Summer 8-13-2013

In Defense of Dynamical Explanation

Shannon B. Nolen
Georgia State University

IN DEFENSE OF DYNAMICAL EXPLANATION

by

SHANNON B. NOLEN

Under the Direction of Daniel Weiskopf

ABSTRACT

Proponents of mechanistic explanation have argued that dynamical models are mere phenomenal models, in that they *describe* rather than *explain* the scientific phenomena produced by complex systems. I argue instead that dynamical models can, in fact, be explanatory. Using an example from neuroscientific research on epilepsy, I show that dynamical models can meet the explanatory demands met by mechanistic models, and as such occupy their own unique place within the space of explanatory scientific models.

INDEX WORDS: Explanation, Dynamical Systems, Mechanisms, Mechanistic Explanations

IN DEFENSE OF DYNAMICAL EXPLANATION

by

SHANNON B. NOLEN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2012

IN DEFENSE OF DYNAMICAL EXPLANATION

by

SHANNON B. NOLEN

Committee Chair: Daniel Weiskopf

Committee: George Graham

Andrea Scarantino

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2012

TABLE OF CONTENTS

1 INTRODUCTION

Mechanisms and mechanistic explanations have without question become the center of recent debates in the philosophical literature on explanation. After numerous and notorious problems with what was once considered the received view on scientific explanation—Hempel's deductive-nomological or 'covering law' model (Hempel 1965)—mechanistic explanation has come to the fore as a potential replacement model of explanation, at least within a large majority of the so-called 'soft sciences,' such as biology and neuroscience. The popularity of mechanistic models as a paradigm of scientific explanation can in part be traced to the widespread use of mechanistic models by scientists in the biobehavioral sciences, and also to the fact that mechanistic models of explanation manage to sidestep many counterexamples that plague law-based conceptions. In fact, the explosion of discussions about mechanisms and their extension to various scientific disciplines has been jokingly referred to as a sort of 'mechanism imperialism' (Weiskopf 2011a), whereby mechanistic explanation threatens to invade and conquer the explanatory domains of other kinds of models, either by calling the explanations that already exist in these domains forms of mechanistic explanation, or by denying these other kinds of models explanatory power at all.

Despite the undeniable explanatory power of mechanistic models in certain scientific disciplines, some philosophers have resisted the move to 'mechanize' all forms of explanation in the life sciences, either by returning to some modified version of law-based accounts (cf. Mitchell 1997; Leuridan 2010), or by arguing for the legitimacy of other forms of explanation that rely on different kinds of explanatory models—such as functional analysis, cognitive models, and dynamical models—in the same scientific domains in which mechanistic models are thought to reign supreme (cf. Weiskopf 2011a; Cummins

1983; Zednick 2011; Chemero & Silberstein 2008).[1] This article will join the ranks of the latter by defending dynamical models from criticisms by mechanists. The point I hope to convey is that dynamical models can meet all the normative demands on good explanations that mechanistic models meet, and as such occupy their own unique place within the space of explanatory scientific models.

In the upcoming discussion, I first review the basics of mechanistic models and their purported explanatory advantages (Section 2). I then describe dynamical models and introduce some of the debates about whether or not dynamical models are or can be explanatory. In particular, I focus on the criticisms leveled at dynamical models in a recent paper by Kaplan and Craver (2011), wherein they argue that dynamical models are not explanatory unless the variables in such models map onto underlying mechanisms (Section 3). I reveal some problems with this mapping constraint and I follow Weiskopf's (2011) treatment of explanatory models that are not mechanistic, arguing that his justification of the explanatory efficacy of cognitive models can be extended in an analogous fashion to dynamical models (Section4). I then anticipate and respond to objections (Section5), and conclude that, despite mechanists' claims to the contrary, dynamical models can and do offer genuine scientific explanations.

[1] From this point forth ,my comments on mechanisms and dynamical models can be assumed to be regarding the life sciences, most especially neuroscience and biology. I do not address whether my arguments apply to other sciences, although I expect that the explanatory import of dynamical models will in fact have similar structure in other fields that make use of such models.

2 MECHANISTIC MODELS AND MECHANISTIC EXPLANATION

Before getting into our discussion about different kinds of explanatory models, a few clarificatory remarks about models are in order. By 'model,' I mean any kind of representation that aims to describe systems or properties situated in the real world. Models can be couched mathematically by representing the world in terms of systems of equations; they can be verbal descriptions (i.e. explanatory texts) of processes aimed at, e.g., elucidating basic principles of a complex system; they can be visual or diagrammatical, as in circuit diagrams used to model conductance properties of neurons—any of these suffices to qualify as a model in the way used in this article.

Generally speaking, mechanistic models describe "how the constituent entities and activities [of a mechanism] are organized to exhibit a phenomenon" (Craver 2007). The phenomenon to be explained—called the *explanandum phenomenon*—is usually some function, or behavior 'F' of a mechanism (or more generally, of a system) 'S.' What mechanistic models explain, then, is S's (the system's) capacity to F (perform some function). Mechanistic explanation proceeds by showing *how* the explanandum phenomenon comes about in virtue of the activities of the entities in the mechanism, as well as its unique temporal, organizational, and structural properties.

An example of a mechanistic explanation is Craver's (2007) description of neurotransmitter release in a typical neuron:

> The mechanism begins … when an action potential depolarizes the axon terminal and so opens voltage-sensitive calcium (Ca^{2+}) channels in the neuronal membrane. Intracellular Ca^{2+} concentrations rise, causing more Ca^{2+} to bind to Ca^{2+}/Calmodulin dependent kinase. The latter phosphorylates synapsin, which frees the transmitter-containing vesicle from the cytoskeleton. At this point, Rab_3A and Rab_3C target the freed vesicle to release sites in the membrane. Then v-SNARES (such as

> VAMP), which are incorporated into the vesicle membrane, bind to t-SNARES (such as syntaxin and SNAP-25), which are incorporated into the axon terminal membrane, thereby bringing the vesicle and the membrane next to one another. Finally, local influx of Ca^{2+} at the active zone in the terminal leads this SNARE complex, either acting alone or in concert with other proteins, to open a fusion pore that spans the membrane to the synaptic cleft. (p.4-5)

What is being explained here is the phenomenon of vesicular neurotransmitter release in a neural cell; what is doing the explaining is the description of a mechanism—the entities, such as proteins, ions, membranes, and vesicles; their activities, such as binding, phosphorylating, and opening; and the unique temporal and organizational features of the mechanism, such as the particular sequence of events and the spatial location and morphological characteristics of the entities. The activities of the entities in the mechanism are what produce, maintain, or underlie the phenomenon in question, and so by describing the mechanism we are explaining how the phenomenon comes about (Craver 2007). This is supposed to be one of the main advantages of mechanistic explanations: they reveal the *productive* relation between the explanans and the explanandum phenomenon; that is, the explanans (i.e. the mechanism) makes the explanandum phenomenon *intelligible* (Leuridan 2010).

