only a small number of instantiations. Usually, the number of laws will dwarf that in conventional scientific modeling.

ii) Since the data already represents all relevant configurations of the phenomenon, there is little need to introduce abstract levels of description, the main purpose of which was a clever reduction of the data while preserving most of the information content. Big-data models thus *lack the hierarchical, nested structure* that is characteristic of most conventional science.

iii) Relatedly, *the explanatory power of horizontal models is much smaller* than that of hierarchical models. After all, models become more explanatory according to most accounts of scientific explanation the more pronounced the hierarchical structure is with each new level of laws or rules constituting a new level of explanation. Consequently, the horizontal models provide little understanding, e.g. the understanding of a language is poor without knowledge of the grammatical structure. This will be further elaborated in Section 5.

iv) *Idealizations and simplifications play only a minor role* in horizontal modeling compared with the hierarchical approach, since these are usually introduced to link different levels of generality. However, crucial modeling assumptions in the horizontal approach result from choices which data to collect and how to analyze it.

v) Traditionally, models have focused on one or several aspects of a phenomenon while leaving out others, i.e. information is reduced in view of a certain purpose. This *perspectival character is less prominent* in horizontal modeling, again since there is no incentive for data reduction.

With respect to the debate on models in philosophy of science, horizontal modeling differs from conventional models of data (Suppes 1962, Mayo 1996) in several ways. Few restrictions are imposed on the extent and kind of data that is gathered and analyzed ('messy data'). Also, there are fewer modeling assumptions guiding the formulation and analysis of the data. Finally, 'no' data is discarded in the modeling process. These differences can be motivated by the fact that conventional models of data work with parametric statistics while horizontal modeling employs non-parametric statistics (cp. Section 6).

Another relevant comparison concerns causal modeling, in particular in terms of causal graphs (e.g. Pearl 2000). In this respect, note that causal relations established by the difference-making account of Section 3b cannot fully be represented in terms of simple graphs, in particular when it comes to complicated Boolean expressions for causal dependencies. Most importantly, the presence of a causal link between parameters often depends on the chosen causal background.

5. Big data's lack of explanatory power

We can now clear up the confusion concerning the causal and explanatory nature of big-data modeling, essentially resulting from a conflation of causation, mechanisms, and explanation. On the one hand, big-data models aim to identify causal relations because they are geared at prediction and manipulation of phenomena. On the other hand, big-data models lack considerable explanatory virtues. This is often pointed out, for example in the following representative quote: 'The correlations [found in big-data-science] may not tell us precisely *why* something is happening, but they alert us *that* it is happening.' (Mayer-Schönberger & Cukier 2013, 21) To understand how big-data-models can be causal but fail to be explanatory, various notions of explanation have to be carefully distinguished as discussed in the philosophy-of-science literature.[15]

[15] Excellent introductions are Psillos (2002) or Woodward (2011).

A good starting point is the following distinction: (i) to explain by giving an argument that derives what is to be explained from a number of general laws thereby relating a phenomenon to other phenomena and achieving unification. For example, the motion of a ball can be explained by Newton's axioms in combination with the law of gravity; (ii) to explain by citing the causes that made something happen as in 'this ink stain is explained by Henry dropping his pen yesterday afternoon'.[16]

In the first case, general laws are explanatory, explanations have the structure of arguments, and they aim at unification. In the second case, causes are explanatory, explanations consist in lists of causes, and they aim at causal stories. Since big-data science is about causal modeling but fails to produce a hierarchical structure, big-data models yield explanation in the second sense but largely fail to be explanatory in the first sense.

