Minds & Machines (2008) 18:475–491
DOI 10.1007/s11023-008-9120-8

Computational Models of Emergent Properties

John Symons

Received: 15 November 2007 / Accepted: 3 May 2008 / Published online: 14 November 2008

Abstract Computational models fail to shed light on general metaphysical questions concerning the nature of emergence. At the same time, they may provide plausible explanations of particular cases of emergence. This paper outlines the kinds of modest explanations to which computational models are suited.

Keywords Emergence · Computational models · Explanation · Cellular automata

Introduction

Computational modeling plays an increasingly important explanatory role in cases where we investigate systems or problems that exceed our native epistemic capacities. One clear case where technological enhancement is indispensable involves the study of complex systems.[1] Even in contexts where the number of parameters and interactions that define a problem is small, simple systems sometimes exhibit non-linear features which computational models can illustrate and track. In recent decades, computational models have been proposed as a way to assist us in understanding emergent phenomena.

The core concern of this paper centers on the following question: Assuming that emergent properties are a genuine feature of the natural world, how might computational models help us to generate explanatory accounts of those properties?

[1] The term 'technological enhancement' in this context is due to Humphreys (2004). There, Humphreys provides a detailed consideration of some of the epistemological implications of increasing reliance on computational modeling for science. Not all such enhancements simply leverage the computational power of machines. Some, for instance the automatic telescopes discussed by Humpreys (ibid 6) serve to overcome non-epistemic limitations such as fatigue or slow reaction times.

J. Symons (✉)
Department of Philosophy, University of Texas, El Paso, TX, USA
e-mail: jsymons@utep.edu

Putatively emergent properties such as the flocking behavior of birds (Reynolds 1987), the adaptive features of the immune systems (Hofmeyr and Forrest 2000) and the characteristic patterns of traffic flow (Schreckenberg et al. 1995) have been given computational models. In what sense (if any) do such models help to explain the respective emergent features under consideration?

The role of models in theories and explanations has been an important topic for philosophers of science since the early 1970s.[2] However, when investigating emergent properties, the nature of the explananda makes the role of the model especially problematic. Given the supposed irreducibility of emergent properties, in order to claim that one has provided computational explanations for emergent properties one must first distinguish computational models or explanations from reductions. If computational models have a reductive character, then the very idea of a computational model of an emergent property would be self contradictory. So, for instance, Joshua Epstein identifies computational models as reductive and contrasts what he sees as the desire of classical emergentists to "preserve a "mystery gap" between micro and macro" with agent-based modeling which "seeks to demystify this alleged gap by identifying microspecifications that are sufficient to generate—robustly and replicably—the macro (whole)" (1999, p. 55) Against Epstein, this paper argues that computational studies of emergence adopt a modest approach to the ambit of their models and that the explanations provided by such models rely on mechanisms whose generality is limited. As such, these models fail to shed light on general metaphysical questions concerning the nature of emergence while nonetheless providing plausible explanations of particular cases of emergence.

While the basic metaphysical problem of emergence will not be solved by computational modeling projects, we have ample reason to believe that specific cases of emergence are amenable to scientific explanation of a more modest sort. Contrary to the claims of Epstein and others, such modest models of emergence can be explanatory without necessarily being reductive.

Computational Emergence

In philosophical discussions of emergence, computational models are usually deployed in support of epistemological characterizations of emergent properties. For

[2] An emphasis on the role of models in scientific explanation has been one of the central characteristics of the semantic tradition in the philosophy of science (Suppes 1962; Van Fraassen 1980). According to the semantic view of scientific theories (esp. Suppe 1977; Van Fraassen 1980) scientific inquiry and explanation is largely a matter of the construction of models. The semantic view of theories emphasizes the abstract structure of theory rather than on the particular syntactic presentation of the theory (Van Fraassen 1972). While models are central to the semantic view of theories, when we consider specifically computational models, we encounter a layer of philosophical challenges of which philosophers in the semantic tradition in the philosophy of science were mostly unaware. In computational modeling, the question of application and implementation are unavoidable. While it is useful to consider theories as abstract structures, in the context of computational modeling as Humphreys points out: "(s)yntax matters… The importance of syntax to applications, and especially to computational tractability, is something that the semantic account of theories, for all its virtues, is essentially incapable of capturing." (2002, p. S3)

instance, Mark Bedau provides an account of what he calls weak emergence in terms of computational simulation (Bedau 1997). Bedau defines weakly emergent features of a system as those which can be derived from the microdynamics of the system only by an exhaustive simulation. Thus, weakly emergent properties are some subset of the properties of a computational simulation which are distinguished by reference to an epistemic agent's inability to predict their appearance without having first run each step in the simulation.[3] Observing runs of the simulation could later lead an agent to inductively predict the appearance of the emergent property, but such predictions are warranted by judgments concerning the reliability and regularity of the behavior of the simulation, they are not based on an understanding of the object whose behavior is being simulated. In light of the possibility of successful inductions without detailed understanding, we could say that weakly emergent properties are characterized such that for all epistemic agents, including omniscient ones, the shortest *non-enthymatic* path towards a prediction of the emergent property has the same computational complexity as the simulation which produces that property. In any event, any attempt to specify what we mean by weak emergence will depend on reference to some properties of the simulation under consideration. Thus, weak emergence is only derivatively a matter of the relationship between an observer and a model. It is primarily a result of the computational complexity of the output of a model.