Proponents of mechanistic explanation can be seen as advancing a variety of claims about what constitutes a good explanation and why mechanistic models are explanatory. A few of them will be central to understanding why mechanists reject the claim that dynamical models can be explanatory, and they are summed up by the following statements (Kaplan & Craver 2011):

(1) Explanations are adequate to the extent that they describe causal mechanisms that maintain, produce, or underlie the explanandum phenomenon.[2]

(2) Mechanistic models have explanatory force in virtue of the fact that they describe causes and mechanisms that produce or underlie the phenomenon.

(3) To describe causal mechanisms, one must describe component parts, their relevant properties and activities, and how they are organized together causally, spatially, temporally, and hierarchically. (605)

These claims form the foundation of the mechanistic paradigm. When appropriately qualified, they assert that a good explanation describes causal mechanisms, that a model that describes causal mechanisms is *ipso facto* explanatory, and that to adequately describe causal mechanisms one needs to know and understand the components of the mechanism, their activities within the mechanism, and the organization of these components and their activities. Thus a good mechanistic model will describe the causal mechanism responsible for the explanandum phenomenon without leaving out any causally relevant components or activities; it will specify the initial and termination conditions of the mechanism; and it will likely also provide information that tells us how it is (often hierarchically) situated in its environment. In addition, Craver and other mechanists (cf. Craver 2006 and 2007; Bogen 2008) claim that one of the main advantages to a good mechanistic model—and also, they argue, an outcome that provides evidence that we are doing genuine explanatory work—is that it will reveal the "knobs and levers" that we may take advantage of to manipulate and control the system.

[2] This claim is intended for the domain of cognitive and systems neurosciences specifically, and the biobehavioral sciences more broadly.

There are a few caveats here worth noting. First, a given model can be a model *of* a mechanism without being a mechanistic model. Presumably, many dynamical models are models of mechanisms (particularly when they are employed in biology and neuroscience), but this does not make dynamical models mechanistic models. What makes a model qualify as a certain kind of model is how the model represents its target, not what the target is *per se*. Mechanistic models are therefore models that represent real-world mechanisms or systems *as* mechanisms, whereas, e.g., a dynamical model might be representing a mechanism *as* a dynamical system. Put another way, all mechanistic models are models of mechanisms, but not all models of mechanisms are mechanistic models.

Second, while mechanistic explanation may be useful for a variety of explanatory goals, most mechanists are careful to argue that mechanistic explanation is most appropriate when the explanandum phenomenon is some system's capacity to perform some function or behavior (i.e. S's capacity to F). Thus the proper targets for mechanistic explanation include such things as a neuron's ability to release neurotransmitters, a lamprey's ability to swim, a virus's capacity to self-replicate, or the capacity of a neural network to generate waves of excitatory activity. It should be fairly obvious from this that many explanations in the life sciences are going to take the form of mechanistic explanation, in no small part because of the kinds of phenomena these sciences seek to explain.

In sum, mechanists argue that mechanistic models are explanatory because they describe the causal mechanism underlying the explanandum phenomenon, and they do this in an intelligible way. Mechanists thus maintain an implicit (though often explicit) commitment to the idea that the only way to explain a phenomenon is to describe the underlying causes which support it. Any putative explanatory model that fails to describe causal mechanisms in the ways mechanistic models do therefore falls short of providing a genuine explanation on this account. In the next section, we will see how some mechanists have applied this critique to dynamical models.

3 DYNAMICAL MODELS AND THEIR CRITICS

Dynamical models can be found across a wide range of sciences, from physics, where they are utilized extensively in thermodynamics and statistical mechanics (Batterman 1998; Stanley 1999); to computational neuroscience, where they illuminate various timing and behavioral properties of individual neurons as well as entire neural networks (Rabinovich et al. 2006; Izhikevich 2007); and even to the nutritional and health sciences, where they have recently been used to model weight gain and changes in body mass over time in efforts to better understand obesity (Chow & Hall 2008). The utility of dynamical models comes not only from the flexibility inherent to using mathematical formulas to represent a virtually limitless variety of system properties, but also from the geometric tools that are a key feature of qualitative dynamical systems analysis, tools that can be used to help us visualize complicated system behavior and make more intuitive mathematical information that would otherwise be opaque.

Generally speaking, dynamical models are mathematical models that are composed of a set of one or more differential or difference equations that contain variables and parameters that stand for different properties of the system being modeled. One of the key features of dynamical models is that they explicitly track the evolution of system variables over time, often with an extremely high level of detail, keeping track of multiple variables and parameters and their mutual influence on one another as the system moves through time. In addition to their general modeling capabilities, dynamical models can be assessed using qualitative dynamical systems analysis, which uses the geometric properties of dynamical systems to predict and organize different kinds of dynamical activity.[3] Dynamical models are thus indispensible for modeling complex nonlinear processes as well as for dissecting subtleties involved in transient or rhythmic processes. Given that transient and rhythmic processes abound in neural systems, dynamical models are essential in computational neuroscience and related fields (cf. Rieke et al. 1997; Rabinovich et al. 2006; Izhikevich 2007).

[3] See Strogatz (1994) for a useful introduction to these methods.