The distinction is manifest in the philosophical debate on the nature of explanation. The classic deductive-nomological model by Carl Hempel and Paul Oppenheim (1948) is a typical example for the first category of explanatory accounts. It construes explanations as arguments deriving the explanandum from a combination of general laws plus a number of boundary and initial conditions. For example, the dynamics of a pendulum can be explained by the laws of a harmonic oscillator plus certain initial conditions. The law of the oscillator in turn is explained by Newton's axioms and Galileo's law of gravity in combination with certain approximations regarding the amplitude of the pendulum. Newton's axioms in turn can be explained by the equations of general relativity given a certain limit. Moving up the hierarchy, further levels of explanation are introduced by showing how a phenomenon fits into a framework of increasingly general laws. Every additional level furnishes unificationist explanation of the level beneath by pointing out similarities and analogies to other classes of phenomena. Such explanatory depth is typical for hierarchical modeling. By contrast, horizontal big-data models fare badly on this type of explanation, exactly because they largely lack a hierarchical structure of increasingly general laws.[17]

Another example of the first category of explanation is the unificationist account developed mainly by Michael Friedman and Philip Kitcher, which directly emphasizes the unifying nature. In Friedman's words: 'this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than with more.' (1974, 15) Philip Kitcher develops the idea further to account for explanation in terms of argument patterns (1989). Harmonic oscillators or Newton's axioms are examples of such patterns. Obviously, in its reductionist spirit the unificationist account is quite similar to the hypothetico-deductive account. Both want to derive an observation or phenomenon from a small number of sufficiently general laws, only the former focuses on argument structure while the latter on the premises.

[16] Note that some overlap can exist between both kinds of explanation, in particular if the causal laws are sufficiently general.

[17] The pendulum has been treated from a big-data perspective (Schmidt & Lipson 2009), deriving the laws of motion from a video covering all relevant states of the pendulum.

The second type of explanatory accounts, causal explanation, can of course only be understood with a notion of causation already in place. Various explications are discussed in the literature, for example statistical relevance or counterfactual accounts. For this article, we will stick with the difference-making approach sketched in Section 3. As should be obvious from the discussion there, big-data models are perfectly suited to yield causal explanations, to tell causal stories why something happened. Note that causal explanation occasionally may not achieve unification at all. After all, the method of difference can identify singular causation.

Let me conclude the section with two remarks: First, scientific understanding usually presupposes unificationist explanation. A phenomenon is understood by linking it to other phenomena and reducing the number of independent assumptions. The smaller this number and the larger the range of phenomena covered, the deeper the perceived understanding of a topic. Obviously, this notion of understanding is intimately tied to hierarchical modeling and to the conscious human cognitive apparatus with its limitations in data storage and processing.

Second, phenomena in the complex sciences, e.g. the social sciences, may not be accessible to unificationist explanation at all, because of an absence of laws of significant generality. Consequently, human understanding of these phenomena may always be considerably impaired. This implies reconsidering the role of human experts in the complex sciences as they are traditionally thought to guide the research process by providing understanding. Big-data science often functions without understanding, as some of the pertinent examples show: machine translation without knowledge of grammar, advertising without classical advertising knowledge, campaigning without in-depth political-science knowledge.

6. Science without equations: The paradigm shift in statistics

The notion of a qualitative change from hierarchical to horizontal modeling is further corroborated by a concurrent paradigm shift in statistics, which has been described as a transition from *parametric* to *non-parametric* modeling (e.g. Russell & Norvig 2010, Ch. 18.8), from data to algorithmic models (Breiman 2001), or from model-based to model-free approaches. Since the shift concerns methodology and not theoretical or empirical content, it differs in important ways from scientific revolutions. Nevertheless, the statistics community currently experiences some of the social ramifications and ‘culture clashes’ that are typical for scientific paradigm shifts as documented for example in Breiman (2001) or in Norvig’s dispute with Noam Chomsky on data-driven machine translation (Norvig 2011).

Several features distinguish parametric from non-parametric modeling, which were in part already mentioned in the comparison between hierarchical and horizontal modeling (cp. Section 4): i) Parametric methods usually presuppose considerable modeling assumptions. In particular, they summarize the data in terms of a ‘small’ number of model parameters specifying for example a Gaussian distribution or linear dependence, hence the name. By contrast, non-parametric modeling presupposes *few modeling assumptions*, e.g. allows for a wide range of functional dependencies or of distribution functions. ii) In non-parametric modeling, predictions are calculated *on the basis of all data*. There is no detour over a parametric model that summarizes the data in terms of a few parameters. iii) While this

renders non-parametric modeling quite *flexible* with the ability to quickly react to unexpected data, it also becomes extremely *data- and calculation-intensive*. This aspect accounts for the fact that non-parametric modeling is a relatively recent phenomenon in scientific method.

