Are Systems Neuroscience Explanations Mechanistic?

Carlos Zednik

czednik@uos.de
Institute of Cognitive Science, University of Osnabrück
49069 Osnabrück, Germany

Paper to be presented at: *Philosophy of Science Association 24th Biennial Meeting (Chicago, IL)*, November 2014

Abstract

Whereas most branches of neuroscience are thought to provide mechanistic explanations, systems neuroscience is not. Two reasons are traditionally cited in support of this conclusion. First, systems neuroscientists rarely, if ever, rely on the dual strategies of decomposition and localization. Second, they typically emphasize organizational properties over the properties of individual components. In this paper, I argue that neither reason is conclusive: researchers might rely on alternative strategies for mechanism discovery, and focusing on organization is often appropriate and consistent with the norms of mechanistic explanation. Thus, many explanations in systems neuroscience can also be viewed as mechanistic explanations.

1. Introduction

There is a widespread consensus in philosophy of science that neuroscientists provide *mechanistic explanations*. That is, they seek the discovery and description of the mechanisms responsible for the behavioral and neurological phenomena being explained. This consensus is supported by a growing philosophical literature on past and present examples from various branches of neuroscience, including molecular (Craver 2007; Machamer, Darden, and Craver 2000), cognitive (Bechtel 2008; Kaplan and Craver 2011), and computational neuroscience (Kaplan 2011). In contrast, one area that has received relatively little philosophical attention is *systems neuroscience*: the study of networks at various levels of brain organization. Do systems neuroscientists, like their colleagues in other branches of the discipline, seek the discovery and description of mechanisms? Answering this question is important for gaining an improved understanding of this exciting and increasingly influential area of research, but also for determining whether the various branches of neuroscience are unified by a common set of epistemic practices and explanatory norms.

Three research traditions can be distinguished within contemporary systems neuroscience. The first seeks the identification and description of networks at various levels of brain organization. The second seeks to reproduce the brain's behavioral dynamics and information-processing capacities through artificial neural network simulations. The third tradition specializes in the development of concise mathematical descriptions of the holistic behavior of biological as well as artificial brain networks. Several philosophical commentators have recently denied that systems neuroscientists in either one of these research traditions provide mechanistic explanations. To a large extent, this denial is motivated by the observation that systems neuroscientists rarely invoke the heuristic strategies of *decomposition* and *localization* that are traditionally associated with mechanistic explanation (Bechtel and Richardson 1993; Silberstein and Chemero 2012; Varela, Thompson, and Rosch 2001). Moreover, the fact that systems neuroscientists often emphasize brain networks' global organization rather than their detailed composition is often considered to be evidence that these researchers have abandoned the mechanistic approach (Silberstein and Chemero 2012). If these commentators are correct, systems neuroscientists are quite unlike their colleagues in other branches of the discipline: they do not provide mechanistic explanations.

In what follows, I briefly introduce the three main research traditions within systems neuroscience (Section 2), and present the main reasons for thinking that researchers working within these traditions have abandoned mechanistic explanation (Section 3). Subsequently, I argue that these reasons are inconclusive (Section 4). For one, systems neuroscientists who eschew the strategies of decomposition and localization may appeal to alternative strategies for discovering mechanisms. For another, focusing on mechanistic organization rather than composition should not be viewed as a departure from mechanistic explanation. Indeed, such a focus is appropriate when a mechanism's organization is the principal determinant of that mechanism's behavior. Insofar as the discipline of systems neuroscience specializes in describing the organization of large network mechanisms in the brain, and insofar as the concept of organization remains poorly understood in philosophy of science, philosophers have much to gain from an improved understanding of mechanistic explanation in this increasingly important branch of neuroscientific research.

2. Three Research Traditions

Systems neuroscience is motivated by the observation that the brain is a complex system of networks of different kinds and at different levels of organization, as well as the observation that these networks are involved in the production of a wide range of behavioral and neurological phenomena. Within systems neuroscience, three conceptually distinct but mutually beneficial research traditions can be distinguished. Whereas the *network tradition* concerns the identification, description and topological analysis of brain networks, the *simulation tradition* specializes in the design of artificial neural network models to simulate the biological brain's behavioral dynamics and information-processing capacities. Both of these traditions are linked to the *complexity tradition*, the aim of which is to develop concise mathematical descriptions of the behavioral dynamics or information-processing capacities of artificial as well as biological brain networks.