According to Zednick (2011), the received view on dynamical explanation is that it is a special case of the well-known *covering law explanation*, originally proposed by Hempel (1965). Covering law explanations proceed by deriving the explanandum phenomenon from the conjunction of a law of nature (loosely understood as a non-accidental counterfactual-supporting generalization) and some set of initial conditions.[4] Dynamical explanation is thought to be organized in exactly this way: the future state of a dynamical system is predicted (or derived) from a set of differential equations and the initial values of the variables and parameters that characterize that system. Thus most have conceived of dynamical explanation as a form of covering-law explanation, with the unfortunate consequence that the well-known flaws of the latter will, by analogy, be extended to the former (Zednick 2011; Kaplan & Craver 2011).[5]

An example of this sort of critique of dynamical explanation can be found in a recent article by Kaplan and Craver (2011). Kaplan and Craver argue that, at least within the domain of systems and cognitive neuroscience, dynamical models explain a phenomenon only when "there is a plausible mapping between elements in the model and elements in the mechanism for the phenomenon" (p. 601).[6] Otherwise, they argue, dynamical models are mere re-descriptions of the phenomenon under study, a sort of 'mathematical biography' that summarizes observable features of a system without explaining *how* those features come about. Their accusation is that unless dynamical models can meet their "3M constraint" (a model-to-mechanism-mapping), such models are *phenomenal* in nature—i.e. their merit does not go beyond empirical adequacy and/or predictive success. Thus dynamical models are phenomenal

[4] The fact that one can *derive* the explanandum in the covering law model is the basis of the association between *prediction* and *explanation*—if we can predict the occurrence of a phenomenon, then we have shown some basic understanding of it on this account.

[5] It turns out that in practice at least, simulated solutions via numerical methods occur more often than analytical solutions. This may have implications for the validity of this conception of dynamical explanation, at least if it is supposed to adequately serve as a model for how scientists *actually* use dynamical models to explain phenomena.

[6] Presumably, Kaplan and Craver would extend this mapping criterion to other domains of science, at least those in which systems can plausibly be interpreted as being mechanisms. My arguments should apply to those domains as well.

models that simply re-characterize the phenomenon—and unlike mechanistic models, they don't explain *how* the phenomenon comes about.

The main line of argument Kaplan and Craver offer against dynamical explanation consists of comparing dynamical explanation to covering law explanation, given the similarities of their structures noted above. The covering law model is notorious for certain objections that have caused it to fall out of favor among contemporary philosophers, not the least of which is the contention that prediction is not sufficient for explanation. The basis for this claim rests on a series of counterexamples that are meant to appeal to intuitions about what is or is not an explanation. To borrow from some common examples that Kaplan and Craver appeal to, a barometer reading may predict the presence of a thunderstorm, but the barometer reading does not *explain* the thunderstorm. A sputtering engine may predict an empty gas tank, but the engine sputtering does not *explain* the empty gas tank. The length of the shadow cast by a flagpole and the angle of the sun may predict the height of the flagpole, but the length of the shadow and angle of the sun do not *explain* the height of the flagpole. What these examples are meant to show is just that our intuitive judgments about prediction and explanation may come apart, and when this happens, it seems to suggest that prediction is not sufficient for explanation.[7]

In addition, Kaplan and Craver note that while explanatory models suffer from the inclusion of irrelevant detail, predictive models do not—irrelevant details do not alter a model's ability to generate accurate predictions (2011). The example they offer is that of the classic physics model that predicts the

[7] These counterexamples are not without problems themselves, and I am not convinced they completely overturn the notion that prediction can be enough for explanation, at least in some cases. The simplistic, everyday nature of the examples only suffices to show that *these* kinds of predictive correlations are not enough to qualify as genuine explanations. However, these examples look very different from the kinds of complex predictive models used in the sciences, models that are perhaps fitted to one dataset and then shown to be predictive for a novel one (e.g. Hodgkin and Huxley's equation for the action potential, various economic models, etc). At best the analogy is strained.

Also, the sputtering engine case is not a genuine counterexample, since the prediction it makes is about what best explains the sputtering of the engine, and therefore what is most likely to the be the cause. Inferring from a sputtering engine to an empty gas tank is a case of inference to the best explanation, and so it cannot serve to help guide our intuitions about when a prediction does or does not constitute an explanation. The prediction in this case is *about* an explanation.

period of a pendulum from the length of the rod and its initial velocity. The prediction of the pendulum's period is the same, even if we include details about what color the pendulum is and what it is made of. Of course these details aren't relevant to determining the period, but they nevertheless do not detract from the predictions made by the model. Thus Kaplan and Craver conclude that since explanatory models and predictive models differ with respect to individual outcomes of including irrelevant detail, explanatory models and predictive models are subject to distinct epistemic norms. And if explanation and prediction are governed by different norms, then explanation cannot be (merely) prediction (p.607).

All of this is relevant to dynamical models in particular, because one of the hallmarks of dynamical models is their predictive power. Given information about initial conditions and parameter values, a dynamical model can predict the future state of the system with a high degree of accuracy. But it is part of the mechanist's critique that 'mere prediction' is not sufficient for explanation, because one can derive the explanandum phenomenon without knowing exactly *how* it came about, which they argue is what only the mechanistic story can provide. This is the basis for Craver's claim that the Hodgkin and Huxley (HH) model of the action potential is a dynamical model that, despite generating highly-accurate predictions, does not *explain* the action potential (Craver 2007 and 2008). (However, this point is up for debate; see Weber 2008 for an alternative interpretation; see Bogen 2008 for continued discussion). Craver argues that the only thing that can explain a neuron's ability to produce an action potential is a mechanistic model that describes all and only the relevant entities and activities that together produce an action potential. However successful the HH formalism has been for predicting the time course of action potentials, it doesn't tell us how a neuron generates one in the first place, and thus it does not explain the action potential on a mechanist's account. As a result, Kaplan and Craver claim that "dynamical models are not part of or an alternative to mechanistic explanations; at best they are descriptive tools for representing how complex mechanisms work" (p.602).

