ON THE BOOK SERIES
VERA BÜHLMANN, LUDGER HOVESTADT

Only one hundred years ago, hardly any scientist of renown would have been unaware of philosophy, and hardly any artist or architect uninformed about up-to-date technology and mathematics. Today, our ability to explain and explicate our own work within a shared horizon of assumptions and values beyond our specific scientific community has, perhaps paradoxically, turned into an inability and resulted to some degree in a kind of speechlessness. Only rarely now is it thought important that we relate our work to, and integrate it with, an overall context that is in itself "on the table" and up for consideration. More and more, that kind of context is taken for granted, without any need for active articulation, refinement, or development. At the same time though, the media are full of news stories about catastrophes, crises, and an impending doom that cannot, it seems, be warded off. Climate change, shortage of resources and population growth, urbanization—this is just to name a few of the critical issues today. Quite obviously, the notion of such an overall context, both implicit and assumed, is extremely strained, if not indeed overstretched today. This all is widely acknowledged—the United Nations Educational, Scientific and Cultural Organization (UNESCO) Division of Foresight, Philosophy and Human Sciences in Paris, for example, launched a discourse on this subject in their 21st-Century Talks and Dialogues under the heading "The Future of Values." The companion book, published in several languages simultaneously in 2004, is structured in three parts, and includes one chapter on the ethical issues of values and nihilism lying ahead, another chapter on technological progress and globalization, as well as a third chapter on the future of science, knowledge, and future studies. What remains strangely implicit, and in that manner ignored here, in a way that is typical of this inarticulacy with regard to an overall context mentioned above, is the societal, scientific, and cultural role that inevitably is ascribed to technology against the backdrop of such discussions, along with the expectations that are associated with that role of technology. In the Metalithikum series, we tend to regard technology in the extended

sense of technics at large. Along with its respective solution-oriented application to the sciences, culture, economics, and politics, we think that technology needs to be considered more fundamentally, especially regarding the semiotic and mathematical-philosophical aspects it incorporates. From this perspective, we see in technology a common factor for facilitating a discourse that seems to have been largely lost from today's discursive landscape, the degree of its disappearance inversely proportional to the increasingly central role technology plays in every domain of our lives. Such a discourse seems crucial if we are to develop adequate schemes for thinking through the potentials of today's technology, something that is in turn essential for all planning. Our stance is an architectural and, in the philosophical sense, an architectonic one. Our main interest centers on the potentials of information technology, and how we can get used to the utterly changed infrastructures they have brought us.

But have our infrastructures really changed substantially? Or is it merely the case that a new level of media networks has emerged on top of technology with which we are already familiar? Are the "new" and digital media simply populating and exploiting, in a parasitic sense, the capacities of modern industrial infrastructures that have brought prosperity and wealth to so many? In his contribution to the UNESCO dialogues, Paul Kennedy was still convinced: "In the Arabic world, 3% of the population has access to the internet. In Africa, it's less than 1%. This situation won't improve as long as the infrastructures remain in their current state. It won't change, as long as these countries lack electrification, telephone wiring and telephones, and as long as the people there can't afford either computers or software. If knowledge is indeed power, then the developing countries today are more powerless than they were thirty years ago, before the advent of the internet." Our experience since then has allowed us to see things a little differently. There are meanwhile as many mobile phones in use worldwide as there are people living on the planet. Six billion people out of a seven billion world population can meanwhile read, write, and calculate (at least in some basic sense). Only three decades ago, this proportional measure was not 6/7th, but 2/5th! We have seen the "Arab Spring" that brought simultaneous political revolutions in several Arabic countries, giving facticity to the cultural impact of digital media, and this to a degree that was unexpected or previously deemed improbable by many. And the credence of this facticity is not harmed, we think, by the fact that since then we have had to witness ongoing fundamentalist reactions as in Syria, where the situation is currently escalating into a veritable civil war. To say that the facticity of the cultural impact of digital media is not impaired in its credence thereby is not to downplay the seriousness of these complex situations. Of course, in political, economic, religious reorientation all is at once at stake, and the idea that technological

III PRE-SPECIFIC MODELING: COMPU-TATIONAL MACHINES IN A COEXISTENCE WITH CONCRETE UNIVERSALS AND DATA STREAMS
VAHID MOOSAVI

VAHID MOOSAVI is a PhD candidate at the the Chair for Computer Aided Architectural Design (CAAD), Swiss Federal Institute of Technology (ETH Zurich), and has been working at the Future Cities Laboratory at Singapore-ETH Centre (SEC) in Singapore since 2012. In 2009 he got his master of science degree in Industrial and Systems Engineering from Amirkabir University of Technology (Tehran Polytechnic). He has worked as a consultant in the field of enterprise planning and transformation for several private and public organizations like the Ministry of Health and the Ministry of Energy in Iran, the Industrial Management Institute of Iran (IMI), and several manufacturing companies. His scientific research interest has always centered around abstract and mathematical modeling and design approaches. In his PhD research, he develops a generic computational modeling framework for urban and architectural design problems, based on Self-Organizing Maps (SOM) and Markov Models as core technologies in coexistence with Urban Data Streams. Some of the research applications he is now focused on are as follows: modeling urban traffic dynamics using GPS data streams, data-driven urban air pollution modeling, modeling the dynamics of real estate markets, spatial function approximation, and multi-variate thematic mapping. www.vahidmoosavi.com.

This article criticizes a majority of traditional approaches in scientific modeling as following an idealist point of view that results from the set theoretical understanding of models based on the assumption of "abstract universals." We discuss the success of set theoretical models as well as some principle limits to applying them in dealing with complex systems. As an alternative to the assumption of abstract universals, we propose a conceptual modeling framework based on "concrete universals" that can be interpreted as a category theoretical approach to modeling. We call this modeling framework pre-specific modeling. We show how a certain group of mathematical and computational methods that work with data streams are able to

operationalize the concept of pre-specific modeling with great benefit and promise.

I   HOW TO APPROACH THE NOTION OF SCIENTIFIC MODELING

Modeling paradigms, as a necessary element of any scientific investigation, act like pairs of glasses that impact the way in which we encode (conceive of) the real world. Therefore, any kind of intervention in real world phenomena is affected by the chosen modeling paradigm and the real phenomena under investigation. In the domain of urbanism and urban design, cities as complex and open environments with dynamic and multidimensional aspects are challenging cases for modeling scholars, as there are many distinct urban phenomena. Figure 1 shows a list of different functional aspects of urban phenomena in an indexical manner.

FIG. 1.
Different functional aspects of urban phenomena.

In addition to the diversity of urban problems, there is a huge variety of competing paradigms for analyzing cities: the city as an ecological phenomenon that is optimally adjusted to an environment (economic, political, cultural) assumed to be "natural" for it; the city as a thermodynamic system that needs to be balanced and that can be controlled; the city as a grammatical text with its own "syntactic laws"; the city as a biological organism following fractal growth patterns. Further, historical perspectives provide additional city models such as the City of Faith, the City as Machine, or the Organic City,[1] and especially, since the advent of computers from the second half of the twentieth century, city as information.[2]

1   See Kevin Lynch, *Good City Form* (Cambridge, MA: MIT Press, 1984).
2   See Manuel Castells, *The Informational City: Information Technology, Economic Restructuring, and the Urban-Regional Process* (Oxford: Blackwell, 1989), 15.

Compared to classical science and its engineering disciplines such as physics, chemistry, and mechanics, urban design, planning, and modeling is a rather young discipline; yet when one does a quick search of the keywords central to this field, one is quickly confused by the number of approaches and the variety of practical problems within the reaches of the discipline. For example, A. G. Wilson's five-volume text on urban modeling is over 2,600 pages long.[3] A broad range of case-based canonization has thus emerged, and applied techniques are developed for specific urban functions such as urban land use, urban transportation, urban economy, urban social patterns, and so on. As a result, the lack of a more abstract categorization of applied techniques makes comparison between them very hard.

Beginning in the mid-twentieth century, general systems theory emerged as one of the main theories for working toward unification of different disciplinary modeling practices.[4] In principle, the underlying idea of systems theory is the promotion of a unified view to modernist-reductionist science, which was diversified around a variety of application and functional domains. Although interdisciplinary collaborations such as making analogies within disciplines (e.g., hydraulic theories to describe biological systems) was not new, systems theory's formalization, as an orthogonal view to classically diversified scientific and practical problems, reached a point in which, according to George Klir, systemic tasks such as modeling, optimization, and simulation emerged as distinct scientific disciplines.[5] However, taking systems theory as a body of knowledge (rather than a specific and singular theory), one could expect a gradual divergence of its methods, starting from its unified principles. The advent of computational methods by Alan Turing in the 1940s and later the democratization of computational methods in the 1980s created a new diversified landscape of system modeling approaches. As a result, after fifty years we encounter a competitive ecosystem of different modeling species with different capacities and trade-offs. Figure 2 shows a list of different modeling methodologies in an indexical manner.

FIG. 2.
Competitive ecosystem of different modeling methodologies.
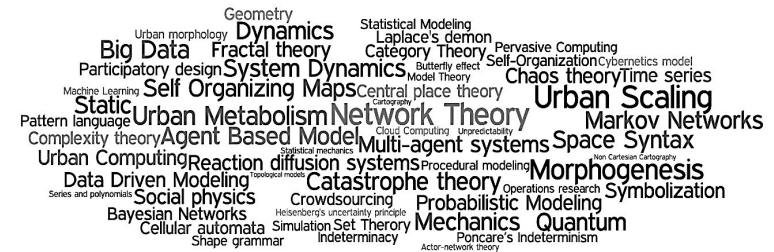
3   A. G. Wilson, *Urban Modelling: Critical Concepts in Urban Studies*, vols. 1–5 (London: Routledge, 2012).
4   See Ludwig von Bertalanffy, *General Systems Theory* (New York: George Braziller, 1993).
5   See George J. Klir, *Architecture of Systems Complexity* (New York: Saunders, 1985).