When computer scientists use the term 'computational emergence' they usually mean to mark some feature of a process which the software designer did not intentionally code into the algorithms which generate the process. As such, emergent features are sometimes described in loose terms as not having been 'hard coded' into the source code of an executable object. In most practical programming contexts emergent properties are either of little interest or are a nuisance to be avoided. However, some computer scientists have proposed actively exploiting features of emergent computing in the attempt to model emergent features of the natural world. For example, Stefanie Forrest writes: "…interesting and useful computational systems can be constructed by exploiting interactions among primitive components, and further, that for some kinds of problems (such as modeling intelligent behavior) it may be the only feasible method." (1990, p. 1) Thus, emergent computing is both a result of the unanticipated interplay of parts of an executable object and a way of modeling the appearance of putatively emergent properties. As such, on Forrest's account, emergence is both a property of the computational model itself and a property of the objects being modeled. While the literature on computational emergence sometimes tends to conflate these two roles for the notion of emergence, the distinction is indispensable if we hope to understand the explanatory role of computational models of emergence.

Forrest's view of computational emergence is intended to be applied to a wide variety of cases. For the purposes of this paper, it suffices to focus on the simplest and most familiar context in which emergence is studied computationally, namely the cellular automata models (CA) that were first described by John von Neumann

[3] Peter Cariani (1991) also emphasized the role of observers in a similar, though less formal manner, when describing what he calls 'computational emergence'.

and Stanislaw Ulam. Their development was part of von Neumann's investigation of the possibility of self-reproducing machines. Von Neumann developed CA as a way of characterizing systems wherein the rule-governed interactions of basic constituents can be observed as they unfold over a series of discrete steps. In order to isolate the problem of self-reproducing machines from other questions (for example, questions concerning the representational relationship between the model and the actual conditions in the physical world) von Neumann stipulated that in a CA, the objects under consideration be isolated and governed solely by the rules of the model. Thus, as Hu Richa and Xiaogang Ru note (2003), standard CA can be characterized in terms of a quintuple set: {Cells, Cell Space, Cell State, Neighborhoods, Rules}. Where *cells* are the basic objects or elements of the CA each having some individual state depending on the rules of the CA. *Cell space* is defined as the set of all cells and their values at some time. *Neighbors* are the set of cells surrounding some any center cell and *rules* are the transition functions of cell states, mapping cell spaces to cell spaces (Hu Richa and Xiaogang Ru 2003, p. 1047). The rules of the CA are defined as being maximally general with respect to the cells in the model and the application of rules updates each cell synchronically.[4]

One especially suggestive and well-known descendent of von Neumann's method is John Conway's *Game of Life* automaton.[5] Conway's automaton generates patterns that exhibit some of the behaviors we intuitively associate with living things. As such, it illustrates von Neumann's insight that simple rules can generate complicated outputs. By stipulating well-defined rules and boundaries for the model, von Neumann's strategy permitted the exclusion of myriad contextual influences that play a role in real biological or physical investigation.[6]

CA like von Neumann and Conway's exemplify what John Holland would later call constrained generating procedures (Holland 1997). Holland treats constrained generating procedures as representations of the interaction of simple mechanisms wherein possible interactions are represented via some transition function. A transition function maps a set of possible states of a system onto itself (1997, pp. 130–131). The interaction of transition functions may give rise to new kinds of regularity some of which, in turn, can also be given fruitful formal representations as components of further interactions.[7]

[4] For the sake of simplicity, I will focus on the individuals-based models. However, it should be noted that many of the most interesting computational models are hybrid, rather than pure CA. So, for example Christina Warrender's models of the peripheral immune system (2004) employ both an agent based and a particle-systems model. Cells involved in the earliest stages of infection are small in number and are not appropriately modeled by continuous representations. However, the number of molecules involved in an infection far outstrips the number of cells and so she models the molecular environment of cells as well as many of the components of each cell state as continuous variables.(2004, 17). While much of my argument will involve a comparison of CA and differential equation models, the argument for modesty can also be extended to the hybrid cases.

[5] Cellular automata became especially well-known in the early 1970's with the appearance of two articles by Martin Gardner in *Scientific American* devoted to the Game of Life.

[6] For von Neumann's thoughts on recursivity and self-replication see Burks A. W. (Ed.) (1970).

[7] Following Holland's explanation of the role of transition function (only slightly modified) we begin with some set of states S $\{s1, s2, s3, \ldots\}$ which is taken to be finite for the sake of computational tractability. A transition function takes as its argument some state of the system in combination with some

Chris Langton and others have described parameters on the space of possible CA rules via formal characterization of the transition functions for CA-like systems (1990). Parameterization of the space of possible rules for CA reveals a variety of important mathematical properties. Perhaps the most interesting property of the CA is that the set of logically possible mechanisms that can be given a CA representation is wider than those that can be modeled via differential equations (Hu Richa and Xiaogang Ru 2003, p. 1049). With this in mind, CA (suitably characterized) may be preferable to differential equations, for certain purposes, because they can capture a very wide set of values, including all the discrete cases that ordinary differential equations miss.