As a consequence of the above claims, Kaplan and Craver propose that the *only* occasion in which dynamical models can be explanatory is when such models meet the 3M constraint. Kaplan and Craver stipulate 3M as follows (Kaplan & Craver 2011):

> (3M) In successful explanatory models in cognitive and systems neuroscience (*a*) the variables in the model correspond to component activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (*b*) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism. (611)

It should be clear from the above that while mechanistic models will inherently meet 3M, many dynamical models will not. In fact, on closer inspection of 3M, one finds that it essentially says that a model is a successful explanatory model if it is a mechanistic model, since *(a)* and *(b)* are *the* characteristic features of mechanistic models. Only if dynamical models map to underlying mechanisms can they be considered explanatory on this account. Under their view, all dynamical models (as well as computational or functionalist models) that embody no commitment to the underlying structure which gives rise to the phenomena are descriptive models. And by merely describing the phenomena (i.e. by being *phenomenal* models), they do not explain it, or so the argument goes. The following quote from Kaplan and Craver summarizes their view succinctly:

> Dynamical models do not provide a separate kind of explanation subject to distinct norms. *When they explain phenomena, it is because they describe mechanisms*. As descriptive tools, they can be used to describe mechanisms phenomenally or mechanistically, correctly or incorrectly, and completely or incompletely. (618; my emphasis)

Mechanists are not at all unclear about what does the explaining when the phenomenon to be explained is S's capacity to F. Mechanistic models explain S's capacity to F because they describe the causal mechanism which sustains that capacity. By contrast, dynamical models are merely descriptive of phenomena, ways of characterizing the phenomenon in compact mathematical form—unless, that is, there is some plausible mapping between the variables in the model and the underlying mechanism. When such a mapping exists, dynamical models are describing mechanisms, and the greater extent to which they are able to describe mechanisms (i.e. correctly, mechanistically, and completely), the greater their explanatory power. This is the view that Kaplan and Craver defend, and it is the view I criticize in the next section.

4 HOLDING THE LINE AGAINST MECHANISTIC IMPERIALISM

Kaplan and Craver's views on dynamical models make it clear why some philosophers have thought of the mechanistic tradition as increasingly imperialistic—in the case of dynamical models, the 3M constraint essentially requires dynamical models to be models of mechanisms in order for them to be considered explanatory. In this section, I challenge Kaplan and Craver's views on dynamical models and the 3M constraint by providing an example of a dynamical model that fails to meet 3M and yet nevertheless is explanatory.

As a reminder, the 3M constraint—the requirement that dynamical models map onto underlying mechanisms in order for them to be considered explanatory—is at bottom an assertion about the explanatory primacy of mechanistic models, because it requires models in neuroscience to do what mechanistic models do in order to achieve any kind of explanatory adequacy. Thus dynamical models, cognitive models, and functional-analytic models all will be deemed explanatorily inadequate unless they properly map and have specified commitments to an underlying mechanism. It's important to note that 3M is not an argument that only mechanistic models can be good explanations—it's an *assertion* to that effect. Of course, Kaplan and Craver provide examples of dynamical models that fail to meet 3M, such as Kelso's HKB model of bimanual coordination, and they rely on intuitive judgments that such models are phenomenal and not explanatory to support the implementation of 3M. But 3M is only presumed to be an arbiter of good explanations in neuroscience because it relies on the presupposition that good explanations in neuroscience are mechanistic explanations. If one is not immediately sold on this idea, then one must dig further to uncover what is supposed to be the thing that makes describing causal mechanisms the *only* way to explain a phenomenon.[8]

[8] The issues here are deep and preclude thorough treatment in this article. If mechanistic explanation is a kind of causal explanation, then it should be clear that our operative notion of cause is paramount. Many of the issues I discuss from this point on in the paper will be affected by the theory of causation one adopts.

Without independent justification, 3M begs the question of explanatory adequacy in assuming that an adequate explanation of a phenomenon has to be a mechanistic explanation. To say that mechanistic explanation is superior over dynamical explanation because only the former describe entities, activities, and organizational features that produce a phenomenon is just to say that mechanistic explanation is superior because it is mechanistic explanation. Of course, mechanists argue that the virtues of mechanistic explanation justify these claims and circumvent the apparent circularity of their arguments. But though the explanatory virtues of mechanistic models are many(cf. Bechtel 2011; Craver 2007 and 2008), other kinds of models can meet these standards as well. Dynamical models, for example, also have numerous explanatory virtues, and can provide genuine explanations in certain contexts, as I'll argue below. If this is right, then what is needed is a way of evaluating the epistemic virtues of explanatory models *independently* of their specific form (i.e. mechanistic, dynamical, cognitive, functional-analytic, etc). Explanatory models can then be evaluated on dimensions of assessment that do not presuppose that a given type of model is superior. Evaluating dynamical models under this neutral framework then becomes key in showing that dynamical models can perform as mechanistic models do in terms of their meeting normative requirements on good models in general, and more importantly, in terms of their being explanatorily adequate.

In a recent paper, Weiskopf (2011) argues that, despite its surge in popularity, mechanistic explanation should not be seen as the only viable form of scientific explanation, and further that other kinds of models can meet the same normative requirements on good models that mechanistic models do. Weiskopf extracts two central dimensions of epistemic evaluation from Craver's (2007) treatment of the norms governing mechanistic models:

> (1) degree of evidential support of the model (what Craver calls "degrees of realism"); and

(2) degree of representational accuracy of the model, evaluated in terms of grain size and correctness.

As Weiskopf rightly points out, these dimensions are equipped to evaluate all types of models, and not just mechanistic ones (Weiskopf 2011). While Weiskopf uses these dimensions of normative assessment to evaluate various cognitive models, I will use these dimensions to evaluate dynamical models. And since these norms were extracted from those used by mechanists to promote mechanistic explanation, they serve as fair arbiters of dynamical models. In the remainder of this section, I show that dynamical models can meet *all* the requirements on good models that mechanistic models meet, without themselves being describable in mechanistic terms (i.e. without being considered mechanistic models themselves). To do this, I will need to show that dynamical models can be highly confirmed models, which speaks to degrees of evidential support, and that dynamical models can be representationally accurate, in that they correctly describe or represent those features of the target system that they aim to model.

To illustrate how dynamical models can meet these epistemic norms and in addition be considered explanatory, it will be useful to work with an example. Here I present a dynamical model that is based on models of neural networks aimed at understanding epilepsy in the brain.