Therefore, the first motivation of this present essay is to find a unifying (abstract) perspective for an assessment of different (interdisciplinary) modeling approaches, while keeping the diversities. Toward this aim, we need to investigate the mathematical and philosophical grounds of scientific modeling.

II   FORMAL DEFINITIONS AND CATEGORIES OF SCIENTIFIC MODELING

Because there is such a wide variety of modeling approaches in different scientific domains, formalizing and theorizing the practice of scientific modeling is an active research area in the philosophy of science. For example, according to Roman Frigg and Stephen Hartmann, there exist the following types of models:

> Probing models, phenomenological models, computational models, developmental models, explanatory models, impoverished models, testing models, idealized models, theoretical models, scale models, heuristic models, caricature models, didactic models, fantasy models, toy models, imaginary models, mathematical models, substitute models, iconic models, formal models, analogue models and instrumental models are but some of the notions that are used to categorize models.[6]

Nevertheless, these categories are not still abstract enough, but rather labels for different (not necessarily exclusive) modeling approaches.

To better understand models, one can look at the interpretation of their roles and functions, and to distinguish the presets on which the different points of view are based. One of the main issues by which models have been extensively discussed is the relation between models and the way of representation of real phenomena under study (the target system). According to Frigg and Hartmann, from a representational point of view there are "models of phenomena" and "models of data,"[7] and within these categories there are subcategories such as "scale models,"[8] "idealized models,"[9] and "analogical models,"[10] like the hydraulic model of an economic system, which are further divided into material analogy, where there is a direct similarity between the properties (or relations between properties) of two phenomena, and formal analogy, where two systems are based similarly on a formalization such as having the same

mathematical equations that describe both systems[11]. Further, one can refer to phenomenological models, which are focused on the behavior of the particular phenomena under investigation rather than on underlying causes and mechanisms.[12] Further still, there are models of theories, like the dynamic model of the pendulum, which are based on laws of motion. Models can also be divided into ontological classes like physical objects, fictional objects, set-theoretic structures, descriptions, and equations. However, these categories of models and modeling approaches overlap and they are descriptive and neutral classifications rather than critical. *They do not give us a measure or a gauge to compare different modeling approaches in terms of their capacities and their limits in dealing with different levels of complexity in real world problems. In this essay I am looking for a way to condition modeling approaches in different levels of complexity to examine their theoretical capacities.*

Among the abovementioned categories, the crucial but somewhat commonly accepted shared property of the majority of traditional scientific modeling approaches is that they are all based on some sort of idealization. In what follows, I explain different aspects of idealization in scientific modeling. The issue of idealizations directs us to the problem of universals, which is an old philosophical issue.[13]

III   IDEALIZATION IN SCIENTIFIC MODELING

In the context of philosophy of science, idealization in modeling has been discussed extensively.[14] In principle, idealization is considered to be equal to an intended (over-)simplification in the representation of the target system. Although there are different ways of explaining or defining the notion of idealization, Michael Weisberg discusses three kinds of idealization that we refer to in this work: minimalist idealization, Galilean idealization, and multiple-model idealization.[15]

Minimalist idealization is the practice of building models of real world phenomena by focusing only on main causal factors. Therefore, as is inferred from its name, minimalist models usually result in very simple elements that are informative enough for further decision making. For example, the aim in the domain of network analytics is to explain complex behaviors that happen in the real phenomena by means of network properties such as centrality measures, integration, closeness,

6   Roman Frigg and Stephan Hartmann, "Models in Science," in *Stanford Encyclopedia of Philosophy,* ed. Edward N. Zalta (Fall 2012), ed. Edward N. Zalta, http://plato.stanford.edu/archives/fall2012/entries/models-science/.
7   Ibid.
8   Max Black, *Models and Metaphors: Studies in Language and Philosophy* (Ithaca, NY: Cornell University Press, 1962).
9   Ernan McMullin, "Galilean Idealization," *Studies in History and Philosophy of Science Part A* 16, no. 3 (1985): 47–73.
10   Mary B. Hesse, *Models and Analogies in Science* (London: Sheed and Ward, 1963).

11   Ibid.
12   Ernan McMullin, "What Do Physical Models Tell Us?," *Studies in Logic and the Foundations of Mathematics* 52 (1968): 385–96.
13   See Gonzalo Rodriguez-Pereyra, "What Is the Problem of Universals?," *Mind* 109, no. 434 (2000): 255–73.
14   See McMullin, "Galilean Idealization"; Leszek Nowak, "Laws of Science, Theories, Measurement," *Philosophy of Science* 34 (1972): 533–48; Michael Weisberg, "Three Kinds of Idealization," *Journal of Philosophy* (2007): 639–59.
15   Weisberg, "Three Kinds of Idealization."

between-ness, etc.[16] As an example, in urban theory, in the city science approach[17] or urban scaling laws[18] the final goal is to find a few main informative factors in cities such as city size or population in order to explain other aspects of cities such as energy consumption in a linear equation. Even though it seems obvious that cities are complex phenomena with many observable aspects and many exceptions, minimalist models attract attention exactly because they identify and state very general rules.

FIG. 3.
Network analytics:
Structure-oriented modeling (minimalist idealization), Central Place Theory (left) and Space Syntax (right).
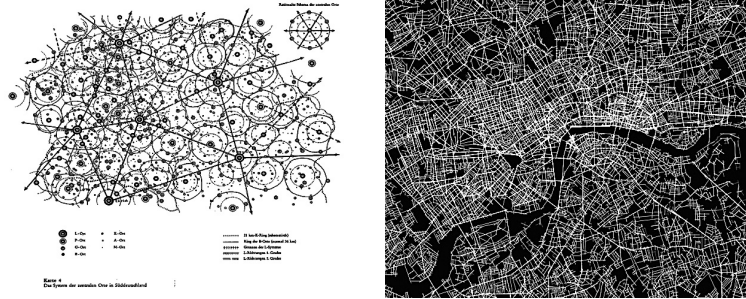
City theories that seek to create archetypal city models are in a way using minimalist idealized models. For example, Lynch's City of Faith, City of Machine, or City as Organism, or Cedric Price's egg analogies of the city (city as boiled egg, city as fried egg, or city as scrambled egg) are characterized by a few urban elements that are informative enough to explain each model and to discriminate that city model from the other models. David Grahame Shane shows how the three abovementioned models could be identified by linear combinations of three recombinant elements, called Enclave, Armature, and Heterotopia.[19]

The second category of Galilean idealization as the most pragmatic type of idealizations happens when the modeler intentionally simplifies the conditions of a complicated situation toward more computational tractability and simplicity. For example, it is common in economic models to assume that agents are rational maximizers, or in transportation models to assume that commuters take the shortest path, or to assume there is no friction in motion models of the particles. The basic idea of Galilean idealization is that by understanding the modeling environment gradually, it is possible to de-idealize or to build more

16 See Mark Newman, *Networks: An Introduction* (Oxford: Oxford University Press, 2010).
17 See Luís M. A. Bettencourt et al., "Growth, Innovation, Scaling, and the Pace of Life in Cities," *Proceedings of the National Academy of Sciences* 104, no. 17 (2007): 7301–06.
18 See Michael Batty, "The Size, Scale, and Shape of Cities," *Science* 319, no. 5864 (February 8, 2008): 769–71.
19 David Grahame Shane, *Recombinant Urbanism: Conceptual Modeling in Architecture, Urban Design, and City Theory* (Chichester: Wiley, 2005).

comprehensive models on top of previous ones. Therefore, the majority of engineering approximation methods such as systems of differential equations or computational fluid dynamics or biological reaction networks are among this category of idealized models. Further, figure 4 shows how the idealization process in a complex phenomenon (here, the agent-based modeling of land-use transportation dynamics of a city) leads to a parametric and feature-based representation of the real phenomena. This layering and parameterization gives the modeler the option to adjust the resolution (levels of details) of the model based on the needs and the purposes of the modeling process and the constraints and limitations, including the availability of data or prior knowledge or time and scale resolutions.
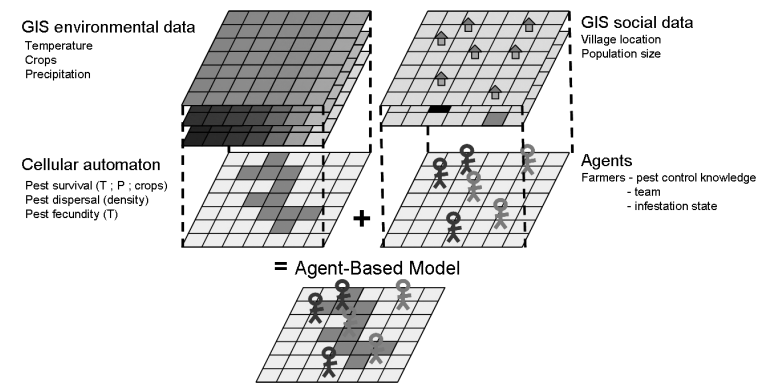

FIG. 4.
Parametricism: Idealization of the interactions between different agencies through layering and parameterization of the real phenomena.

The third category of idealization, multiple-model idealization, results from those models that consist of several (not necessarily compatible) models or several models with different assumptions and different properties. This type of idealization is in fact a combination of two other idealizations and it can be very useful when understanding the final output (the behavior) of the model is more important than knowing the underlying mechanisms of the target phenomena. For example, in weather forecasting, ensemble models, which include several predictors with different parameters or even different structures, are used to predict weather conditions.[20]

Further, from a systemic and functional point of view there are many models in which idealization is happening in (one) main aspect(s) of real phenomena. To just name a few: static or dynamic models, structure-oriented idealization (in network models), process-oriented idealization

20 See Tilmann Gneiting and Adrian E. Raftery, "Weather Forecasting with Ensemble Methods," *Science* 310, no. 5746 (2005): 248–49.