While such formal systems merit attention in their own right, we may still ask whether, they shed light on cases of emergence in the world beyond the model. In the case of Holland's approach, emergence via constrained generating procedures is such that given a system governed by more than one kind of rule ranging over the behavior of a simple set of elements we can sometimes find a set of initial conditions such that some macroproperty of interest is generated by the system in question. What is the relationship between the formal properties of the model itself and the explanandum that serves as its putative subject matter? Surely, for instance, Conway's Game of Life has relatively little explanatory content when it comes to understanding the emergence of real biological systems. Von Neumann was sensitive to the peculiar status of computational models. He wrote for instance: "The formalistic study of *automata* is a subject lying in the intermediate area between logics, communication theory, and physiology. It implies abstractions that make it an imperfect entity when viewed exclusively from the point of view of any one of the three above disciplines." (1966, p. 91) One might ask whether the abstraction that concerned von Neumann is any more severe in the case of automata than it is in the case of other scientific models. The next section unpacks some of the distinctive properties of the abstraction or independence of CA and related computational models.

Computational Models as Scientific Models

Unlike a controlled experiment in biology or chemistry, a computational model concerns only an abstract representation of the form of the system under consideration. In explanations which depend on a computational model, the gap between the explanandum and the explanation differs in important ways from ordinary scientific contexts. As von Neumann recognized, the abstract nature of

Footnote 7 continued
input at a time and gives as a value a state of the system. For any input of type j there will be an associated set of possible input values Ij. Thus, $Ij = \{ij1, ij\,2, ij\,3, \ldots\}$, where $ij2$ names state number 2 of the input j. Given k types of input for the system there will be k sets of possible input values $\{I1, I2, I3, \ldots Ik\}$. The set of all combinations for the system is given as the product of the sets $I1 \times I2 \times I3 \ldots \times Ik$. Now, the transition function can be defined as f: $(I1 \times I2 \times I3 \ldots \times Ik) \times S \rightarrow S$ and the temporal dynamic of the system can be defined as $S(t+1) = f(I1(\mathrm{t}), I2(\mathrm{t}), I3(\mathrm{t}), \ldots Ik(\mathrm{t}), S(\mathrm{t}))$. The iteration of f generates the state trajectory of the system.

computational models of emergence makes the question of their application more problematic than might be the case for other scientific models. Computational models of emergence, like those proposed by Holland, are formal systems of elements, classes and functions or operations such that within the parameters of the system, emergent features follow from the definition of the model. While constrained generating procedures are purely formal and can be studied as abstract mathematical objects, once we treat these computational processes as models, implementing them on a computer and observing their consequences via some graphical representation or animation, we have taken an additional conceptual step. Treating the CA as a model of some part of the natural world takes us beyond the formal characteristics of the coded object itself and raises some challenging philosophical problems for the modeler. Qua model, the CA have the paradoxical characteristic of being both applied to some domain while at the same time being tightly insulated systems that are defined by their own maximally general internal rules.

Like most scientific models CA are generally conceived such that interactions with the world beyond their boundaries are excluded or tightly controlled. CA are exaggerated versions of this traditional scientific modeling strategy insofar as cells are characterized in the most minimal form possible, such that their only relevant features are those that are subject to the rules governing the model. Günter Küppers and Johannes Lenhard (2005) note that even though a computational model may be based on theoretical models, e.g. a system of non-linear partial differential equations, "they require further steps of formal treatment because they have to be implemented into a computer. In this sense simulation models are partly independent from the underlying theoretical model."(2005, p. 7) This independence partly results from the need to provide a computational implementation, but it also results from the process of abstraction and insulation that is involved in crafting a computational model. For example, the objects in a CA are (or are composed of) cells whose behavior is determined solely by the rules of the model. In traditional scientific modeling, it would be a mistake to claim that the identity and of the objects under consideration is completely exhausted by the rules or laws governing the model. For instance, in a gravitational model, while the equations might apply to point masses in a system, the objects whose behavior is of interest are extended physical objects rather than the point masses that figure in our equations. Insofar as they are intended as models of some real physical situation, the nature of the objects treated by a gravitational model is not exhausted by the laws governing a model in which equations govern point masses. The intended application of a gravitational model is a system of extended bodies rather than a set of point masses. It would be perverse, for instance to claim that gravitational models should be evaluated by reference to their success in applying to systems of point masses. It is trivially true that the intended application of the model determines our interpretation of the objects which feature in the model itself. That the properties and behavior of the explananda are only partly captured by the simulation is what permits the possibility of improving the model.

Strikingly, when we lose our grip on the difference between the objects as they are figure in nature and the objects as they figure in the model, the model risks losing any explanatory import. If we were to restrict our understanding of the model such that we regarded it as simply a system which outputs generalizations solely on the basis of the way the model was stipulated or defined, then it is in danger of becoming trivialized.

One of the reasons for the semantic turn in philosophy of science was that axiomatic models of science were subject precisely to this kind of trivialization. Recall that one of the advantages of models-based or semantic approaches over the axiomatic approaches to science is that the axiomatic approach risks prematurely fixing some point in the history of a theory rather than acknowledging the development of improved theories. This was especially true in the biological sciences. As Alex Rosenberg notes "When we try to frame the theory of natural selection into an axiomatic system, the result (fails to) reflect the full richness of Darwin's theory… In particular, the theory's assertion that the fittest among competing organisms is one easy to deprive of explanatory force if we define "the fittest" as those which survive and reproduce." (Rosenberg 2000, p. 100) Philosophers of biology were drawn to the models-based approach to science insofar as it seems adequate to the idea that well-established theories offer models which are subject to refinement and local modification.