Various kinds of computational models have recently been used to simulate epileptiform behavior in neural networks, with the long-term aim of understanding how certain network and neural properties can give rise to epileptic seizures (Netoff et al. 2004; Ursino et al. 2006). Because epileptiform activity in the brain seems to arise from patterns of connectivity and related network-level properties, researchers have been most interested in modeling epilepsy at the level of neural networks, trying out various kinds of networks and varying parameters such as synaptic strength in order to simulate the seizing and bursting activities characteristic of epilepsy.

Netoff and colleagues wanted to explore how epileptiform behaviors (i.e. seizing and bursting) in the hippocampus may be a function of the general connectivity of the neural networks in which they occur. Given the fact that bursting activity originates in the CA3 region of the hippocampus and seizing originates in the CA1, Netoff and colleagues hypothesized that certain sorts of connectivity patterns could explain the differences in network behavior seen in these two regions in previous electrophysiological studies. More specifically, they hypothesized that: (1) the CA3 exhibits bursting because it has more long distance neural connections than CA1, thus allowing for the recurrent excitation that is the signature of network bursting; and (2) the CA1 exhibits seizing activity due to fewer long distance neural connections and less recurrent excitation, which allows activity to spread throughout the network more slowly, which in turn ensures the existence of a sustainable supply of excitable CA1 neurons.

To test these hypotheses, the Netoff model (NM) simulates networks of excitatory neurons with a small-world network pattern of connectivity, which is characterized by mostly local connections and a few random long-distance connections.[9] As a result of this topology, NM can be characterized by just three parameters: the number of nodes (N); the proportion of long-distance connections that are randomly rewired (r); and the proportion of nodes to which each neuron synapses (k).

While the majority of detail need not concern us here, there a few things to note about the NM. First, the dynamical model is built on a general network model, which is composed of a specific number of nodes (neurons), long-distance connections (neurons that synapse out of their own immediate 'neighborhood'), and a proportion of the total number of neurons to which each neuron synapses. Each node in the network model represents a neuron, and the activity of the neuron is represented by an ordinary differential equation (one differential equation per neuron in the network), which then essentially transforms the network model into a stochastic dynamical model. Netoff and colleagues then tested several different kinds of neuron models (i.e. different kinds of differential equations standing in for

[9] Small world networks are the behind the famed 'Six degrees of separation' phenomenon, known to some as 'Six degrees of Kevin Bacon.'

each of the neuron nodes), to show that the behavior of the network and the effect of the connectivity pattern did not depend on the kind of neuron model used—network behaviors of bursting and seizing occurred strictly as a result of the connectivity of the network, and not as a result of the details of the individual neurons. This led to one more iteration in the NM development, as once the stochastic model showed that only network connectivity mattered, Netoff and colleagues were able to reduce the model to a discrete-time dynamical system with only one or two dimensions. Reducing the model in this way—a major advantage of dynamical systems modeling—allowed them to use qualitative dynamical systems analysis to understand how waves of excitatory activity propagated across the network (Netoff et al. 2004).

The NM verified the predictions of the scientists: just by changing the connectivity parameters in the model, networks could be induced to exhibit both bursting and seizing behaviors. Previously gathered data suggested a high degree of excitatory connections in the CA3, which is the location in the hippocampus known to exhibit bursting behavior. Tuning the parameters of the NM so that the connectivity of the network mimicked that of increased excitatory connections caused the network to exhibit bursting activity, consistent with experimental preparations of CA3. Tuning the parameters to mimic connectivity observed in CA1 also produced the seizing activity characteristic of CA1. In their own words, the dynamical model "explains how specific changes in the topology or synaptic strength cause transitions from normal to seizing and then to bursting" (8075). NM is thus able to offer an explanation for a specific phenomenon—namely, that different kinds of epileptiform activities originate in different regions of the hippocampus (i.e. CA1's capacity to seize and CA3's capacity to burst)—and its results are robust across at least three different neuron models.

Assessing the NM in terms of degree of representational accuracy is fairly straightforward. As a reduced 1-D model of wave propagation through a network, the NM certainly leaves out a variety of details about the system under study. But as Weiskopf (2011) notes, and the robustness of NM across

three different neuron models shows, such details are irrelevant to the modeling purposes of NM *and* to its ability to explain network behavior. Despite the fact that the model is a reduced model with minimal parameters to constrain it, the NM is able to accurately represent the connectivity patterns in both the CA1 and the CA3 and produce the associated epileptiform behaviors that occur in these regions of the hippocampus. What the NM represents it does so accurately, and details about the individual neurons in the network prove to be irrelevant and so are not failures on this dimension of normative assessment.

Assessing the NM on degree of evidential support is somewhat more complicated, as it can be difficult to interpret what constitutes evidence when simulations are involved particularly because we often simulate systems that are not transparent to us. Nevertheless, it's at least in principle possible, and in many cases, very likely, that dynamical models can be confirmed by data they accurately predict. In the NM example, the model *does* predict the appropriate shifts found in excitatory networks between normal, bursting, and seizing behavior. And we might also say that, had we not know what kinds of epileptiform activity the CA1 and CA3 regions would exhibit, but we did know their connectivity patterns (i.e. sparse or dense), then we *could* have used NM to predict the typical kinds of epileptic activity that is characteristic of each of these regions.