(such as system dynamics,[21] system of differential equations), rule-based idealization (such as cellular automata[22] or fractals[23]), and decentralized interactions (such as agent based), all are placed in the abovementioned categories of idealizations.

FIG. 5.
System dynamics: process-oriented idealization.
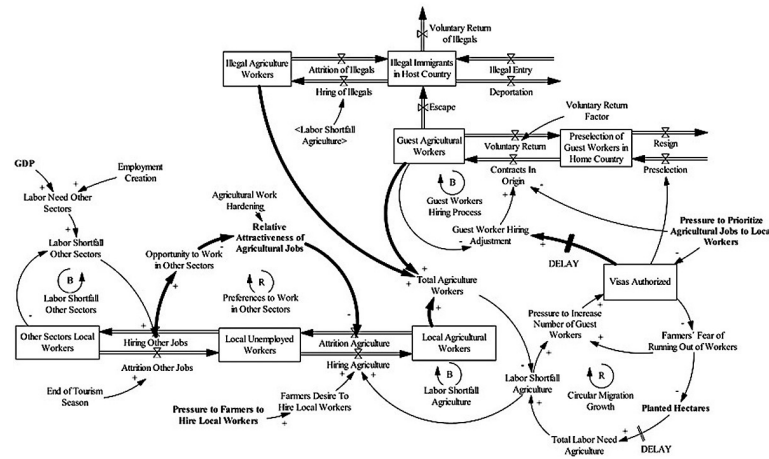
However, considering the size and the variety of parameters and aspects in the target phenomena, idealized models create a dichotomy, where on one extreme the models are all general, simple, and tractable, and on the other, models become complicated, specific, and high-resolution. In fact, multiple model idealization becomes necessary whenever the selected parameters and aspects of the target system in each individual model (out of Galilean idealization for example) are not sufficient, but also add more aspects to an individual model, either making it more complicated or resulting in model inconsistency. This issue seems to be a never-ending debate in many scientific fields, including biology, ecology, economics, and cognitive and social science, where one group believes in the explanatory power of models and the other group believes in model accuracy and the level of details comparing to the real phenomena.[24]

21   John Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World* (New York: McGraw Hill, 2000)
22   Waldo Tobler, "Cellular Geography," in *Philosophy in Geography*, ed. Stephen Gale and Gunnar Olsson (Boston: D. Reidel, 1979), 379–86.
23   Benoit Mandelbrot, *The Fractal Geometry of Nature* (New York: W. H. Freeman & Co., 1982).
24   See Matthew R. Evans et al., "Do Simple Models Leads to Generality in Ecology?," *Trends in Ecology & Evolution* 28, no. 10 (2013): 578–83; Matthew R. Evans et al., "Data Availability and Model Complexity, Generality, and Utility: A Reply to Lonergan," *Trends in Ecology & Evolution* 29, no. 6 (2014); Mike Lonergan, "Data Availability Constraints Model Complexity, Generality, and Utility: A Response to Evans et al.," *Trends in Ecology & Evolution* 29, no. 6 (2014).

Although idealized models have been applied successfully in many classical modeling problems, this type of debate cannot be fruitful in dealing with complex systems as long as there is no abstraction from the current paradigm of scientific modeling (i.e. idealization). Analogically, an onion-like model of numbers explains what I mean by the abstraction in the concept of modeling. For example, with natural numbers (or more generally, integers) one can never grasp the richness of proportions and fractions in rational numbers (e.g., 2.6, which is neither 2 or 3 from a natural number perspective), while the introduction to the concept of rational numbers as the ratio of two integer numbers (e.g., 26/10) solved this problem. Therefore, by choosing 1 as the denominator, one can show that all the integers are rational numbers; while with rational numbers we have new capacities in addition to integers. Similarly, if we take an idealized model as an arbitrary representation of real phenomena by adding several of them together (which is the case in multiple model idealization), we still cannot grasp the whole complexity. Therefore, our hypothesis is that an abstraction to the concept of modeling is needed in order to conceptually encapsulate all the potential arbitrary views in an implicit way.

However, I do not claim that one can introduce a new concept as such, but in fact in this work I am trying to identify and discover new aspects of a potential body of thinking in scientific modeling.

In order to highlight this conceptual abstraction from the current idealization paradigm, first we need to explain the notion of universals, including abstract and concrete universals, followed by our interpretations of these concepts in relation to the notion of scientific modeling.

In the next section, after presenting the connections between the notions of idealization and abstract universals, I will formally describe the concepts of abstract universals and concrete universals, which can be interpreted as set theoretical and category theoretical definitions of these two notions.[25] Further, I will show how the concept of concrete universals from category theory can open up a new level of modeling paradigm.

IV   UNIVERSALS AND MODELING

In the majority of texts written about idealization in the domain of scientific modeling, the notion of idealization is equal to simplification and the elimination of empirical details and deviations from a general theory that is the base for the final model. At the same time, the word "ideal" literally comes along with "those perfections that cannot be fully realized." For example, circle-ness as a property is an ideal that cannot be fully realized, and any empirical circular shape has, to a degree, the circle-ness property.

25   David P. Ellerman, "Category Theory and Concrete Universals," *Erkenntnis* 28, no. 3 (1988): 409–29.

FIG. 6.
Enso (circle of Zen):
Toward the ideal
circle.

Therefore, the idealization process in scientific modeling can be explained as a form of purification of empirical observations toward a set of given (assumed) ideal properties. In statistical data analysis, it is always assumed that collected empirical data follows a normal distribution function. Thus, one can convert the empirical data to a normal distribution function and utilize it in a "normalized" manner that results from the machinery of this ideal mathematical representation (i.e. the normal distribution function). Applications of idealizations in many mathematical approaches such as linear algebra are enormous. For example, a Fourier transformation (figure 7) can be seen as a form of idealization by which any observed time-varying data can be reconstructed (approximately) by a set of time-varying vectors (a set of pure sinusoidal waves with different frequencies and phases). From this perspective, any waveform phenomenon is a linear combination of a set of ideal prototypes.

FIG. 7.
Fourier decomposi-
tion: Any observed
form is a linear
combination of some
ideal cyclic form.
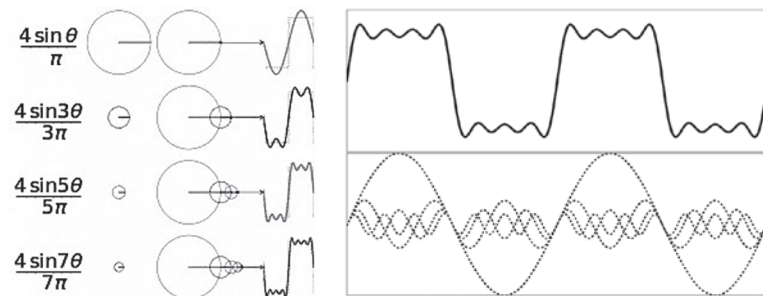
However, these ideal forms (a wave with a certain frequency in the case of the Fourier analysis) as the set of aspects (properties) of real phenomena are abstract. This means that there is no concrete (empirical) instance that fully matches one or several of these a priori, ideal properties. From this point of view, idealized models are models that are based on the notion of abstract universals.

The notions of "universals" and "property" are old topics in philosophy that can be approached differently, namely through realism, idealism, or nominalism.[26] However, in this work I focus on the distinctions between concrete and abstract universals in relation to the paradigms of scientific modeling.

According to David Ellerman, "In Plato's Theory of Ideas or Forms ($\varepsilon\iota\delta\eta$), a property F has an entity associated with it, the universal u_F, which uniquely represents the property. Therefore, an object X has the property F i.e. $F(X)$, if and only if it participates in the universal u_F to a degree (μ)."[27] For example, "whiteness" is a universal and the set of white objects that participate in "whiteness" property (i.e., with different degrees of whiteness) are represented by this property. Further, "Given a relation μ, an entity u_F is said to be a universal for the property F (with respect to μ) if it satisfies the following universality condition: For any x, x μ u_F if and only if $F(x)$."[28]

This condition is called universality, and it means that the universal is the essence of that property.

In addition to universality, a universal should be unique. "Hence there should be an equivalence relation ($\approx$) so that universals satisfy a uniqueness condition: If u_F and u_F' are universals for the same $_F$, then $u_F \approx u_F'$."[29] Therefore, any entity that satisfies the conditions of universality and uniqueness for a certain property is a universal for that property. Now, if a universal is self-participating, it is called a concrete universal; if it does not have self-participatory properties, it is an abstract universal. For example, whiteness is an abstract universal as there is no empirical (concrete instance) to be "whiteness." In language models, being a "verb" is a property that can be assigned to many words, but "verb" itself is an external definition and it is not self-participating in the sets of concrete verbs. The same argument goes for the above example of the Fourier analysis and ideal forms.

On the other hand, defining a property as being part of set A and set B has a concrete universal, which is the intersection of two sets A and B ($A \cap B$). It means that any object from set A and B (including all the potential subsets) that has this property (being part of A and B) participates in the intersection set $A \cap B$, and since $A \cap B$ is participating in itself, then it is a concrete universal.

Further, Ellerman shows that modern set theory is the language of abstract universals and how category theory can be developed as the mathematical machinery of concrete universals. Finally, he summarizes:

> Category theory as the theory of concrete universals has a different flavor from set theory, the theory of abstract universals. Given

26 See Rodriguez-Pereyra, "What Is the Problem of Universals?"
27 Ellerman, "Category Theory and Concrete Universals," 410.
28 Ibid.
29 Ibid., 411.

the collection of all the elements with a property, set theory can postulate a more abstract entity, the set of those elements, to be the universal. But category theory cannot postulate its universals because those universals are concrete. Category theory must find its universals, if at all, among the entities with the property.[30]

In the past few decades there have been many theoretical works to further the new field of category theory in terms of this fundamental difference between set theory and category theory. For example, currently the main categorical approaches in mathematics are topos theory and sheaf theory, which are generalizations of topology and geometry to an algebraic level.[31] It seems that applications of these general frameworks in different domains should be one of the main future research areas in the field of modeling. On the other hand, Ellerman concludes:

> Topos theory is important in its own right as a generalization of set theory, but it does not exclusively capture category theory's foundational relevance. Concrete universals do not "generalize" abstract universals, so as the theory of concrete universals, category theory does not try to generalize set theory, the theory of abstract universals. Category theory presents the theory of the other type of universals, the self-participating or concrete universals.[32]

Now that we have defined the concepts of abstract and concrete universals, we need to formalize two different approaches of modeling, which are based on these notions of the universal.