Computational models risk losing this advantage insofar as they are subject to an unusually severe kind of "screening off" of interfering conditions at a variety of levels. This partly explains what some authors have described as the independence or autonomy of computational simulations in scientific inquiry (Humphreys 1991; Galison 1996). For computational models, this independence becomes a challenge when we consider the interpretation of the objects in the model. Unlike, for instance a gravitational model which idealizes extended masses as point masses for the purposes of the providing manageable equations, the characteristics of the objects in the CA are less obviously separable from the rules governing the CA. There is no sense in which we can better approximate the characteristics of some particular object via the behavior of a cell in a CA, insofar as the cell and the rules of a CA are interdefined.

By way of contrast, consider how scientists describe what it means to model with differential equations. In a traditional model employing differential equations, it makes sense to think of a modeling cycle consisting of roughly four steps (Fulford et al. 1997, p. 2). First, the system under consideration is observed and the relevant quantities and relations are determined. This involves simplification and abstraction, but it is still closely tied to the act of measuring some feature of the system in question. Second, we represent the relations between measured quantities as equations. We solve the equations and interpret the solutions as answers to questions concerning the original system. Then finally, we determine whether our predictions about the system in question (our interpretation of the solutions for our equations) make sense in light of our experimental investigation of the system. As Fulford et al. note, there is a cyclical process to modeling of this kind insofar as failures at various stages of the modeling process can lead us to refine previous stages. Of course, if the four-stage process is successful, then the model can be used to make further predictions of the system in question. While this description of what is involved in modeling with differential equations is highly simplified, the contrast with what we see in CA models is clear.

The most obvious difference between modeling with CA and differential equations is the lack of any quantitative results in the CA cases. This means that CA simulations provide qualitative analogies to the natural systems in question rather than measurable quantities. These analogies can be extremely useful and may give us some insight into methods for controlling the phenomenon in question. However, in practice and in application, generalizations that we derive from these models will rest entirely on the analogy between the system under consideration and the simulation. This is one of the reasons that the success of a computational model is generally judged not by its predictive power, but by the degree to which it imitates the known behavior of a target system. So, for instance as Küppers and Lenhard note, Norman Phillips' classic climate model was judged by the extent to which it was able to reproduce observed global flow patterns in the atmosphere rather than with respect to the adequacy of the six partial differential equations that formed the initial inspiration for his computational model. In particular "one criterion was the complex pattern of the so-called surface westerlies, winds blowing continuously north(wards from) the equator…The decisive criterion for success was the adequate imitation of the phenomena, i.e. the flow patterns, not the derivation from theoretical principles." (2005, p. 4)

This analogical feature of CA modeling distinguishes it in another important way from traditional models that employ the kind of cyclical approximation methods described above. For instance, if we are interested in grasping the basic laws of nature via our models, then we hope that the objects as depicted by our model will share properties in common with the objects as they exist in nature. The more successfully the equations allow us to predict and control the behavior and properties of the objects of interest, the closer we are to grasping the laws governing that object. The 'objects' which feature in the CA are defined in terms of the rules of the model and there is no sense of a direct comparison between the object, taken as part of the model and the object in the world.

Even in cases where our differential equation model ranges over objects which are quite different from the objects in nature which we are trying to understand, (think, for example, of using a gravitational model which ranges over point masses to understand the behavior of planets) the critical difference with computational models is the role of the model objects in the differential equation model as stand-ins for the object in nature. Thus, while a gravitational model might employ point masses, they can be understood to stand as proxies for planets. By contrast, in CA, insofar as we can still talk about objects in a model standing for objects in nature, this relationship is mediated by the analogy between the model as a whole and the relevant portion of nature under consideration.

In models using differential equations we can diminish the space between the object in the model and the explanandum by adjusting the equations governing that object as a result of our measurements or as a result of revising our view of the relationship between quantities. In the case of CA it becomes more difficult to understand how we could provide incrementally more accurate descriptions of the target objects. In CA (whether we see cells or patterns of cells as the objects featuring in the model) we lack a comparable sense in which we can consider improving the correspondence between the objects in the model and the objects in nature. If we identify the CA with its quintuple and in particular its rules, then we will simply be unable to modify the CA to get a more accurate representation of some phenomenon. While that identification is common for classifying CA, it is obviously necessary to go beyond that characterization when they are used as scientific models. For example, it is surely appropriate to allow modifications in the rules for the CA to improve fit just as it is acceptable to change parameter values in differential equation models.