Non-parametric models allow for novel ways to deal with complexity: iv) A crucial shift occurs from *equation modeling* to *algorithmic modeling*. Conventional parametric modeling in terms of equations, describing for example functional dependencies or distribution functions, already presupposes that the picture has been reduced to a small number of parameters and to relatively simple functional relationships. By contrast, algorithmic modeling does not have such restrictions. It relies less on sophisticated mathematics and more on a brute-force execution of a large number of steps, when for example an algorithm searches a large data-base for similar cases. *Algorithmic models consist of the original data and an algorithmic procedure to derive predictions from the data*.

v) As discussed in Section 5, the complexity of non-parametric models prevents a deeper understanding of the phenomena. Thus, there is a *shift in epistemic values* regarding the aims of modeling. Non-parametric modeling is geared almost exclusively at prediction and manipulation and rarely at understanding in terms of general laws or rules. By contrast, parametric modeling usually emphasizes understanding. While parametric modeling often correlates with a *realist* and *reductionist* viewpoint, non-parametric modeling has *instrumentalist* and *pluralist* connotations. The instrumentalist attitude is for example apparent in the wide-spread use of ensemble methods that combine different models even if these start from mutually contradictory assumptions. Presumably, this shift in epistemic values is at the root of the mentioned divide between the different 'cultures' of statistical modeling.

Let me illustrate these changes by means of two examples, first the comparison between parametric and non-parametric regression and second between parametric and non-parametric density estimation. In a parametric univariate linear regression problem, one has reasonable grounds to suspect that a number of given data points $(x_i;y_i)$ can be summarized in terms of a linear dependency: $y = ax + b$. Thus, two parameters need to be determined, offset b and slope a, which are usually chosen such that the sum of the squared deviations $\sum_{i=1}^{n}(y_i - (ax_i + b))^2$ is minimized.

In non-parametric regression, the data is not summarized in terms of a small number of parameters a and b, but rather all data is kept and used for predictions (Russell & Norvig 2009, Ch. 18.8.4). A simple non-parametric procedure is *connect-the-dots*. Somewhat more sophisticated is locally weighted regression, in which a regression problem has to be solved for every query point x_q. The y_q-value is determined as $y_q = a_q x_q + b_q$ with the two parameters fixed by minimizing $\sum_{i=1}^{n} K(d(x_q, x_i))(y_i - (a_q x_i + b_q))^2$. Here, K denotes a so-called kernel function that specifies the weight of the different x_i depending on the distance to the query point x_q in terms of a distance function $d()$. Of course, an x_i should be given more weight the closer it is to the query point.

The generalization to higher dimensions is straight-forward though for next-neighbor methods an important issue arises that has been termed the 'curse of dimensionality'. With increasing

number of dimensions, i.e. of predictor variables, the average distance between neighboring points rapidly becomes very large of order $(1/N)^{1/n}$, where N is the total number of points and n the number of dimensions. Consequently, the data points will almost always be sparsely distributed in many dimensions.[18]

Let us briefly reflect how these regression methods illustrate the differences between parametric and non-parametric modeling i) to v). While in parametric regression, linear dependency is presupposed as a modeling assumption, the non-parametric method can adapt to arbitrary dependencies. In parametric regression, the nature of the functional relationship has to be independently justified by reference to a theoretical context, which prevents an automation of the modeling process. Certainly, non-parametric regression also makes modeling assumptions, e.g. a suitable kernel function must be chosen that avoids both over- and underfitting. However, within reasonable bounds the kernel function can be chosen by cross-validation. Since often, predictions turn out relatively stable with respect to different choices of kernel functions, an automation of non-parametric modeling remains feasible.

While non-parametric regression is more flexible than parametric regression, it is also much more data-intensive and requires more calculation power. Notably, in the parametric case, a regression problem must be solved only once. Then all predictions can be calculated from the resulting parametric model. In the non-parametric case, a regression problem must be solved for every query point. In principle, each prediction takes recourse to all the data. While the parametric model consists in a relatively simple mathematical equation, the non-parametric model consists in all the data and an algorithmic procedure for making predictions.