The network tradition in systems neuroscience aims to identify and describe brain networks of different kinds and at different levels of organization. Thus, networks have been identified at the level of individual neurons within a population, at the level of neural populations (e.g. cortical columns) within a cortical region, and at the level of the brain as whole: networks of cortical regions. At each one of these levels, researchers in the network tradition must first determine how to identify a network's elements. Sometimes this choice is principled, as when biological principles are used to individuate nerve cells or cortical regions. At other times the choice is highly pragmatic, as when the network elements are chosen to correspond with voxels in fMRI data or with the placement of electrodes in electrophysiological studies. After having identified a particular set of elements in this way, researchers in the network tradition must choose which kind of connections to pay attention to. Thus, networks might be defined over anatomical links such as synapses, but might also be defined over causal or functional links, typically operationalized as correlated activity over time. Once these choices have been made, brain networks can be identified via a variety of mapping and imaging techniques. These range from invasive methods such as histological studies of nerve cells and synaptic connections, to non-invasive methods such as structural, functional and diffusion MRI, among others (for a detailed overview of these methods, and of the network tradition in systems neuroscience as a whole, see: Sporns 2011).

Once brain networks have been identified, researchers in the network tradition typically represent and study these networks by invoking the mathematical framework of *graph theory*. A graph theoretical representation of a brain network characterizes the network's elements as nodes in a graph, and the anatomical, functional or causal links between these elements as (possibly weighted and/or directional) connections between individual nodes. Such graph theoretical representations facilitate the study of a network's global and local organization or *topology*. Thus, large-scale brain networks of all kinds have been shown to exhibit *small-world* topologies, that is, a high-degree of local clustering with short average path-lengths (Bassett and Bullmore 2006). Similarly, researchers have demonstrated the existence of *hub nodes* of relatively high degree, network *motifs* (small sub-graphs that are repeated throughout a network), and *modules*, i.e. densely interconnected communities of nodes with relatively sparse links to nodes in other communities (Sporns, Honey, and Kötter 2007; Sporns and Kötter 2004;

Meunier, Lambiotte, and Bullmore 2010). To date, these and other graph theoretic concepts have been used to understand the topology of brain networks in *C. elegans* (White et al. 1986; Izquierdo and Beer 2013), as well as cats and macaque monkeys (Sporns and Kötter 2004), and are likely to prove instrumental in the eventual success of the *Human Connectome Project* (Sporns, Tononi, and Kötter 2005) and other human brain mapping initiatives.