However, the difference between the two approaches has less to do with the possibility of tweaking the model as a whole (clearly this is possible in both cases) but rather with the distinct role of objects in the models. This difference becomes clear when we consider the role of initial conditions in the two types of model. Generally, when we consider the definition of a CA, we treat it as independent of the choice of possible initial conditions. However, in many cases, the choice of initial conditions will determine whether or not some putative object will figure in the model (in *Life,* for instance, only some initial configurations will produce gliders). Given that the "objects" of interest depend on the choice of initial conditions, and given that CA are defined independently of that choice, we can conclude that CA are not models of those objects per se. Put bluntly, even though, point masses are very different from planets, the motion of the planets is the target of an explanation involving such models. Such models begin from the assumption that there are objects with distinctive properties. The goal of standard scientific models is to get a better representation of the properties or behavior of those objects. In order for CA models to play a similar role, we would have to include the relevant initial conditions in the definition of the CA. Restricting the CA to some subset of initial conditions in order to make the CA accommodate some target object would involve a very different kind of modeling project than traditional definitions of CA have envisioned.[8]

Kuhn et al. (2003) distinguish between verification and validation of a computational model where ""verification is "building the system right," while validation is "building the right system." If we have a set of requirements, we can verify, formally or informally, that the system implements the requirements. But validation is necessarily an informal process. Only human judgment can determine if the system that was specified and built is the right one for the job." (2003, p. 1) The task of validating the computational model is absolutely dependent on human judgment insofar as we must judge that the analogy holds between the patterns in the model and the patterns in the object.[9] In the terms employed by Kuhn et al. if we have not built the right system for the job, then we simply do not have a model. By contrast, in other areas of science we have an extremely high tolerance for a lack of fit between the model and the entities being modeled. A Newtonian model of the solar system, for instance, is literally false, insofar as it excludes many small factors,

[8] I am very grateful to Paul Humphreys for his suggestions here.

[9] For some discussion of the technical aspects involved in verification, validation and certification of computational models see Balci 2003.

(friction, comets, electrical fields) oversimplifies the objects under consideration (treating planets as point masses) and contradicts our understanding of relativistic effects. What permits us to make fruitful use of Newton's model in technological other practical aspects of life is that we do not rely on the truthfulness set of all statements that we could, in principle, derive from the model. Instead, we exploit the approximate correspondence between the objects as mentioned in the model and the real entities under consideration.

Metaphysics and Explanation

In the case of the study of emergence, the purpose is not to approximate or predict the behavior of some object, rather the goal is to understand the conditions which might give rise to some new object or property. For scientists interested in emergence, as we have seen, CA permit a simulation of how the interplay of distinct constraints (at various levels) or rules can govern the behavior of the components of the model so as to give rise to properties which can themselves be subject to transformations of various kinds. The study of how constraints interact makes involves a somewhat different approach to the way models function. Here, it is less clear what values we are measuring insofar as we are not directly concerned with the properties of target objects for instance. In the case of emergence, the explanandum is the very appearance of an entity or property rather than its behavior.

Often, the kind of complex systems of interest to scientists studying emergent properties would be intractably difficult to understand without the capacity to exclude interfering conditions. The practice of screening off interfering conditions leads to generalizations that hold within the context of the model or at best *ceteris paribus* (all else being equal). *Ceteris paribus* clauses, or provisos are a feature of many areas of scientific investigation. However the kind of screening off that we see in CA models is of a more extreme kind than we find in, for example a traditional model based on differential equations insofar as the objects of the CA cannot be meaningfully considered apart from the rules governing the model. In traditional scientific models, there is the assumption that the objects under consideration have some standing independently of the model and that the goal of modeling is to get a more accurate account of their properties. Some philosophers of science, most notably Nancy Cartwright have argued that all scientific inquiry and explanation takes the form of the construction of models in which the objects under consideration are in large part constituted by their role in these models. She contends that the generalizations that we derive from scientific models (including models of fundamental physical phenomena) have limited generality. Such generalizations provide local truths concerning very restricted domains. Without agreeing to her more general claim concerning the generality of physics, it seems quite clear that CA models fit her characterization precisely.

Computational models of the kind discussed here have some distinctive features. They work by analogy with (or by imitating) the portion of the natural world in question, they have limited generality, they are holistic and as I will argue below, they are ontologically agnostic. These features have significant implications for the attempt to provide computational accounts of emergence.

This is not the place to defend the scientific status of *ceteris paribus* generalizations. However, given that computational models of emergence have the kind of modest character described above, then the explanations that they support will have a limited ambit. In order to determine what the limits of this kind of explanation might be, this section argues that the kinds of mechanical explanations offered in for instance, complexity theory, are not equivalent to ontological reduction in any sense which would contradict the metaphysical position of an emergentist. This section briefly describes the metaphysical problem of emergence, in order to show that the kind of computational models under consideration here have relatively little relevance.

The metaphysical problem of emergence is easy to appreciate. Common sense tells us that putatively emergent features of the natural world—things like organisms, minds, economies and nation states—are real. Thus, we are inclined to think that a world without such things would have a smaller inventory of real objects than the actual world. Likewise, for example, it is intuitively evident that the appearance of conscious mental life at some point during the course of natural history meant the arrival of something genuinely new. Metaphysical reflection quickly departs from common sense, noting that ontological novelty seems to entail an unacceptably paradoxical form of downward causation and arguing that there are not likely to be as many kinds of object as common sense leads us to believe. Since the microstructural components are thought to be able to do the job of the macro-objects, ontological parsimony encourages us to avoid counting macro-objects in our inventory.

Most emergentists are physicalists and so, like the common sense ontologist, they must answer the challenge of epiphenomenalism for emergent properties. Stated more precisely, the challenge is to understand how an emergent property of a system can act on its constituents? Surely this requires that by acting on its constituents, the emergent property changes the very things that make it what it is? If so, then wouldn't the identity of the organism be changing in such a way as to make it impossible to say that it is acting on itself? Taken in its strictest sense, it looks like the idea of systems acting on their own constituents reduces to absurdity. While constitution and identity are distinguishable notions, the implicit contradiction in cases of emergent properties acting on their constituents leads many philosophers to conclude that the putative causal powers of emergent properties are always causally preempted by the properties of their underlying constituents.