As a reminder, the above two dimensions of model assessment are meant to place the NM on a continuum that stretches from weak to strong evidential support, and from low to high representational accuracy—they do not (nor do they aim to) determine whether or not a given model is *explanatory*. So it's important to note that while a good model may do well in terms of representational accuracy and evidential support, it may not turn out to be explanatory for a given phenomenon (e.g. Kelso's HKB model of bimanual coordination). Conversely, a model that is explanatory for some phenomenon may

not be well confirmed or may employ filler terms that prevent the model from being considered a good model on these dimensions of assessment.[10]

Now, it should be clear from the above that the NM does well on the normative dimensions of model assessment, so the question now becomes whether or not NM is explanatory. Recall that the explanatory target for mechanistic models is some system's ability to perform some function or behavior (i.e. S's capacity to F). In order for dynamical models to compete for explanatory adequacy in the same domain as mechanistic models, dynamical models need to have the same kinds of explanatory targets as mechanistic models (if their respective explanatory targets differ in kind, then there is no need for comparing their putative explanatory status). In the NM, the explanatory target is the capacity of the hippocampus to exhibit differential epileptiform network behaviors in its CA1 and CA3 regions (seizing and bursting, respectively). A proper explanation of this phenomenon then, would involve at minimum some description of *how* different bursting and seizing patterns come about and/or *what* features of the system are responsible for giving rise to the bursting and seizing patterns we observe. And the NM gives precisely this sort of explanation of the bursting/seizing phenomenon in epilepsy. The connectivity of the network and the resulting dynamics *explain* the different epileptiform capacities of the CA1 and CA3 regions: by describing the network dynamically, the NM reveals what features of the system are responsible for giving rise to bursting and seizing (and the transitions between them). Moreover, the fully articulated model shows—by appeal to these crucial system features—*how* the different patterns of bursting and seizing come about from, e.g., the interplay in network topology and synaptic strength. At this more intuitive level of evaluating what constitutes an explanation of a phenomenon, the NM appears to do the same sorts of explanatory work as mechanistic models do—it shows what system ele-

[10] Of course, this is not to say that a model that does poorly on the two normative dimensions of model assessment has no value or utility whatsoever—it may just be a fledgling model, equivalent to what Craver terms a 'mechanism sketch'(2007). Rather, the claim is that the more well-confirmed and representationally accurate the model is the better that model will be at doing what models do—providing a true representation of the target phenomenon that helps make its workings intelligible to us.

ments are important for generating the explanandum capacity and how those elements interact to produce it. From this perspective, it seems perfectly natural to call the NM an explanatory model for the different epileptiform capacities of the hippocampus.

In addition to being considered a bona fide explanation in general, what is supposed to be characteristic of good explanatory models is that they are able to answer a range of counterfactuals and have the potential to help us manipulate and control the system they model.[11] Being able to answer questions about what would have happened had things been different is the hallmark of an explanatory model, and the NM (and no doubt many dynamical models in general) will be able to do this. Given different kinds of network connectivity, NM can answer how the network will behave if we were to change such parameters. In addition, it predicts the kinds of activity we might see in networks with properties different from the CA1 and the CA3. (Note that the NM is able to do this *without* being committed to exact mechanistic details of the neural substrates underlying the pattern—more on this below.)

At this point, mechanists might raise the objection that the NM, if it meets the necessary criteria for explanatory adequacy, does so because it satisfies 3M. That is to say, the parameters in the model *do* in fact correspond to real features of the underlying system, and if this is right, then NM meets 3M, and whatever explanatory status it might have attained is just a result from its describing the appropriate neural mechanism. The very fact that the connectivity parameters were tuned to approximate observations in the CA1 and CA3 seems to suggest that this is in fact the case.

This objection is well-taken, and it points to some crucial issues regarding what exactly makes something a mechanism, and further, what makes something a *mechanistic model*. First, on the basic

[11] In the absence of any uncontroversial theories of what an explanation *qua* explanation actually *is,* it seems reasonable to rely on our intuitive judgments of when a phenomenon has been explained by a putative model, as well as the maxim "explanations are as explanations do." In other words, good explanations will answer a wide range of counterfactuals about the explanandum phenomenon, and they may also reveal ways to manipulate and exploit the system giving rise to the phenomenon (presumably, the latter here is a result of the explanation involving *causal* features of the system). If a given model is able to do these things, then we might well consider that model explanatory. These issues are obviously crucial to what will ultimately determine whether or not dynamical models are explanatory, but space in this article prohibits further analysis. I leave the reader to his own judgments about my suggestions here.

understanding of mechanism, mechanisms are at bottom both componential (i.e. composed of localizable entities) and causal (i.e. the components in the mechanism perform activities that are the constitutive causes of the phenomenon). What makes something a mechanistic model, besides the fact that it represents a mechanism *as* a mechanism, is the unique kind of one-one mapping between components and activities in the model, and components and activities in the mechanism. Because any kind of model can potentially be *of* a mechanism, the key feature of mechanistic models is that they represent mechanisms as mechanisms, and doing this requires that each part of the mechanism has a corresponding part in the model, *mutatis mutandis* for each (causally relevant) component activity. Therefore in order for the NM to meet 3M and thus effectively become a mechanistic model, it would need to map part-for-part to the underlying system in a direct and systematic way.

The NM, however, contains no such mapping. While it may be true that the connectivity parameters correspond to characteristics of the neural network, this is not enough to be considered a one-one mapping of model to mechanism. What the NM represents is *not* the components of the underlying neural system—e.g. the specific neurons, their specific synaptic properties, the dominant pathways among them, etc. (refer to mechanistic explanation given in Section 2)—but rather a few key *quantities* that represent statistical properties of the neural system. The difference here is subtle, but crucial. The total *number* of nodes, the *proportion* of total nodes synapsed on, and the *proportion* of long distance connections made are the key parameters that define the NM and that are jointly sufficient to bring about or produce the differences seen in neural firing in CA1 and CA3. These parameters represent *statistical* features of the underlying system, and this fact blocks NM from meeting 3M in two ways.

First, the amount of connectivity in a network is, again, a *quantity* that characterizes the system, and quantities are not 'parts' of a mechanism, properly speaking. Parts of a mechanism are in essence the entities doing the activities. When each causally relevant entity in a system (and its associated activity) are represented in a model that aims to explain a capacity, that model is said to be mechanistic. Re-

call that Craver's mechanistic example of a neuron's capacity to release neurotransmitters makes extensive use of biological entities jostling around in such a way that they give rise to the vesicular release of neurotransmitters in a neuron (2007). Contrasting this case with that of the NM, it is at the very least a stretch to allow the connectivity parameters to qualify as entities performing activities in the same sort of way. Yet this is what is required if NM is to meet the 3M constraint.