Consider density estimation as a second example (Russell & Norvig 2009, Ch. 20.2.6). The parametric approach makes an explicit assumption about the nature of the distribution function, for example a Gaussian distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. This distribution is determined by two parameters, the mean μ and the standard deviation σ, which are chosen such that a best fit with the data is achieved.

A simple non-parametric approach is the histogram method, where the parameter space is partitioned into cells of equal volume dV and the number k_i of all N data points is counted for each cell i. The density is given by $f(x) = k_i / N\, dV$, where k_i is the number of data points in the same cell as the query point x. A closely related often more effective non-parametric method is k-nearest-neighbors, where the same formula is used but k is now fixed and one determines the minimal volume dV surrounding the query point x such that k points are included. The parameter k should be chosen in a way to avoid overfitting, but still be sufficiently sensitive. A suitable value can again be chosen by cross-validation allowing for straight-forward automation of the non-parametric approach.

Again, in the parametric case the data is summarized in terms of a model characterized by a few parameters μ and σ resulting in a simple formula, while the non-parametric method makes no assumptions about the nature of the distribution function and is thus much more

[18] Note that this curse of dimensionality does not automatically apply to all big-data algorithms. To the contrary, it occasionally turns out helpful to artificially increase the dimensionality of the parameter space in methods like decision trees or support vector machines (Breiman 2001, 208-209).

flexible. On the other hand, the non-parametric method is very data-intensive since it uses the original data points to make predictions. The difference between the two types of models is striking: While parametric models usually are simple equations, the non-parametric models consist in the original data plus an algorithm to derive predictions from the data.

7. Conclusion: The new science of complexity

The horizontal modeling based on non-parametric statistics will in the coming years greatly extend the causal knowledge in the complex sciences. Opportunities lie for example in medicine and epidemiology when dealing with complex diseases like allergies, asthma, and cancer or in ecology when trying to understand complex processes like the recent worldwide decline in bee populations. Presumably, more effective ways of management will become possible through big data in both economics and politics. However, there are also considerable dangers concerning potential abuses especially in the social sciences, where most of the large data sets are currently collected.

The knowledge established by big-data methods will consist in a large number of causal laws that generally involve numerous parameters and that are highly context-specific, i.e. instantiated only in a small number of cases. The complexity of these laws and the lack of a hierarchy into which they could be integrated prevent a deeper understanding, while allowing for predictions and interventions. Almost certainly, we will experience the rise of entire sciences that cannot leave the computers and do not fit into textbooks.

Acknowledgments

I acknowledge helpful discussions with Mathias Frisch as well as comments and criticism from audiences at Enschede, Munich, and Berlin. The research was carried out within the Poiesis project, generously funded by the BMW Foundation Herbert Quandt, the Gerda Henkel Foundation and the Institute for Public Knowledge of New York University. I am grateful for much inspiration from the other fellows.

References

Anderson, Chris. 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *WIRED Magazine* 16/07. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Bacon, Francis. 1620/1994. *Novum Organum*. Chicago, Il: Open Court.

Beisbart, Claus. 2012. “How can Computer Simulations Produce new Knowledge?” European Journal for Philosophy of Science 2:395-434.

Beisbart, Claus & John Norton, 2012. “Why Monte Carlo Simulations are Inferences and not Experiments.” *International Studies in Philosophy of Science* 26:403–422.

Breiman, Leo. 2001. “Statistical Modeling: The Two Cultures.” *Statistical Science* 16(3):199-231.

Callebaut, Werner. 2012. “Scientific perspectivism: A philosopher of science’s response to the challenge of big data biology.” *Studies in History and Philosophy of Biological and Biomedical Science* 43(1):69-80.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Friedman, Michael. 1974. “Explanation and Scientific Understanding.” *Journal of Philosophy* 71:5-19.

Frigg, Roman & Julian Reiss. 2009. “The Philosophy of Simulation: Hot New Issues or Same Old Stew?” *Synthese* 169(3):593–613.

Gramelsberger, Gabriele. 2010. *Computerexperimente. Zum Wandel der Wissenschaft im Zeitalter des Computers*. Bielefeld: Transcript.