If we accept the causal inheritance principle,[10] any powers, over and above the powers of constituents are not logically possible. Thus, on the standard view, a non-trivial model of downward causation (and, by extension, of emergence) can make sense only given a weakened conceptual interpretation of emergent properties, wherein, as Jaegwon Kim writes: "we interpret the hierarchical levels as levels of

[10] If a functional property *E* is instantiated on a given occasion in virtue of one of its realizers, *Q*, being instantiated, then the causal powers of this instance of *E* are identical with the causal powers of this instance of *Q*.

concepts and descriptions, or levels within our representational apparatus, rather than levels of properties and phenomena in the world." (Kim 1999, p. 33). Kim, like Trenton Merricks (2001) and others, places a significant burden on the conceptual operation of the mind while relegating the rest of the natural world to the status of a single basic stuff. This stuff manifests itself through its homogenous causal power, but insofar as we judge it to be individuated, it is our judgments that do the individuating. For Kim and Merricks then, the world is split in two with minds on one side and stuff on the other. While we may identify new patterns and phenomena for instrumental or other subjective reasons, these can only be shown to be non-arbitrary and have some basis in non-mental reality given that they make a difference in nature. However, on the Kim/Merricks view, any candidate powers that we might identify with emergent properties are preempted by the powers of nature's single basic stuff.[11]

Those of us with metaphysical inclinations might wonder whether Holland-style models prove the in-principle reducibility of emergent properties to a more basic set of physical laws. So, for instance, one general line of objection to the claim that there are computational models of emergence is to argue that the very existence of a computational model of some phenomenon is tantamount to its reduction. Since emergent properties are supposedly irreducible, a computational model of emergence might seem like a self-contradictory notion and instead, computational models of a putatively emergent phenomenon in the natural world might be taken as proof that the phenomenon is reducible rather than emergent. The oxymoron argument rests on a mistaken interpretation of the explanations that these models provide.

As mentioned in Part One Joshua Epstein argues that what he calls the generative sufficiency of agent-based computational models is equivalent to explanatory sufficiency which, in turn is equivalent to reductionism. "it is precisely the generative sufficiency of the parts (the microspecification) that constitutes the whole's explanation! In this particular sense, agent-based modeling is reductionist." (1996, p. 55) By identifying reductionism, generative sufficiency and explanatory sufficiency Epstein has left no room for a non-reductive explanation. The sense in which explanations can be non-reductive is relatively obvious. As Putnam famously noted (1975, p. 295), there are often good reasons for finding non-reductive explanations preferable to reductive accounts. To use Putnam's example; explaining why a round peg of 1 inch diameter fails to fit through a square hole with a 1 inch diagonal is clearly not a task for quantum physics.[12] It is a mistake to identify

[11] Symons 2001 proposes a response to this argument against emergent properties.

[12] Rueger points out that there is an objective reason, namely "The peculiar 'singular limit' relation between the micro and macro descriptions of the hole-and-peg system" that "the macro description seems to tell us 'something different' than the micro description, and why the former is not reducible to the latter in a way that would support a reductive explanation. In such singular limit cases, regular perturbation theoretic approaches to solving the problem break down, and the solutions provided by singular perturbation schemes automatically introduce different levels or scales into the problem, making it clear, for instance, how the macroscopic explanation considered by Putnam manages to bring out "certain relevant structural features of the situation" which are invisible in the micro description. Explanations based on such singular limit relations thus defy the very aim of reductive explanations, viz., to give an account at a single 'basic' level (2001, p. 504).

explanation per se with reductionism. In a footnote, Epstein acknowledges that his view of reductionism is one among many: "The term "reductionist" admits a number of definitions. We are not speaking here of the reduction of theories, as in the reduction of thermodynamics to statistical mechanics." (1996, p. 56) By calling agent-based computer modeling reductionist, Epstein's intention is to avoid what he sees as the classical emergentist's commitment to mysterious gaps between macro and micro states of a system and to remain faithful to broadly physicalist ontological principles.

The important difference between traditional reductionist projects in the philosophy of science and the kinds of explanations of emergent phenomena that result from computational modeling in scientific is that the rules governing these models are constitutive of the model and its denizens independently of the connection between the model and the lower-level or more general theoretical framework. This is another example of the kind of independence of computational models. For instance, the conditions that define computational models of emergent phenomena are not equivalent to the kind of reductionist bridging laws that it was hoped could connect target phenomena or target generalizations to some lower-level ontological or theoretical framework. The difference lies in the relationship between the conditions defining the model and the lower-level theory. Specifically, the computational modeler's approach is agnostic with respect to the relationship between the configuration of simple rules/initial conditions that constitutes the simulation in question and the lower-level theory. Some lower-level theory that the reductionist would take as the base-level theory for a metaphysically satisfying reduction is not within the purview of the computational model.

Computational models are likely to satisfy our ordinary demands for explanation insofar as they provide a simulation or an analogy which permits the possibility of some measure of control over the system in question. This limited kind of explanation would leave deeper metaphysical questions about the individuation, reality and even the physical reducibility of emergent properties untouched.