Second, in order for NM to be mechanistic, the relationship between parts in the system and parts in the model needs to be direct, i.e. one-to-one. A mechanistic model represents all causally relevant components in a mechanism, their associated activities, and any other organizational features that may be important to the mechanism's operation. To achieve this, the model must represent each properly circumscribed part *as that part*, which requires that no two distinct and causally relevant parts in the actual mechanism are lumped together or are otherwise not differentiated in the model of that mechanism. But the mapping from the statistical quantities in the NM to the underlying neural system is indirect—any number of different *actual* underlying neural configurations (that would in fact be causally responsible for epileptiform behavior in the hippocampus) might generate the appropriate statistical quantities that do the explanatory work *in the model*. In other words, a variety of actual neural configurations are compatible with a given set of parameter values, which makes the mapping from each connectivity parameter to the underlying neural system far more indirect than what mechanistic models require—such a mapping does not under any reasonable interpretation qualify as one-to-one. This, in combination with the point that quantities are not properly understood as parts, makes it difficult for mechanists to argue for the claim that NM meets 3M, and that *this* might be the reason for its explanatory status.

Even if the NM fails to meet 3M and yet is considered explanatory, mechanists may still object that NM is just what Craver would call a 'mechanism sketch'—i.e. the NM is aiming to fully model a mechanism, but falling short. On Craver's view, a mechanism sketch is a model of a mechanism that is

incomplete in the sense that it leaves gaps and employs black boxes or filler terms for entities and processes that we do not yet know enough about (2007). The idea is that as we learn more about the system under study, we can flesh out the model, filling in missing details and cashing out black boxes in the model. If the NM turns out to be a mechanism sketch, then mechanists can argue that its failure to meet 3M is a result of its being a mechanism sketch, and that further research will expand the NM into the territory of being a complete mechanistic model, in which case it *would* be capable of meeting 3M. But a close look at the NM shows that this cannot be the case. NM does abstract away from many of the details at the neural level, and it omits mechanistic details that are not relevant to explaining the capacity of the CA1 and CA3 to seize and burst. However, it does *not* employ any filler terms or black boxes; nowhere in the model is a term that is meant to be cashed out once we learn more about the system, and the NM has no free parameters either. With respect to its intended explanatory goal, the NM is complete; it contains all and only those parameters crucial to generating the key behavior of the system, and it leaves no gaps that might be filled in by future research. Without gaps, black boxes, and filler terms, the NM cannot plausibly be considered a mechanism sketch, and so this objection fails to be persuasive.

What the NM and the above discussion shows is that at least some dynamical models, despite being non-mechanistic, and irrespective of whether or not they map to underlying neural mechanisms, are able to achieve representational accuracy, evidential support, and answer a range of counterfactuals about the phenomena they explain. If such models are able to achieve what mechanistic models can when measured against these criteria, there is no reason to think that they cannot be explanatory or that mechanisms must be involved in order for them to be explanatory.

Now, at this point, someone might raise an objection like the following: “It may be true that dynamical models are able to explain certain features of a dynamical system, but what explains the dynamics of the system itself is going to be some kind of mechanistic story, or at the very least a causal explanation involving the components of the system. If there is a causal explanation for the dynamical

properties of the system, then anything explained by the dynamical properties will be (in principle) explainable by the components in the causal explanation, rendering the dynamical properties superfluous to any genuine explanation. In other words, if a physical mechanism explains the dynamics, then the dynamics cannot in turn do any explanatory work, or if they do, it is only because they provide a convenient shorthand for what is actually going on at the physical level. The dynamical explanation does not stand on its own."

I see this objection as closely related to issues regarding the autonomy of special science explanations, and more generally, to questions about the transitivity of explanation.[12] It seems that mechanists want to say that if a mechanistic model explains the occurrence of some phenomenon, then a dynamical model of the same phenomenon cannot in turn be explanatory, in large part because the mechanistic model presumably explains why the phenomenon has the dynamics that it does. In other words, the claim seems to be that if A explains B (i.e. the mechanism explains the dynamics), then B is not or cannot be explanatory.

There is no reason to think that explanation is a transitive relation, because often what does the explanatory work are factors that are highly contextual, as mechanists are ready to admit. If A explains B, and B explains C, there is no *a priori* reason to assume that the context of A and the context of C are similar enough for A to be an adequate explanation of C. The very existence of the special sciences speaks to the autonomy of macro-level explanation (Fodor 1974), and we ought not automatically assume that because physical mechanisms explain the dynamics of a system, that those physical mechanisms themselves can explain a particular bifurcation, phase transition, or other dynamical phenomenon. Any mechanist who claims that because the mechanism explains the dynamics, the dynamics cannot in turn be explanatory opens herself up to the same criticism—namely, that since certain regularities

[12] Thanks to Dan Weiskopf for directing my attention to this point; and also to Elliot Sober, who later pointed out to me the absence of a proper discussion of this issue.

in the world may explain the existence of mechanisms, mechanisms themselves would not be explanatory.

Dynamical models, like the NM model discussed above, often will omit lots of detail in exchange for understanding what the crucial variables in a system are, but this need not detract from the accuracy with which they may represent the phenomena they target. Such models will also often have a high degree of evidential support, as did the Hodgkin and Huxley model when it was originally proposed. And since dynamical models track change over time, whether that change is continuous or discrete, they provide a wealth of information about what would happen if things had been different. The details of dynamical explanation remain to be fleshed out on another occasion, but the above discussion should prod us into taking a closer look at dynamical models and the explanatory potential they have to offer.

5 CONCLUSION

In the biological sciences, the phenomena under study tend to be the behavior of complex systems with fairly structured internal organizations (often hierarchically organized, spatially and temporally) and full of autonomous or semi-autonomous sub-components, each with their own relevant processes, etc. Mechanistic models are indispensible in getting explanations of biological phenomena off the ground, and often the assumptions made in dynamical models depend crucially on details gleaned from a mechanistic understanding of the relevant processes. But when we want to know more about a system, how differences in the amount and rates of change in the various entities and activities contribute to the occurrence of the overall phenomenon, dynamical models can provide deep insights into the structure of a system and how it evolves through time. To belabor points about needing a mapping to exist for a dynamical model to be explanatory is to misunderstand what dynamical models are attempting to explain and how they contribute to an overall understanding of the phenomena under study. By trying to make dynamical models into mechanistic models, some mechanists miss the unique information dynamical models provide and conflate the explanatory targets of these two kinds of models. I have argued herein that dynamical models can sometimes be explanatory, in virtue of the fact that they meet all the norms for explanatory models that mechanistic models do.