Gray, Jim. 2007. “Jim Gray on eScience: A Transformed Scientific Method.” In Tony Hey, Stewart Tansley & Kristin Tolle (eds.). *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf

Halevy, Alon, Peter Norvig & Fernando Pereira. 2009. “The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems* 24(2):8-12. http://www.csee.wvu.edu/~gidoretto/courses/2011-fall-cp/reading/TheUnreasonable%20EffectivenessofData_IEEE_IS2009.pdf

Hartmann, Stephan, 1996. “The World as a Process: Simulations in the Natural and Social Sciences,” in R. Hegselmann, et al. (eds.), *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, Dordrecht: Kluwer, 77–100.

Hempel, Carl & Paul Oppenheim. 1948. “Studies in the Logic of Explanation.” *Philosophy of Science* 15:135-175.

Hey, Tony, Stewart Tansley & Kristin Tolle. 2009. *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research.

Humphreys, Paul. 2004. *Extending Ourselves. Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press.

Issenberg, Sasha. 2012. *The Victory Lab: The Secret Science of Winning Campaigns*. New York, NY: Crown.

Kitcher, Philip. 1989. “Explanatory Unification and the Causal Structure of the World.” In Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation*, 410–505. Minneapolis: University of Minnesota Press.

Laney, Doug. 2001. “3D Data Management: Controlling Data Volume, Velocity, and Variety.” Research Report. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Leonelli, Sabina (ed.). 2012a. Data-driven Research in the Biological and Biomedical Sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1).

Leonelli, Sabina. 2012b. “Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies.” *International Studies in the Philosophy of Science* 26(1):47-65.

Mackie, John L. 1965. “Causes and Conditions.” *American Philosophical Quarterly* 12:245-265.

Mayer-Schönberger, Viktor & Kenneth Cukier. 2013. *Big Data*. London: John Murray.

Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.
Mill, John S. 1886. *System of Logic*. London: Longmans, Green & Co.
Mitchell, Sandra. 2008. *Komplexitäten. Warum wir erst anfangen, die Welt zu verstehen.* Frankfurt a.M.: Suhrkamp.
Norvig, Peter. 2011. "On Chomsky and the Two Cultures of Statistical Learning." http://norvig.com/chomsky.html
Norvig, Peter. 2009. "All we want are the facts, ma'am." http://norvig.com/fact-check.html
Pearl, Judea. 2000. *Causality. Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.
Pietsch, Wolfgang. 2013. "The Nature of Causal Evidence Based on Eliminative Induction." In P. Illari and F. Russo (eds.), forthcoming special issue of *Topoi*. http://www.wolfgangpietsch.de/pietsch-eliminative_induction.pdf
Psillos, Stathis. 2002. *Causation and Explanation.* Durham: Acumen.
Rohrlich, Fritz. 1991. "Computer Simulation in the Physical Sciences." In Arthur Fine, Mickey Forbes, and Linda Wessels (eds.), *PSA 1990, Vol. 2*. East Lansing: Philosophy of Science Association, 507-518.
Russell, Stuart & Peter Norvig. 2009. *Artificial Intelligence.* Upper Saddle River, NJ: Pearson.
Schmidt, Michael & Hod Lipson. 2009. "Distilling Free-Form Natural Laws from Experimental Data." *Science* 324(5923):81-85.
Siegel, Eric. 2013. *Predictive Analytics.* Hoboken, NJ: John Wiley & Sons.
Spirtes, Peter, Clark Glymour & Richard Scheines. 2000. *Causation, Prediction and Search.* Cambridge, MA: M.I.T. Press.
Suppes, Patrick. 1962. "Models of Data", in Ernest Nagel, Patrick Suppes and Alfred Tarski (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford: Stanford University Press, 252–261.
Winsberg, Eric. 2010. *Science in the Age of Computer Simulation.* Chicago, Il: University of Chicago Press.
Winsberg, Eric. 2013. "Computer Simulations in Science", *The Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/archives/sum2013/entries/simulations-science/
Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
Woodward, James. 2011. "Scientific Explanation." *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/