Mechanism

It is possible to distinguish reductionism in its ontological and theoretical forms from a commitment to mechanistic explanation. Mechanistic explanations are central to scientific practice in the investigation of the source of some phenomenon or in the study of processes or sequences of events which have some salience. The notion of mechanism at play is far broader than the familiar push–pull model of mechanical interaction. A non-reductive account of mechanism has been articulated by Peter Machamer et al. (2000). Their presentation of mechanism focuses on what they see as the explanatory practices of biologists, however, it applies equally well to the more abstract computational models under consideration here. They argue for a view in which mechanism is understood in terms of the organization of entities and activities in ways that are "productive of regular changes from start or set-up to finish or termination conditions." (2000, p. 3) As such, they emphasize the discovery and characterization of activities in systems under investigation.

The mechanistic approach favored by these philosophers is closely related to the kinds of explanations afforded by computational models described above insofar as the mechanism under investigation has a formal character which can be discovered and described independently of a detailed understanding of the components. So, for instance, mechanism is conceived as a regular transformation from one set of states to another. Such an approach is fully compatible with the account of transition functions that Holland and others provide.

> …what makes it regular is the *productive continuity* between stages. Complete descriptions of mechanisms exhibit productive continuity without gaps from the set up to termination conditions. Productive continuities are what make the connections between stages intelligible. If a mechanism is represented schematically be A → B → C, then the continuity lies in the arrows and their explication is in terms of the activities that the arrows represent." (2000, p. 3)

CA can be understood as providing an approach to mechanism which focuses on characterizing the arrows of productive continuity in an abstract form. While advocates of the mechanistic approach to scientific explanation are eager to emphasize that their view is grounded in real case studies in the biological sciences, they provide an account which is quite consonant with more abstract reflection on the "arrows" of the kind that CA permit.

Clearly, mechanistic models of explanation are distinguishable from ontological reductionism insofar as the latter is an attempt to provide a maximally general inventory for the natural world, while the former is a far more modest response to a particular epistemic or technological need. Mechanisms can be discovered and usefully employed in science and technology without having any firm grasp on the foundational or ontological landscape.

It is important to recognize the contrast between computational models that serve modest explanatory goals and the view that scientific explanation aspires to maximal generality. The model for the maximalist approach to scientific explanation would be one which takes the basic laws and ontology of physics as providing a descriptively adequate picture of the natural world. A model of explanation along these lines is generally assumed in the metaphysical debate over the preemption of emergent properties by basic physical properties.

By contrast, Holland-style models provide explanations of putatively emergent phenomena by bracketing possible interfering conditions without being committed to any specific metaphysical framework. No solution of the metaphysical problem of emergence is assumed by computational modelers. Because the project of providing mechanistic explanations is largely agnostic with respect to basic metaphysical questions, it can provide illuminating mechanistic explanations of particular emergent phenomena.

In very practical terms, the scientific study of emergence often demands that we abandon hope of providing the kind of maximal accounts that are assumed in metaphysical debates over preemption. If we examine the kinds of accounts that we can expect from complexity theorists, we find that they rule out maximal accounts almost as a working hypothesis. Complexity theorists generally emphasize the goal of modeling the results of interactions in systems with a large number of parts. 'Largeness' (unlike, for example, "larger than") is not a well-defined property, and by using it, what complexity theorists seem to intend are systems whose number of parts is large enough that the interaction of parts leads the system to exhibit interesting macro-properties but not so large that it would be more adequately analyzed using the tools of thermodynamics (E.g. Bar Yam 1997, p. 10). Thus, the very nature of the medium-scale areas of investigation that fit the interests of complexity theorists, requires restrictions on the generality of the types of explanation that can be provided.

This should not surprise or worry us unduly. In practice, providing an explanation of the appearance of some emergent macroproperty is usually a local project which depends, in large part, on the epistemic needs of the interested parties. Broader questions concerning the reality or the metaphysical status of putatively emergent explanandum are beyond the purview of the kinds of explanations that satisfy such local epistemic demands.

Most complexity theorists begin with instances of medium scale complexity; weather, traffic epidemics etc. By contrast, a maximal interpretation would not include the kind of screening off or ceteris paribus conditions which would permit a focus on medium scale problems. Instead, the research strategy of the maximalist would involve the search for CA rules that generate an output which includes a representation which is analogous to the entire universe. In practice, rather than providing explanations of obscure cases of medium scale complexity, the maximalist goal might involve cataloguing the range of simple mechanisms and observing their impressive behaviors.[13] Put in simple terms, in the maximal case, a scientist would abandon the idea of science as the discovery of generalizations. Since generality is one of the most important differences between explanation and mere description, this should strike us as an extravagant sacrifice. In the search for a maximal model we would be involved in a cataloguing expedition, recording and organizing the outputs of simple programs in the search for patterns that resemble the natural world—a project akin to a botany of simple machines. The explanatory value of such a cataloguing expedition is minimal and has arguably already been completed. To get a sense for the problem with cataloguing expeditions, imagine some recursive function with a random output. One can be absolutely certain that there are all kinds of interesting patterns in the output of such a function, the question is, how to find them and how to determine their relevance. Such a function would be completely useless with respect to understanding or explanation and would constitute a failure in the epistemic project of scientific inquiry. In the context of computational models of emergence, especially in complexity theory, the trouble with maximal models is that they fail to provide explanations of the phenomena in question, in large part because it does not permit the kind of screening off that is necessary to detect and model these phenomena.