As stated earlier, idealized models are models that are based on the notion of abstract universals and consequently idealized models can be interpreted as set theoretical models. In the next section, by focusing on the idea of representation in idealized models, I show their theoretical consequences and their limits in dealing with complex systems, with the definition of the abstract universal being crucial. Next, I show another conceptual representational framework that is matched with the concept of concrete universals. Further, I will introduce an alternative line of modeling to idealized modeling.

V   SPECIFIC MODELING: MODELS BASED ON ABSTRACT UNIVERSALS

The fundamental difference between abstract and concrete universals is the issue of self-participation. In terms of modeling and representation, in those models based on abstract universals, the definition of the common property of the target system is a priori, given in a meta-level. This means that in an empirical setup, we have an externally given

30   David Ellerman, "On Concrete Universals: A Modern Treatment Using Category Theory"; available online at SSRN, http://ssrn.com/abstract=2435459, here p. 6.
31   See Saunders Mac Lane and Ieke Moerdijk, *Sheaves in Geometry and Logic: A First Introduction to Topos Theory* (New York: Springer, 1992).
32   Ellerman, "Category Theory and Concrete Universals," 16–17.

idea about the set of properties (aspects) of the real phenomena under study at the beginning of the modeling process. As an example, if we are comparing many concrete objects (e.g., several apples), we first need to define a set of specific properties (such as size, color, taste, etc.) to construct a representation of apple-ness. Therefore, apple-ness is reduced to this external setup. We call this approach specific modeling as it is based on a set of specific properties of the target system. In relation to the idealization process, the level of details in terms of the number and variety of properties is the choice of the modeler. If the modeler considers few aspects of the target system the model becomes simple and if he or she selects many aspects or properties, the model becomes complicated.

Figure 8 shows the concept of idealized representation in specific modeling schematically. Each circle in this figure stands for a concrete object. These objects are symbolic, which means that they can stand for anything—be it people, cars, companies, buildings, streets, neighborhoods, cities, websites, protein networks, networks of words in a corpus of texts, or people and their activities in a social network. Therefore, in the first step, we need to define our abstract universals, which leads to a set of selected features of the real objects. These features are shown by rectangles. As a result of these universal features, the concrete instances of the object are assumed to be independent from each other, as they will be all compared indirectly by an abstract class definition, which acts as an external reference.
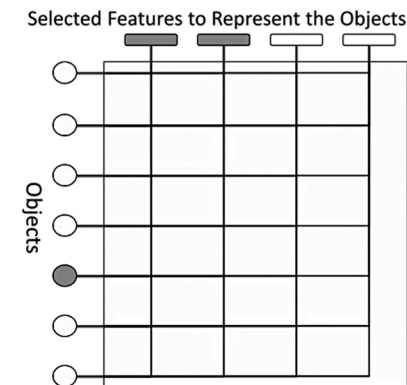

Selected Features to Represent the Objects

Objects

This is the underlying notion of rationality started in sixteenth century by René Descartes and it should be mentioned that it offers a fantastic mechanism and an abstract language for axiomatization of different phenomena. Nevertheless, there are fundamental limits to this approach of modeling in dealing with complex phenomena, with many different properties, where the specific models need to define an arbitrary set of properties.

Within the literature of scientific modeling, the majority of discussions on the issues of scientific modeling approaches are bounded to models based on abstract universals and the differences of different idealization processes. Among the few investigations, Richard Shillcock discusses the fundamental problems of modeling in the domain of cognitive science from the perspective of universals. He notes: "Cognitive science depends on abstractions made from the complex reality of human behaviour. Cognitive scientists typically wish the abstractions in their theories to be universals, but seldom attend to the ontology of universals."[33] Later he explains several fundamental problems in the domain of cognitive science by reviewing the different aspects of abstract and concrete universals. In what follows I present some of the fundamental issues of the models that are based on abstract universals.

V.I.I   GÖDEL'S INCOMPLETENESS THEOREM AND ARBITRARINESS OF MODELS BASED ON ABSTRACT UNIVERSALS

In models based on abstract universals, the universal properties are not self-participating. Intuitively, one can argue that in any level of abstraction, members of a set are concrete and the set itself is abstract with regard to its members. Therefore, the first modeling step is the decision about the set of properties that define (represent) the object of inquiry. To have a set of concrete instances (e.g., set of red apples), one needs a super-set that defines the ideal properties of that class (the apple-ness and the red-ness). This requirement (brought forward by Plato) initiates a never-ending hierarchical process of defining abstract universals for the higher order classes (e.g., a set for colors). As a result, one can argue that in practical modeling domains, from a meta-level above, models are based on assumed or commonly agreed properties of the target system. This problem can be explained by Gödel's incompleteness theorem, that is to say we only can make a consistent system if it is based on an unproved truth (the incomplete model); if the model is complete (everything based on proofs), it cannot be consistent.[34] This beautiful theorem simply says that any model that is based on abstract universals is in a way arbitrarily consistent, but not simultaneously complete. The same argument holds for the case of Russell's paradoxes and naive set theory.[35]

---

33   Richard Shillcock, "The Concrete Universal and Cognitive Science," *Axiomathes* 24, no. 1 (2014): 63–80, here p. 63.

34   Panu Raatikainen, "Gödel's Incompleteness Theorems," *Stanford Encyclopedia of Philosophy,* ed. Edward N. Zalta (Winter 2014), http://plato.stanford.edu/archives/win2014/entries/goedel-incompleteness/.

35   See Ellerman, "Category Theory and Concrete Universals."

V.I.II   CURSE OF DIMENSIONALITY IN COMPLEX SYSTEMS

Models based on abstract universals have been successfully applied in many practical domains such as classical physics, medicine, and engineering. Nevertheless, they reach a computational limit in dealing with complex systems. This limit is directly related to their quest for explicit representation of the target systems through a set of specific properties. Assume that we measure the complexity of a system (i.e., a real phenomenon) as a function of the number of its potential properties and the relations between those properties.[36] In this scenario, in comparison to a building a wooden chair is less complex, and the same relation holds for a building in comparison to a city.

As a result, by increasing the number of potential properties and their interrelationships, and consequently the exponential growth in the number of combinations, the space of modeling (i.e., potential specific models) expands in an exponential manner. This phenomenon is called the curse of dimensionality introduced by Richard Bellman in 1961.[37]

Consequently, in a complex system, any endeavor toward an explicit representation (which is the case in specific modeling) leads either to a complicated model (models with lots of redundancy and lack of explanation) or to very simple and minimalistic idealizations. Figure 9 shows this issue diagrammatically.
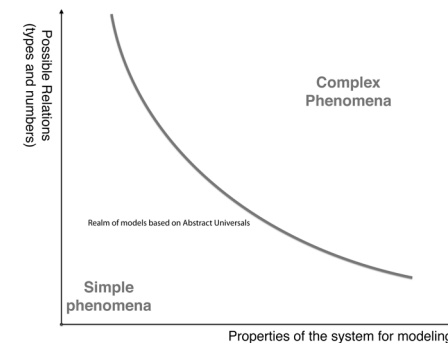
FIG. 9.
The curse of dimensionality and idealized models based on abstract universals.

### V.I.III   FROM PARTICULAR TO GENERIC AND THE CONCEPT OF "ERROR"

In the idealization process of particular objects there is no longer a unique identity dedicated to a particular (concrete) instance, but rather the identity of that particular case is realized as a combination of globally defined properties (see figure 8). In other words, in models based on

---

36   See Klaus Wassermann, "That Centre-Point Thing—The Theory Model in Model Theory," in *Printed Physics—Metalithikum I,* ed. Vera Bühlmann and Ludger Hovestadt (Vienna: Ambra, 2013), 156–88.

37   Richard Bellman, *Adaptive Control Processes: A Guided Tour* (Princeton, NJ: Princeton University Press, 1961).

abstract universals the particular object is considered as an instance of a (fictitious) generic object. Along the same line, Shillcock says: "The concrete universal is a universal, but it has all the richness of the particular. Whereas an abstract universal can be defined as something abstract (typically seen as a property) that inheres in many other different things, a concrete universal is an entity in which many other different things inhere."[38] Consequently, constructing the notion of a generic object through the lens of abstract universalism, we impose a limit to empirical deviations and treat them as errors. For example, assuming linearity as an ideal property in a system, the generic object has a purely linear behavior while other objects are erroneous (they deviate from the line) to a degree. Figure 10 shows this issue in the case of linear regression. In a two-dimensional linear system, it is assumed that for any observation (i.e., a concrete object via its *x* and *y* dimensions) there is a linear relation in the form of *y=ax + c*. Therefore, those points that don't fall in a common line have a degree of error in comparison with the ideal line.