I do not want to suggest that dynamical models are superior to mechanistic models or that mechanistic models are overrated. On the contrary, it seems clear that science benefits most from a rich interaction between both kinds of models, each providing separate but unique explanatory information that, when taken together, gives scientists the deepest possible understanding of the phenomena they seek to explain.

6 REFERENCES

Anderson, H.K. (2011). Mechanisms, laws, and regularities. *Philosophy of Science, 78*(2), 325-331.

Auyang, S.Y. (1998). *Foundations of Complex-System Theories: In Economics, Evolutionary Biology, and Statistical Physics.* Cambridge: Cambridge University Press.

Batterman, R.W. (1998). Why equilibrium statistical mechanics works: Universality and the renormalization group. *Philosophy of Science*, *65*, 183-208.

Batterman, R.W. (2000). Multiple realizability and universality. *British Journal for the Philos ophy of Science, 51*, 115-145.

Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science, 66*, 261-274.

Bechtel, W. (2009). Generalization and discovery by assuming conserved mechanisms: Cross-species research on circadian oscillators. *Philosophy of Science, 76*, 762-773.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science, 78*(4), 533-557.

Block, N. (1997). Anti-reductionism slaps back. *Philosophical Perspectives, 11*, 107-132.

Bogen, J. (2008). The Hodgkin-Huxley equations and the concrete model: Comments on Craver, Schaffner, and Weber. *Philosophy of Science, 75*(5), 1034-1046.

Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of Science, 76*, 1-27.

Chow, C.C., & Hall, K.D. (2008). The dynamics of human body weight change. *PLOS Compu tational Biology, 4*(3), 1-11.

Craver, C.F. (2006). When mechanistic models explain. *Synthese, 153*, 355-376.

Craver, C.F. (2007). *Explaining the Brain*. Oxford: Oxford University Press.

Craver, C.F. (2008). Physical law and mechanistic explanation in the Hodgkin and Huxley model of the action potential. *Philosophy of Science, 75*(5), 1022-1033.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge: MIT Press.

Fodor, J. (1974). The disunity of science as a working hypothesis. *Synthese, 28*(2), 97-115.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Per spectives, 11*, 149-163.

Funkhouser, E. (2007). A liberal conception of multiple realizability. *Philosophical Studies, 132*, 467-494.

Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Izhikevich, E.M. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* Cambridge: MIT Press.

Jackson, F., & Petit, P. (1992). In defense of explanatory ecumenism. *Economics and Philos ophy, 8*, 1-21.

Kaplan, D.M., & Craver, C.F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78*(4), 601-627.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phe nomenological Research, 52*, 1-26.

Klein, C. (2008). An ideal solution to disputes about multiply realized kinds. *Philosophical Studies, 140,* 161-177.

Ladyman, J. (2008). Structural realism and the relationship between the special sciences and physics. *Philosophy of Science, 75*, 744-755.

Leuridan, B. (2010). Can mechanisms really replace laws of nature? *Philosophy of Science, 77*(3), 317-340.

Loewer, B. (2009). Why is there anything except physics? *Synthese*, *170*, 217-233.

Mitchell, S.D. (1997). Pragmatic laws. *Philosophy of Science, 64*, 468-479.

Netoff, T.I., Clewley, R., Arno, S., Keck, T., & White, J.A. (2004). Epilepsy in small-world net works. *The Journal of Neuroscience*, *24*(37), 8075-8083.

Polger, T. (2008). Two confusions concerning multiple realization. *Philosophy of Science, 75,* 537-547.

Polger, T. (2009). Evaluating the evidence for multiple realization. *Synthese, 167*, 457-472.

Privman, V. (1998). Universality. In *Encyclopedia of Applied Physics*, Vol. 23, p. 31-46. New York: American Institute of Physics.

Rabinovich, M.I., Varona, P., Selverston, A.I., & Abarbanel, H.D.I. (2006). Dynamical princi ples in neuroscience. *Reviews of Modern Physics, 78*(4), 1213-1265.

Rieke, F., Warland, D., van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the Neural Code.* Cambridge, MA: MIT Press.

Schaffner, K.F. (2008). Theories, models, and equations in biology: The heuristic search for emergent simplifications in neurobiology. *Philosophy of Science, 75*(5), 1008-1021.

Shapiro, L. (2000). Multiple realizations. *Journal of Philosophy, 97*, 635-654.

Shapiro, L. (2008). How to test for multiple realization. *Philosophy of Science, 75,* 514-525.

Sober, E. (1999). The multiple realizability argument against reductionism. *Philosophy of Science, 66*, 542-564.

Stanley, H.E. (1999). Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Reviews of Modern Physics, 71*(2), 358-366.

Stewart, I. (1995). *Nature's Numbers.* New York: Basicbooks.

Strogatz, S.H. (1994). *Nonlinear Dynamics and Chaos.* Perseus Books Publishing.

Ursino, M., & La Cara, G.E. (2006). Travelling waves and EEG patterns during epileptic seizure: Analysis with an integrate-and-fire neural network. *Journal of Theoretical Biology, 242*, 171-187.

Weber, M. (2008). Causes without mechanisms: Experimental regularities, physical laws, and neuroscientific explanation. *Philosophy of Science*, *75*, 995-1007.

Weiskopf, D. (2011a). Models and mechanisms in psychological explanation. *Synthese, 183*(3), 313-338.

Weiskopf, D. (2011b). The functional unity of special science kinds. *British Journal for the Philosophy of Science, 62*(2), 233-258.

Wilson, M. (2010). Mixed-level explanation. *Philosophy of Science, 77*, 933-946.

Wimsatt, W.C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality.* Cambridge: Harvard University Press.

Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, Part I: A counterfactual account. *NOÛS, 37*(1), 1-24.

Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, Part II: Plumbing explan atory depth. *NOÛS, 37*(2), 181-199.

Woodward, J. (2003). *Making Things Happen*. New York: Oxford University Press.

Zednick, C. (2011). The nature of dynamical explanation. *Philosophy of Science, 78*(2), 238-263.