Acknowledgments I am very grateful for the assistance that I have had in writing this paper. Paul Humphreys, Clifford Hill and most especially Philippe Huneman have provided challenging and insightful critical comments on earlier versions of this paper. Participants in colloquia and talks at

[13] I take this to be the goal of Ed Fredkin (1990) and Wolfram (1984).

University College Dublin, the IHPST in Paris and the London School of Economics have raised many excellent points which were of great help in writing the paper.

References

Balci, O. (2003) Verification, validation, and certification of modeling and simulation applications. In *Proceedings of the 2003 Winter Simulation Conference* (pp. 150–158). Piscataway, NJ: IEEE.

Bar Yam, Y. (1997). *Dynamics of complex systems*. Reading MA: Addison Wesley.

Bedau, M. (1997) Weak emergence. *Noûs*, Supplement: Philosophical perspectives, 11. *Mind, Causation, and World, 31*, 375–399.

Burks, A. (1970). *Essays on cellular automata*. Champaign IL: University of Illinois Press.

Cariani, P. (1991). Emergence and artificial life. In C. G. Langton, C. Taylor, J. D. Farmer & S. Rasmussen (Eds.), *Artificial life II*. Redwood City, CA: Addison-Wesley.

Epstein, J. (1999). Agent-based computational models and generative social science. *Complexity, 4*(5), 41–60.

Forrest, S. (1990). Emergent computation: self-organizing, collective, and cooperative phenomena in natural and artificial computing networks. *Physica D, 42*, 1–11.

Fredkin, E. (1990). Digital mechanics. *Physica D, 90*, 254–270.

Fulford, G., Forrester, P., & Jones, A. (1997). *Modelling with differential and difference equations*. Cambridge: Cambridge University Press.

Galison, P. (1996). Computer simulation and the trading zone. In P. Galison & D. J. Stump (Eds.), *The disunity of science: Boundaries, contexts, and power*. Palo Alto: Stanford University Press.

Gardner, M. (1971). Mathematical games: On cellular automata, self-reproduction, the Garden of Eden and the game "life". *Scientific American,* pp. 120–123.

Holland, J. (1997). *Emergence: From chaos to order*. Redwood City, CA: Addison Wesley.

Hofmeyr, S., & Forrest, S. (2000). Architecture for an artificial immune system. *Evolutionary Computation, 7*(1), 1289–1296.

Humphreys, P. (1991). Computer simulation. In A. Fine, M. Forbes, & L. Wessels (Eds.), *PSA 1990* (Vol. 2, pp. 497–506). East Lansing: Philosophy of Science Association.

Humphreys, P. (2002). Computational models. *Philosophy of Science, 69*(2002), S1–S11.

Humphreys, P. (2004). *Extending ourselves computational science, empiricism, and scientific method*. Oxford: Oxford University Press.

Hu R., & Ru, X. (2003) Differential equation and cellular automata models. In *IEEE Proceedings of the International Conference on Robotics, Intelligent Systems and Signal Processing* (pp. 1047–1051).

Kim, J. (1999). Making sense of emergence. *Philosophical Studies, 95*, 3–36.

Küppers, G., & Lenhard, J. (2005) Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation 8*(4). http://jasss.soc.surrey.ac.uk/8/4/3.html.

Kuhn, R., Craigen, D., & Saaltink, M. (2003) *Practical application of formal methods in modeling and simulation*. National Institute of Standards and Technology. csrc.nist.gov/staff/kuhn/kuhn-craigen-saaltink–03.pdf. Last accessed 28 March, 2007.

Langton, C. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D, 42*, 12–37.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Merricks, T. (2001). *Objects and persons*. Oxford: Clarendon Press.

Putnam, H. (1975). *Mathematics, matter and method. Philosophical papers* (Vol. 1). Cambridge: Cambridge University Press.

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model, in Computer Graphics, *21*(4). In *SIGGRAPH '87 Conference Proceedings* (pp. 25–34).

Rosenberg, A. (2000). *The philosophy of science: A contemporary introduction*. New York: Routledge.

Rueger, A. (2001). Explanations at multiple levels. *Minds and Machines, 11*, 503–520.

Schreckenberg, M., Schadschneider, A., Nagel, K., & Ito, N. (1995). Discrete stochastic models for traffic flow. *Physical Review E, 51*, 2939–2949.

Suppe, F. (1977). *The structure of scientific theories*. Urbana: University of Illinois Press.

Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 International Congress* (pp. 252–261). Stanford: Stanford University Press.

Symons, J. (2001). Emergence and reflexive downward causation. *Principia – An International Journal of Epistemology, 6*, 14–39. 2002.

Van Fraassen, B. (1972). A formal approach to the philosophy of sicence. In R. Colodny (Ed.), *Paradigms and paradoxes* (pp. 303–366). Pittsburgh: University of Pittsburgh Press.

Van Fraassen, B. (1980). *The scientific image*. Oxford: Oxford University Press.

Von Neumann, J. (1966). *Theory of self-reproducing automata*. Champaign IL: University of Illinois Press.

Warrender, C. (2004). *Modeling intercellular interactions in the peripheral immune system.* Ph.D., Thesis Computer Science, University of New Mexico.

Wolfram, S. (1984). Cellular automata as models of complexity. *Nature, 311*(4), 419–424.