FIG. 10.
Introduction to the concept of error: The deviation of particular objects from the ideal line.
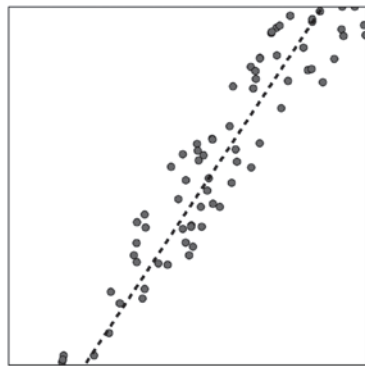
VI   PRE-SPECIFIC MODELING: MODELS BASED ON CONCRETE UNIVERSALS

In this section, we investigate the potentials for a new level of abstractions in paradigms of scientific modeling. This is not in opposition to specific modeling (i.e., models based on abstract universals), which is a common approach in social science and the humanities, rather it is based on the notion of concrete universals from category theory. As discussed in the previous section, models that we call specific models are based on a priori defined or selected abstract universals. Further, they have certain theoretical limits and issues in dealing with complex systems. Therefore, my hypothesis is that if any specific model is like an arbitrary view to the real

38   Shillcock, "The Concrete Universal and Cognitive Science," 71.

phenomena, there should be a category of models that encapsulates all the potential specifics in an implicit way. We call this approach pre-specific modeling, which was originally introduced by Vera Bühlmann.[39]

If specific modeling can be theorized by set theory and abstract universals, pre-specific modeling should be supported by the concepts of category theory and concrete universals. In order to establish the building blocks of pre-specific modeling, we need to focus on fundamental assumptions of the specific modeling.

VI.I   DEDEKIND CUT: WHEN A PARTICULAR OBJECT IS REPRESENTED BY THE NEGATION OF ITS COMPLEMENT

In specific modeling, when one defines the abstract universal in terms of a set of specific properties, a parametrical generic object will be conceptualized directly. The individual objects can then be reconstructed (analyzed) or generated (synthesized) by changing the values of those specific parameters in the generic object. As a fundamental example, we refer to number theory and the definition of rational numbers as the ratio of two integers *m* and *n*, where *n* is not equal to 0. In this case, any specific rational number, *q*, can be directly represented by infinite pairs of (*m, n*) integer values, where *m/n = q*. In other words, *q* is graspable directly and independently from the other rational numbers. However, as we know, the rational numbers are countable and they are only a small fraction of the whole space of the real numbers. As a result, this approach reaches a point in some cases where no one can define a real number on its own.
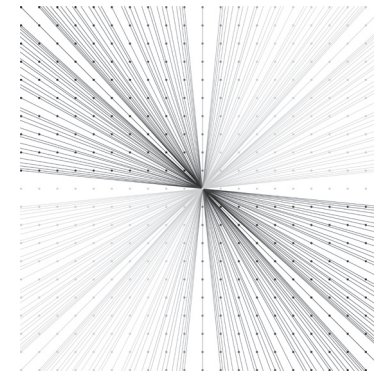

FIG. 11.
Rational numbers cannot fill the space of real numbers. Each line corresponds to one rational number.

For example, $\sqrt{2}$ cannot be touched by the abovementioned procedure. In general, this is the case for irrational numbers. A different method for the definition of irrational numbers is required. In the late nineteenth century, Richard Dedekind came up with a different

39   Vera Bühlmann and Martin Wiedmer, eds., *Pre-Specifics: Some Comparatistic Investigations on Research in Art and Design* (Zurich: JRP|Ringier, 2008).

conceptual definition of irrational numbers, known as the Dedekind cut.[40] Intuitively, a Dedekind cut is a unique way of representing an irrational number by its complementary set. He defined a cut for a specific number, *b*, as the space between two ordered sets of rational numbers *A* and *B*, where all the elements of *A* are less than all the elements of *B* and further all the elements of *A* are smaller than *b* and all the elements of *B* are equal or greater than *b*. Definitely, if *b* is a rational number, the union of set *A* and *B* is the whole number space of real numbers, *U*, and if *b* is an irrational number, *b* is equal to *U* minus the union of sets *A* and *B* *(A∪B)*. For example, in order to define √2, *A* is the collection of all negative rational numbers and the collection of every nonnegative rational number whose square is less than 2 and *B* is the collection of all positive rational numbers whose square is larger than 2.
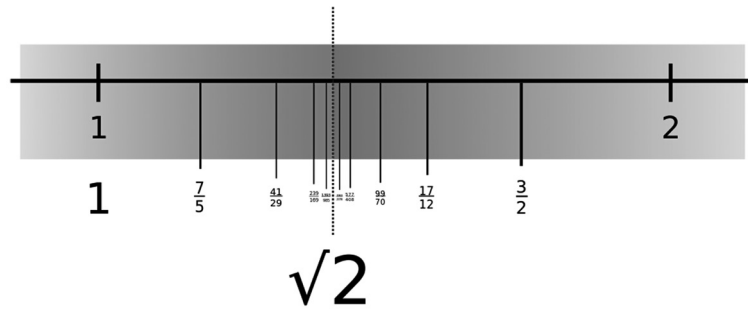
Further, two irrational numbers can be compared through their corresponding cuts—if their cuts are equal, these numbers are identical. By this definition each specific irrational number is represented uniquely as the negation of its complement, while it is not directly touchable. This is opposite to that of a rational number, where it can be directly pointed out. We think that regarding the issue of object representation, the Dedekind cut importantly implies that it is possible to introduce an alternative approach for representation of the objects to what is common in specific modeling (shown in figure 8).[41]

VI.II   FROM GENERIC TO PARTICULAR: OBJECT-DEPENDENT REPRESENTATION

The core aspect about any modeling paradigm is how real phenomena are represented. As figure 8 shows, by selecting the set of representational properties of the real phenomena in specific modeling, each

individual object is represented directly. In other words, the identity of a particular object is defined independently of the other (concrete) objects as long as we have a global axiomatic set up (i.e., those selected properties) to define the generic object. Here, the generic object is the abstract universal, in which with different parametric values one can instantiate or approximate a particular object. Referring to the example of number theory, this is similar to the case of rational numbers, where a specific number can be generated as the ratio of two integers.

Now, imagine an empirical representation of concrete objects in network-based representation in which nodes of connectivity, which are to specify in multidimensional ways, represent objects. For example, the number of cars that pass from one street to another, or the relation established by two individuals who select the same restaurant, or the relation between two cities that host offices of the same company, or the number of times a specific word has appeared after another specific word. In distinction to parametric representation of objects, the identity of an object is defined directly in terms of the relations it maintains with the other objects. The main difference between the two approaches is that in the feature (property) based approach, the specific identity of objects is assumed independently, while in the network-based representation, the identity of objects is regarded as pre-specific, and is specified purely relationally, out of the connectivity, which is observable. Two objects are considered identical if they share the same sets of relations with other objects.

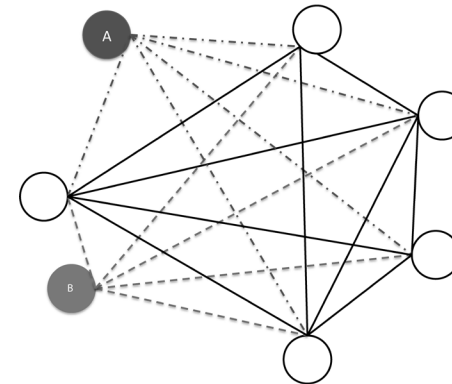Figure 13 shows the representation of concrete objects in a network-based approach.

In specific modeling, each property has an abstract universal, but in object—object relations, each concrete object is a property (for example A-ness for concrete object A) and thus we have concrete universals for each property, as each object has an identity relation with itself.

---

40   Richard Dedekind, "The Nature and Meaning of Numbers," in *Essays on the Theory of Numbers*, trans. W. W. Beman (Chicago: Open Court Publishing, 1901 [1898]).

41   For more detailed discussions around the conceptual and philosophical underlying ideas of the Dedekind approach, refer to what Ludger Hovestadt mentions as centered voids in Vera Bühlmann and Ludger Hovestadt, ed., *EigenArchitecture* (Vienna: Ambra, 2014).

Assuming each concrete object as a feature, we can represent each object via its relation to the other objects. In comparison with the definition of irrational numbers by the Dedekind cut, here too the identity of a particular concrete object is defined as the negation of the identity of the other particulars. Note that here there objects that are not yet defined—unlike the case of parametric object representation.

This setup, as shown in figure 13, is an object-dependent representation that is conceptually scalable with the size of empirical objects. While in specific modeling the size of parameters is independent to the number of concrete objects (i.e., observations), in object-dependent representation, by adding one concrete object we directly add one new aspect for the representation of other objects. This aspect makes pre-specific modeling suitable for working with large amounts of data. Further, in many areas today, the conditions for these types of representation hold, as we have an emergent network of connected instances that can be used for the representation of the object of inquiry.

In section 8 I present two main technical frameworks that support the concept of pre-specific modeling—two main applications of object-dependent representation. However, as I mentioned before implicitly, object-dependent representation and pre-specific modeling in general are data driven, as the setup shown in figure 13 is based on concrete objects. The role of data in pre-specific modeling is different than classical empirical research when one assumes to have an a priori generic object. As pre-specific modeling is proposing a new modeling framework it demands another notion of data, one that is different than traditionally designed observations and measurements.

VII   MASSIVE UNSTRUCTURED DATA STREAMS: AN INVERSION IN THE NOTION OF MEASUREMENTS AND DATA PROCESSING

In classical scientific modeling, theories and a priori representations define what should be measured and observed. According to Bas C. van Fraassen, a measurement outcome is always achieved relative to particular experimental setup designed by the user and characterized by his theory.[42] Similarly, as we showed in the case of specific modeling, by selecting abstract universals we limit the set of potential observable aspects of real phenomena. For example, when dealing with a pendulum model and using Newtonian laws of gravity as a theoretical model to describe the foundation of the motions of particles, data and measurements can only empirically validate or propose minor modifications. Therefore, classical data has always played a marginal role in the process of modeling.