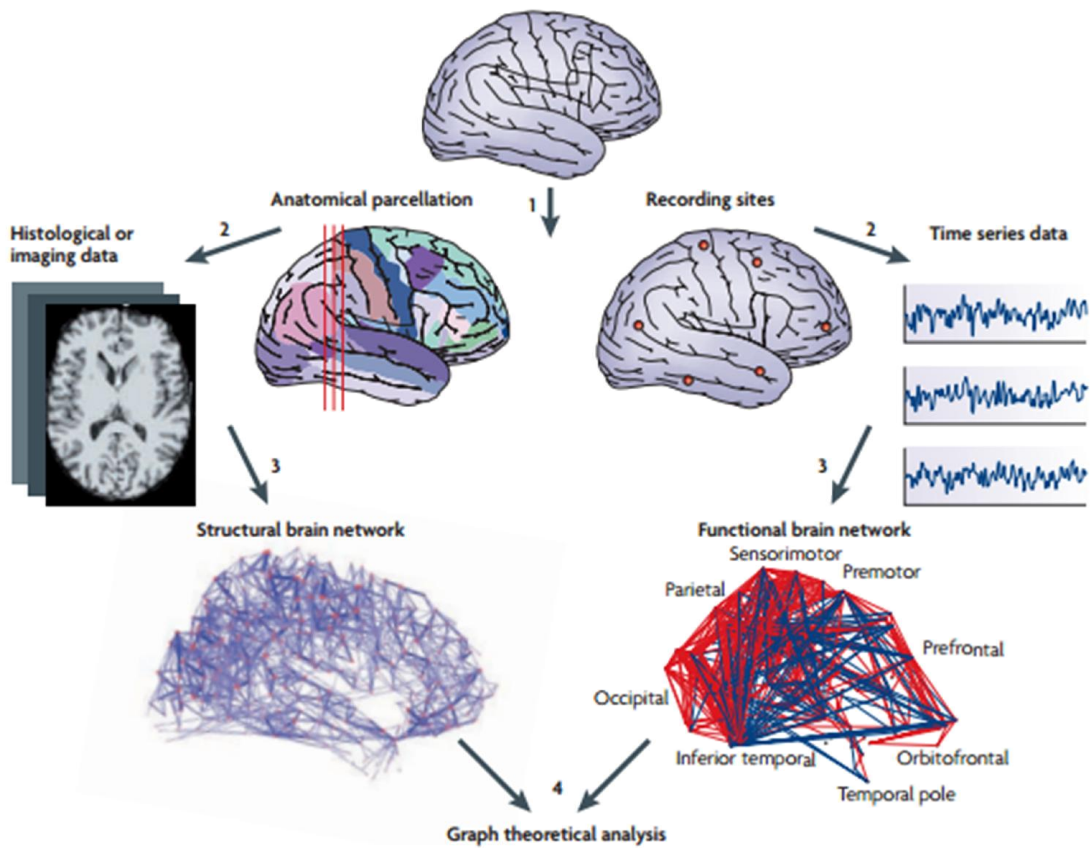

Figure 1. A schematic representation of model-development in the network tradition. The left pathway represents the development of structural network models; the right pathway represents the development of functional network models. Reprinted from Bullmore & Sporns (2009).

Whereas the network tradition seeks the description and topological analysis of brain networks, the simulation tradition aims to develop artificial neural network models to reproduce the brain's behavioral dynamics or information-processing capacities. Thus, the primary purpose of these models is to understand the parameters under which brain networks might produce oscillations, synchronized firing patterns, and robustness to perturbation, as well as to understand when such networks are particularly good or bad at processing and integrating information. Although the relevant parameters are often local—determining e.g. the way in which individual network elements transform inputs to outputs—perhaps the most interesting research in this tradition seeks to understand the influence of topological parameters. Thus for example, Perez et al. (2011) have explored the extent to which random, small-world and scale-free networks exhibit local and global patterns of synchronization. Similarly, Tononi & Sporns (2003) have studied the relationship between the degree of modularity in a network and the degree of information integration—the ease by which information can be transmitted between any two network elements. In general, whereas the network tradition in systems neuroscience seeks to describe the kinds of networks that actually exist in the brain, the simulation tradition seeks to understand what these kinds of networks can do.

The third main research tradition within the discipline of systems neuroscience can be termed the *complexity tradition*. Building on decades of research in the discipline of complexity science, this tradition within systems neuroscience aims to develop concise mathematical models of behavior and information-processing in biological and artificial networks. These models are typically developed using the mathematical concepts and methods of *dynamical systems theory* and *information theory*. Whereas the former can be used to concisely characterize a network's behavior over time, the latter shows how information is processed and integrated. One of the key features of these models is their relative simplicity. Whereas a particular network may consist of myriad reciprocally and non-linearly connected elements, its behavioral dynamics is often quite simple, exhibiting periodic oscillation or stability over time, and its information-processing efficient. The complexity tradition in systems neuroscience seeks to mathematically characterize these simple behaviors, which are often also characterized as the "emergent" properties of brain networks.

Although most research projects in systems neuroscience can be associated with either one of these three research traditions, it is not uncommon to see researchers combine the methods and results of two or more of them. Thus, researchers working in the simulation tradition increasingly seek to replace highly unrealistic artificial neural network models with "biologically inspired" network models that are rooted in the findings of the network tradition. Similarly, although many of the mathematical models developed in the complexity tradition were originally developed to characterize the global behavior of artificial neural networks, some of these models are proving themselves to be exceedingly useful for understanding the behavioral dynamics or information-processing of the biological brain. Notably, these combined research efforts are likely to be particularly influential in the future, since they apply the analytic power of computer simulations and mathematical analyses to increasingly accurate and biologically plausible models of the brain.