In addition to this conceptual setup, measurement and observation have, historically, been very expensive, and this pushed modelers toward more structured, designed, and optimized experiments and observations. Taking into consideration data in specific modeling, figure 14 shows the classical process of modeling. As shown in the diagram, abstract universals (or the definition of the generic objects) are always the first and the primary element of modeling processes; the data—including its structure (i.e., the selected properties of the real environment) and its size (to be statistically enough)—has a supporting role in model tuning and model validation. This diagram shows that since the data is the secondary element, after a certain level of observation the model quality (in terms of accuracy for example) becomes stable, as we have enough data to tune the system.
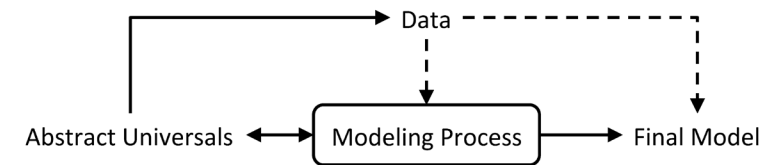
FIG. 14
The classical modeling process (specific modeling).

Nevertheless, considering computational technologies as the dominant factors in shaping and directing the area of scientific modeling of the last century, the landscape of measurements and data processing has been changing dramatically. In a recent article, I discuss three levels of computational capacities, known as computing power, computational and communicational networks, and data streams.[43] The first level deals with computing power in terms of numerical simulations in comparison with analytical approaches. Historically, there have been different technologies of computation starting with mainframes, moving to the democratization of computing through personal computers and microcomputers, which are still getting faster and more powerful at an exponential rate. The primary function of computing power is numerical simulation, even though computers have been isolated or with limited communication abilities. Although computers and their simulation power opened up new possibilities for better understanding of the real world phenomena in 1960s and 1970s in many fields, for a while during the late 1970s, these computational models got data hungry and their demand for data was higher than what was available for model tuning and validation. This produced some skepticism about the applications of computational models to real world problems.[44]

---

42   Bas C. van Fraassen, "Scientific Representation: Paradoxes of Perspective," *Analysis* 70, no. 3 (2010): 511–14.

43   Vahid Moosavi, "Computational Urban Modeling: From Mainframes to Data Streams," presented at the AI for Cities Workshop, AAAI Conference, Austin, Texas, 2015.

44   See Douglas Lee, "A Requiem for Large Scale Modeling," *Journal of the American Institute of Planners* 39, no. 3 (1973): 163–78.

However, alongside the developments within computing technologies, advancements in communication technologies gradually opened up another capacity for modelers, which can be considered as the second level of computational capacities. In this level, where computing power was not scarce, it was the communication between computing systems that turned important. Therefore, new phenomena such as networks of sensors, mobile phones or computers, and the Internet started to emerge. Gradually, considering the amount of embedded systems in many real world applications, computers as computing machines became the ground to introduce new functions that were emerging on top of computational networks.

As a byproduct of these networks of computing and communicating machines, the amount of digital data started to increase as well. Starting from the mid-1990s, technical terms such as "data mining" and "database management" emerged in parallel to a focus on methodology to explore digital data (mainly structured data) among modelers. As can be seen, data began to emerge around this period, but this data was a byproduct of designed measurements and sensory systems. It is important to note that by this time the notion of data had not changed from its old notion—collected data was still structured and followed by the modeler's choices. In fact, the data still was the secondary element, rationally determined by the given properties of the target system. However, what had changed dramatically was the amount of digitally collected data. It started to grow quantitatively on top of the communicating and computational networks across disciplines.

Finally the third level, for which we think we have a suitable notion of data for pre-specific modeling, emerged only recently. With rapid advancements both on the level of computing power and the networks of computing systems, and a rapid growth in social media, we have encountered a new stage in which on top of ubiquitous computing and communicating systems, a new level of abstract phenomenon has started to emerge. We have begun to experience exponential growth in the amount of information available, together with the mobile computing devices most people use on a daily basis. This is often called a data deluge. Next to the challenges these changes bring, we can also see how new areas for research and practice are emerging.[45]

It seems clear today that the classic paradigm of observation and data gathering has changed radically. Data is produced on an everyday basis,

45   To just mention a few recent publications: ; Adam Greenfield, *Everyware: The Dawning Age of Ubiquitous Computing* (Berkeley, CA: New Riders, 2006); Uwe Hansmann et al., *Pervasive Computing: The Mobile World* (Berlin: Springer, 2003); Nathan Eagle and Alex Penxtland, "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing 10, no. 4 (2006): 255–68*; Eric Paulos, R. J. Honicky, and Ben Hooker, "Citizen Science: Enabling Participatory Urbanism," in *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*, ed. Marcus Foth (Hershey, PA: Information Science Reference, IGI Global, 2008).

from nearly any activity we engage in, and accumulates from innumerable sources and formats such as text, image, GPS tracks, mobile phone traces, and many other social activities, into huge streams of information in digital code. These unstructured and continuous flows, which can be called urban data streams, can be considered as a new infrastructure within human societies.

This notion of data is opposed to its classical notion, where data was produced mainly as the result of designed experiments to support specific hypothetical models or when data was transmitted via defined semantic protocols between several inter-operating software. These new data streams are the raw materials for further investigations; and similar to computing power, they hold new capacities for modeling. As a result of this new plateau, we are challenged to learn new ways to grasp this new richness.

These massive, unstructured urban data streams induce an inversion in the paradigm of modeling from specific modeling, and they match the concepts of pre-specific modeling and models based on the concept of concrete universals.

Therefore, as an alternative to previous modeling process, pre-specific modeling is mainly based on the coexistence of unstructured data streams representing particular objects and self-referential representation of concrete universals (figure 15). As shown in the next section, opposite to a specific modeling paradigm, where data has a limited use in modeling process and after a certain level of data size, the performance of models become stable in this level of data-driven modeling and object-dependent representation. Adding more data will improve the quality of the final model. This is the power of concrete universals that are scalable with data size, unlike parametric models, in which after a certain size of data, the parametric state space of the generic object becomes full and reduces the discrimination power of the model in dealing with a variety of circumstances in complex systems. By using concrete universals, each new instance will introduce a new aspect in relation to the other concrete instances. Therefore, it is scalable with data size.
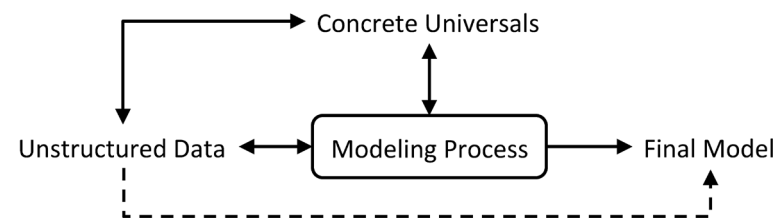

FIG. 15
Pre-specific modeling in coexistence with unstructured data and concrete universals.

In the next section, I present a certain category of mathematical and computational methods that support the notion of pre-specific modeling.

Even though I've presented several examples of pre-specific modeling, so far the explanation has remained on a conceptual level. Here I present a more technical discussion of two computational methods that fit very well with the concept of pre-specific modeling.

VIII.I   MARKOV CHAINS

Andrei Markov is among the greatest mathematicians of the twentieth century; he has made numerous contributions toward forming probability theory, but his major work is the concept of Markov chains, which he introduced in 1906. In engineering and applied scientific domains today, many people know Markov chains as a kind of memoryless dynamic model: once we have a sequence of random variables (x_1, x_2, x_3, … , x_{t-1}, x_t), the state of the system at the next step (x_t+1) depends only on the previously observed state (x_t). These processes create a chain of random activities, where there is a probabilistic link between adjacent nodes. Here, we assume the case of discrete time processes with a finite number of states, but in principle one can assume to have continuous time and continuous state space. Further, the chain is called homogeneous if the conditional distributions of x_t+1 given x_t were independent of time steps. And, assuming more sequential dependency, one can construct highe-order chains from the state at step t dependent upon its n previous steps, if the order of the chain is n. Nevertheless, in the case of first-order chains, assuming the mutual dependencies between any two potential states, a Markov chain can be represented with a directed graphical network, where each node corresponds to a state and the edges between two nodes correspond to two conditional probabilities. Figure 16 shows an example of a Markov chain with three states.
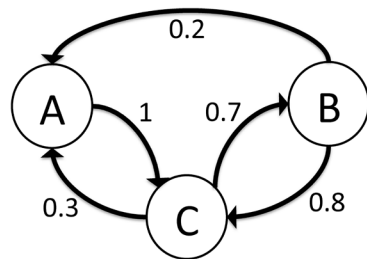
FIG. 16
Traditional
representation of
a Markov chain
in modeling of a
dynamic system.

In the domain of dynamic systems, Markov chains and their many different versions have been studied and applied in diverse applications for the simulation of dynamic systems or for the study of steady state

conditions. They have also been applied successfully to sequence and time series prediction and classification.

Markov's brilliant idea for the representation of a complex phenomenon such as natural language in a purely computational manner is particularly relevant here. Before going into his approach, let us unpack the traditional concept of language models. One of the major and to some degree dominant concepts of linguistic models is based on the notion of abstract universals. In this approach to modeling, which is in accord with Noam Chomsky's,[46] a spoken language can be modeled by means of a set of semantic and syntactic laws of the specific language. Therefore, writing and speaking correctly by an individual means that there is a system of production in his or her mind that produces the instances of that language following his or her ideal model. This is one of the best examples of specific modeling based on the concept of abstract universals. However, Shillcock notes that because natural languages are complex evolving systems, trying to identify the ideal model of a live language is always a process of catch-up.[47] Therefore, considering the evolution and the exceptions and the number of different languages all over the world, this approach has never been successfully applied in a computational model.

Now let us refer to the experiments of Markov in 1913, which are among the first linguistic models that follow the concept of concrete universals.

In what has now become the famous first application of Markov chains, Markov studied the sequence of 20,000 letters in Alexander Pushkin's poem "Eugeny Onegin" to discover the conditional probabilities of sequences of the letters in an empirical way. What follows is the less-discussed way of interpreting Markov chains, i.e. not from the traditional viewpoint of dynamic systems but rather via the empirical representation of concrete objects. Figure 17 shows the underlying concept of object-dependent representation in Markov chains. Suppose that we have a defined number of symbols in a specific language (e.g., all the observed words in the English language). Now imagine for each specific word in the collection of our words, we consider all the words that have appeared N steps before and N steps after that specific word in all of our collected texts and we count the total number of occurrences. Next, by normalizing the total number of occurrences for each position before and after each specific word, we can find the empirical ratio of having any word after or before that specific word. Now, assuming these relations between all of the words exist, we have an object-dependent

46   Noam Chomsky, "Linguistics and Brain Science," in *Image, Language, Brain*, ed. Alec P. Marantz, Yasushi Miyashita, and Wayne O'Neil (Cambridge, MA: MIT Press, 2001), 13–23.
47   Shillcock, "The Concrete Universal and Cognitive Science."

representation for each word based on its relation with all the other words. This would be a huge pre-specific representation of concrete objects. It is pre-specific because compared to the abovementioned language models, in this mode of representation of words there is no specifically given semantic or syntactic property (e.g., synonym structures or grammatical rules) to the words. The whole network is constructed out of summation and division operations. However, as Markov says, "many mathematicians apparently believe that going beyond the field of abstract reasoning into the sphere of effective calculations would be humiliating."[48]
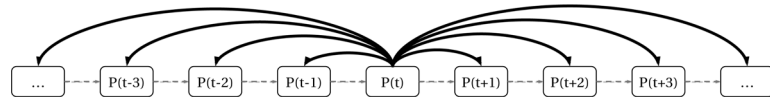
Here again we have a self-referential setup, where concrete instances are implicitly represented by their relations to the other instances. As a result, if two particular words have the same function in that language, they will have similar relations with the other words. This was a big claim in 1906, when there wasn't even enough computing power to construct these relational networks.

As Claude Shannon later noted, even after almost forty years Markov's proposed modeling framework was not practically feasible, since it demands a large number of observations and relatively large computational power.[49] Nevertheless, as mentioned in section 7, the recent rapid growth in computation power has changed the situation dramatically and a similar approach, distributed representation, has attracted many researchers and practitioners.[50] Further, new applications of neural probabilistic models of language are becoming popular, while classical approaches in natural language processing are still struggling for a real world and scalable application.[51]

The PageRank algorithm, used in Google searches, is another important application of Markov chains that follows the concept of representation on objects in a concrete level.[52] Around the year 2000, due to

48 Markov, quoted in Gely P. Basharin, Amy N. Langville, and Valeriy A. Naumov, "The Life and Work of A. A. Markov," *Linear Algebra and Its Applications* 386, no. 14 (2003): 3–26.

49 Claude E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, no. 3 (1948): 379–423.

50 Yoshua Bengio, Aaron Courville, and Pascal Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, no. 8 (2013): 1798–828.

51 Yoshua Bengio et al., "A Neural Probabilistic Language Model," *Journal of Machine Learning Research* 3 (2003): 1137–55.

52 See Sergey Brin and Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* 30 (1998): 107–17.
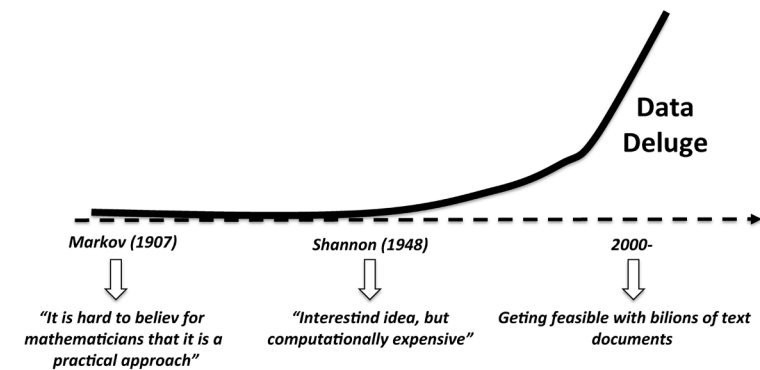
exponential growth in the number and diversity of websites, the ranking of search results yielded by Internet search engines was becoming a critical issue. Prior to PageRank, most solutions sought to define a set of features for each website and then apply a scoring logic based on these features. In other words, the starting point for the ranking system was the act of defining a generic website, represented with a set of abstract universals. Consequently, every particular page would be a point in this parametric space of the generic website. Statistically, what happens is that if we increase the size of observations (i.e., the number websites), No, the ratio of the number of parameters (dimensions), Np, to the number of observations, Np/No, quickly gets close to zero. This means that the parametric space becomes full, and the discernability of different websites from each other becomes impossible. Therefore, considering the size and the diversity of websites across the Internet, the process of defining features and assigning values to each page was a bottleneck and theoretically limited.

With the Markov-based procedure developed in PageRank, Google did not improve the classic approach, but rather changed the paradigm. They changed the basic assumption of centralized ranking and simply assumed that individuals know best which pages are related and important for them—better than any axiomatic or semantic order could know. They looked on a micro-scale at how individuals link important websites to their websites, and based on these live streams of data, they constructed, continuously and adaptively, a Markov chain, mapping how people likely surf the Internet, and thereby they constructed a probabilistic network of connections within the pages. They defined the importance of a website as the result of the importance of the websites connected to that website, which is a self-referential equation with no externally imposed feature set. In terms of statistical representation, what happens is that this object-dependent representation is scalable to the size of the

observations (websites), since each concrete object brings its representation with itself and acts as a new parameter for the representation of all the other pages (for simplicity's sake, assume a binary relation of a website to the other pages). This practical application conceptually is aligned with the definition of the Dedekind cut, where instead of representing an object directly, each object is represented as the negation of its complement.

FIG. 19
A part of Google matrix: I=HI. Everything is represented based on every other thing.
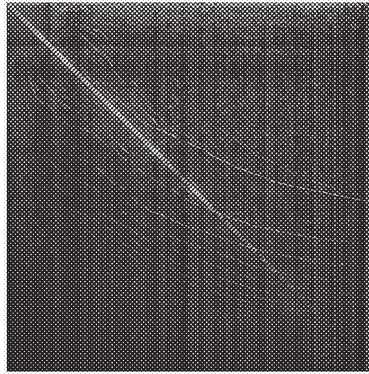
Fortunately, theories from linear algebra are available to solve this self-referential equation. The values of the first eigenvector of the constructed Markov matrix are used as the ranking of the pages. With the same methodology, it is possible to model similar problems in other fields. For example, a Markov chain based on available GPS trackings of cars can be used for modeling traffic dynamics in an urban street network.[53]

FIG. 20
A Markov chain-based representation of traffic networks from a sequence of GPS trackings of cars in Beijing.

VIII.II SELF-ORGANIZING MAP

Following the same line of argumentation for the issue of representation in complex systems that we had for Markov chains, there is another powerful data-driven, pre-specific modeling method called the Self-Organizing Map (SOM).[54] As a well-known method in machine learning, the SOM has a very rich literature with a diverse set of applications.[55] According to the literature, the SOM is a generic methodology that has been applied in many classical modeling tasks such as the visualization of a high-dimensional space,[56] clustering and classification,[57] and prediction and function approximation.[58] During the past three decades there have been different extensions and modifications to the original algorithm that was introduced by Teuvo Kohonen in 1982. For example, one can compare the SOM with other clustering methods or with space transformation and feature extraction methods such as the Principal Component Analysis (PCA).[59] It is possible to explain and compare the SOM with vector quantization methods.[60] Further, it is possible to explain the SOM as a nonlinear function approximation method and to see it as a type of neural network methods and radial basis function.[61]

However, in this work I present two main aspects of the SOM in relation to the idea of pre-specific representation and in comparison with other modernist mathematical approaches, which are based on the notions of ideals and abstract universals.

53 See Vahid Moosavi and Ludger Hovestadt, "Modeling Urban Traffic Dynamics in Coexistence with Urban Data Streams" (paper presented at 2nd ACM SIGKDD International Workshop on Urban Computing, 2013).

54 See Teuvo Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps," *Biological Cybernetics* 43, no. 1 (1982): 59–69.

55 See Teuvo Kohonen, "Essentials of the Self-Organizing Map," *Neural Networks* 37 (2013): 52–65.

56 Juha Vesanto, "SOM-Based Data Visualization Methods," *Intelligent Data Analysis* 3, no. 2 (1999): 111–26.

57 Alfred Ultsch, "Self-Organizing Neural Networks for Visualization and Classification," in *Information and Classification: Concepts, Methods and Applications*, ed. Otto Opitz, Berthold Lausen, and Rüdiger Klar (Berlin: Springer, 1993), 307–13; Juha Vesanto and Esa Alhoniemi, "Clustering of the Self-Organizing Map," *IEEE Transactions on Neural Networks* 11, no. 3 (2000): 586–600.

58 Guilherme de A. Barreto and Aluizio F.R. Araujo, "Identification and Control of Dynamical Systems Using the Self-Organizing Map," *IEEE Transactions on Neural Networks* 15, no. 5 (2004): 1244–59; Guilherme de A. Barreto and Gustavo M. Souza, "Adaptive Filtering with the Self-Organizing Map: A Performance Comparison," *Neural Networks* 19, no. 6 (2006): 785–98.

59 Hujun Yin, "Learning Nonlinear Principal Manifolds by Self-Organizing Maps," in *Principal Manifolds for Data Visualization and Dimension Reduction*, ed. Alexander N. Gorban et al. (Berlin: Springer, 2008), 68–95.

60 Christopher M. Bishop, Markus Svensén, and Christopher K. I. Williams, "GTM: The Generative Topographic Mapping," *Neural Computation* 10, no 1 (1998): 215–34; Teuvo Kohonen, "Improved Versions of Learning Vector Quantization" (paper presented at the IJCNN International Joint Conference on Neural Networks, 1999).

61 See Barreto and Araujo, "Identification and Control of Dynamical Systems Using the Self-Organizing Map"; Barreto and Souza, "Adaptive Filtering with the Self-Organizing Map."

As discussed in section 5, in specific modeling the observations of any real phenomena are encoded in a generic object being represented by a set of given parameters. The underlying idea of pre-specific modeling is how to relax the modeling process from any specific and idealistic representation of the real phenomena—or, how not to depend on the generic object. For example, as mentioned before, in a Fourier transformation we assume that any dynamic behavior can be reconstructed and re-presented by a set of ideal cyclic forms. As figure 21 shows, an observed signal can be decomposed or can be approximated as a linear summation of some ideal waves. In other words, here we assume that there is a generic sinusoidal wave (as an ideal behavior) with a parametric setup, and by changing the parameters of this generic function there are different instances of waves. Finally, an observed signal can be represented as a summation of these ideal waves as follows:
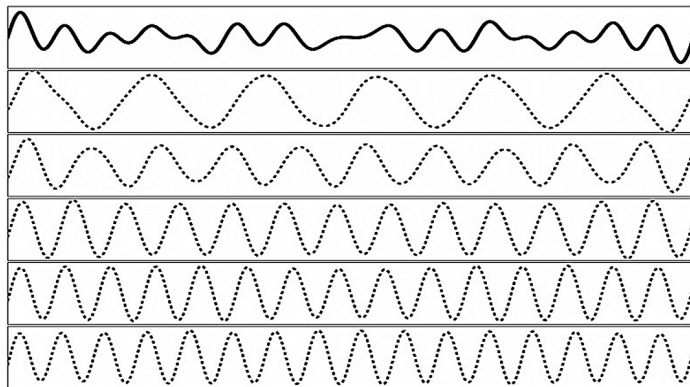
FIG. 21
Reconstruction of
an observed signal
(top row) based
on a parametric
dictionary of ideal
waves in Fourier
decomposition.

$$s(t) = a_o/2 + \sum_{n=1}^{m} [a_\pi \cos(nwt) + b_\pi \sin(nwt)]$$

Therefore, the Fourier transformation is among the specific models that are based on abstract universals. As discussed, although powerful and useful in many classical engineering and scientific applications, this approach of idealized modeling has fundamental limits in dealing with complex (multifaceted) phenomena. The opposite idea or the complementary idea in pre-specific modeling is that based on the concept of concrete universals, it might be possible to establish a self-referential setup using concrete objects (i.e., the observations) to model real phenomena without any external representation or any external control. In the domain of machine learning, this is the underlying idea of unsupervised learning. Interestingly, the SOM corresponds with the idea of representation based on concrete universals. Compared to Fourier decomposition, shown in figure 21, if we train a SOM with enough observations, we get a dictionary of potential dynamic forms collected via real observations (figure 22). In assuming each of the prototypical forms in the trained SOM as a word or a letter in a language, a trained SOM can be used as pre-specific dictionary for the target phenomena. In other words, in terms of signal processing, assuming a fixed segmentation size, the observed signal can be translated into a set of numerical indexes (i.e., the index of the matching prototypes in the SOM network with each segment of the observation vector)—further, these indexes will be used for further steps of modeling. The main point is that unlike the case of Fourier decomposition, here there is no external axiomatic setup for the transformation of observations into codes; the whole encoding system provided by the SOM is based on concrete observations.
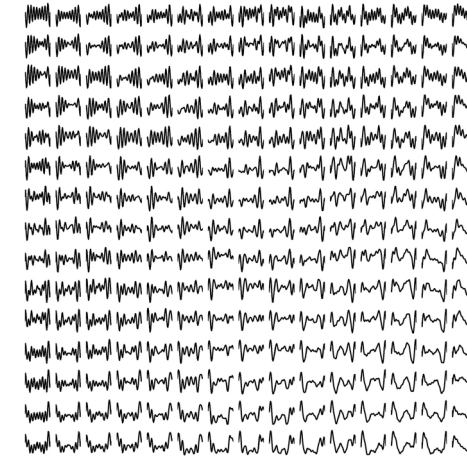

FIG. 22
Self-organizing
maps: Constructing
a pre-specific
dictionary of
dynamic forms from
a large collection of
observed signals.

Further, in a certain topology of SOM networks, the final indexes can be used to transform a multidimensional dynamic system into a one-dimensional symbolic dynamic system. In this case, the indexes of the SOM can be considered as "contextual numbers."[62]

In the field of computer vision and speech processing there has been a growing trend of methods that are based on the idea of representation that outperform many classical pattern recognition methods only in coexistence with a large amount of observations based on feature engineering. Classically, feature engineering means that in order to

62   Vahid Moosavi, "Computing with Contextual Numbers," *arXiv preprint* (2014).

develop a pattern recognition model (i.e., in an image classification problem), one first needs to design a feature space to transform the images by that and then to develop a classification model on top of the engineered features. In the example of the Fourier analysis, the frequency and phase difference are the features. On the other hand, in this new category of modeling, sometimes called representation learning, there is no specific and separate feature-engineering task before adjusting the classification or prediction method.[63] Among these algorithms is the sparse coding algorithm, which has conceptual similarities with the SOM.[64] The principle idea of sparse coding is that if the original observations are n dimensional vectors, one can find an over-complete set of vectors (i.e., K vectors, where $K>>n$) to reconstruct the original observations with a linear and sparse combination of these K vectors. While it looks similar to methods such as PCA[65] or Independent Component Analysis (ICA),[66] sparse coding (similar to the SOM algorithm) does not produce a global transformation matrix. In PCA for example, all the n orthogonal basis vectors proportionally (according to their corresponding eigenvalues) contribute to the representation of all of the original observations, but in the SOM and sparse coding we have a kind of "distributed representation," in which each original observation is directly represented by a few specific prototypes (basis vectors). In other words, in the SOM each prototype is an object, which is not true for each principal component in PCA. They are from different worlds.

Further, this encoding approach can be applied in a hierarchical process. For example, in the case of image processing it can be applied to small patches of an image, where each patch will be indexed to a few codes and the next level (for example the whole image) will then be represented by new codes constructed on top of the previous codes. In fact, the output of one step is used as input for the next layer. Therefore, the whole image is analyzed by multilevel sparse codes. This simple idea of coding in an unsupervised approach has been applied in many practical applications and it has been claimed that it works better than the wavelet decomposition method.[67] I should note that the wavelets act similarly to the Fourier series, but they are more advanced, since there is no longer the assumption that the underlying ideal waves are

stationary. Figure 23 shows an example of a sparse coding algorithm applied to image patches.

63   See Bengio, Courville, and Vincent, "Representation Learning."
64   See Bruno A. Olshausen, "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images," *Nature* 381, no. 6583 (1996): 607–9.
65   Karl Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, no. 11 (1901): 559–72.
66   Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis* (Malden, MA: John Wiley & Sons, 2001).
67   See Olshausen, "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images."
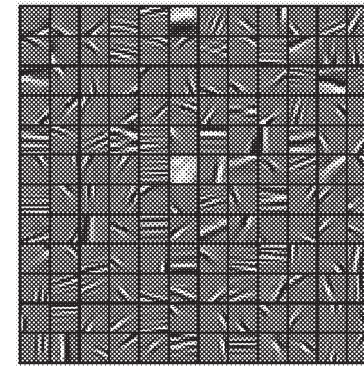
FIG. 23
Sparse coding: Learning data driven image dictionaries.[67]

VIII.II.II COMPUTING WITH INDEXES BEYOND IDEAL CURVES

Another property of the SOM is its unique disposition for structural learning. Figure 24 shows the main difference between the SOM and a classical way of relation (function) modeling. In simple terms, the primary goal of relation modeling is to find the relation between two dimensions based on a set of observations. In a classical way of modeling, one needs to fit a curve (a fixed structure) to a data set, while minimizing the deviations (errors) from the selected curve. In other words, the selected curve represents the logic that idealizes the observed data into a continuous relation. The SOM assumes that the logics (the argument that integrates cases) can be extracted from within the observed data—and it conserves all the logics (arguments) according to which it clusters the cases. What is optimized in such modeling is not how the data fits to logic, but how the logic, which is being engendered, can accommodate as much as possible from the data.

In this sense, in an analogy to a governance and decision-making system, we might say that the classical approach in curve fitting is a democratic setup, in which there is a global structure, tuned locally by the effect of individual votes. On the other hand, the SOM provides a social environment, in which each individual instance is not reduced but is kept active in its own individuality, while individuals can be unified into local clusters, if necessary.

68   This image is taken from James Hughes, Daniel J. Graham, and Daniel N. Rockmore, "Quantification of Artistic Style through Sparse Coding Analysis in the Drawings of Pieter Bruegel the Elder," *Proceedings of the National Academy of Sciences 107, no. 4,* (2010) 1279–83.

FIG. 24
Computing with
indexes beyond ideal
curves.
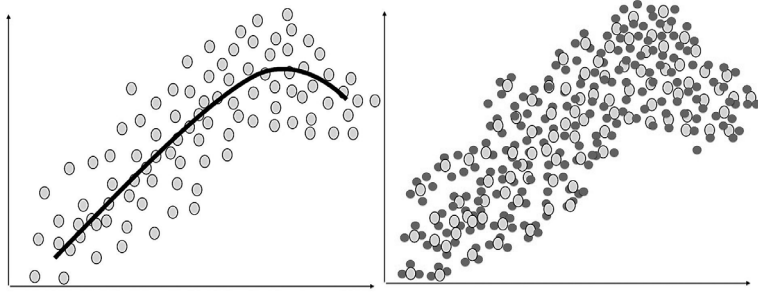
Here again, the final model of the real phenomena is an abstraction of
any potential specific model and it does not import any axiomatic or
semantic specificity. Further, if the real environment is dynamic and
evolving, and if we can assume the availability of dynamic data streams,
then the SOM evolves along with the environment. In other words, pre-
specific models are models in coexistence with data streams.