3. Abandoning Mechanistic Explanation

To what extent do researchers working within these three research traditions provide mechanistic explanations of behavioral and neurological phenomena? Mechanistic explanations center on descriptions or models of the mechanisms responsible for the phenomena being explained. Although there have been many different formulations of what constitutes a mechanism, the basic idea is that of an organized system of parts (or "entities") and operations (or "activities") whose changing properties over time exhibit a phenomenon of explanatory interest (Bechtel and Richardson 1993; Machamer, Darden, and Craver 2000; Bechtel and Abrahamsen 2010; Craver 2007). Notably, although most early contributions to the literature emphasize the characteristically diagrammatic nature of mechanistic explanations, many recent treatments acknowledge the possibility and prevalence of mathematical mechanism-descriptions. Thus, Kaplan & Craver (2011) have recently proposed a *model-mechanism-mapping constraint* (3M) intended to capture the requirements all mechanistic models must satisfy for them to be used in mechanistic explanations:

> "3M: In successful explanatory models…(*a*) the variables in the model correspond to components, 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." (Kaplan and Craver 2011, 611)

Before a mechanism can be described, it must be discovered: Researchers in neuroscience do not typically know in advance what the component parts and operations of a brain mechanism might be. According to Bechtel & Richardson's (1993) celebrated account of mechanism discovery, neuroscientists typically rely on the heuristic strategies of *decomposition* and *localization* to identify a mechanism's component parts and operations, and thus, to develop a description or model of that mechanism. That is, they analyze the target phenomenon into a series or complex of relatively simple operations, break apart the physical system from which that phenomenon arises into a collection of smaller parts, and then study in detail the behavior of individual parts in order to link those parts to particular operations. Because of their perceived importance to mechanistic explanation in several disciplines, decomposition and localization are now often considered to be "the *sine qua non* of mechanistic explanation" (Silberstein and Chemero 2012, 3).

It is generally agreed that brain networks are mechanisms in the above sense. For one, they can clearly be viewed as organized systems of parts and operations. For another, insofar as many brain networks can be associated with particular behavioral or neurological processes, they can be said to exhibit various phenomena of explanatory interest. But although brain networks are mechanisms, some commentators have questioned whether the researchers who study these networks seek mechanistic explanations, as opposed to other kinds of explanations. Their questioning usually centers on the models being developed in systems neuroscience: what kinds of models these are, what kinds of properties they describe, and how these models are developed.

Unlike their colleagues in other branches of neuroscience, systems neuroscientists rarely invoke the heuristic strategies of decomposition and localization. This can be observed in all three of the research traditions introduced above. In the complexity tradition, researchers do not attempt to describe a brain network's component parts and operations at all, but just seek to describe its overall behavior. Thus, the mathematical models being developed in this tradition are often *phenomenological models* rather than mechanistic models: they accurately describe the phenomenon being explained, without representing the causal structures and processes responsible therefore (Mauk 2000; Weiskopf 2011).

In the simulation tradition, the brain's behavioral dynamics and information-processing capacities are reproduced by way of artificial neural network models. Although these models can be construed as mechanistic models because they describe a network's component elements and connections, they are not usually developed via the heuristic strategies of decomposition and localization. In this tradition, the phenomenon being modeled is rarely analyzed into a series of localizable operations, and it is rare to see systems neuroscientists characterize the contributions of individual parts to the network mechanism's overall behavior. Thus, in Bechtel & Richardson's assessment, the simulation tradition

> "emphasizes systems whose dynamic behavior corresponds to the activity we want to explain, but in which the components of the system do not perform recognizable subtasks of the overall task…The overall architecture of the system—and especially the way components are connected—is what explains cognitive capacities, and not the specific tasks performed by the components. We have abandoned decomposition and localization." (Bechtel and Richardson 1993, 222–223)

Thus, rather than describe the contributions of a network's individual elements and connections by deploying the strategies of decomposition and localization, researchers in the simulation tradition focus on the contribution of the network's organization to its overall behavior.

Researchers working in the network tradition within systems neuroscience similarly focus on overall organization rather than individual components. Although developing a network model of a biological brain network involves physically decomposing the brain or brain area being studied to determine the individual elements of the model, researchers rarely seek to describe in detail the structural features of these elements. Indeed, more often than not such network elements correspond to spatially-defined voxels or segments, rather than biologically-defined cortical structures that might plausibly be viewed as a mechanism's "working parts" (Craver 2007). In addition, it can be hard to interpret the connections between the elements of such a network model as a mechanism's component operations. Although some studies have begun to identify causal interactions between brain structures, far more is known about the brain's structural and statistical connectivity. Thus, although it may be known that a brain network's elements are structurally linked and/or statistically interrelated, it remains generally unclear exactly how these individual elements interact, and thus, unclear how they each contribute to the behavioral or neurological phenomena of explanatory interest.

In each one of the three research traditions, the focus on overall behavior and organization rather than detailed composition is often by necessity, rather than by choice. It has long been known (and current research in the network tradition has confirmed) that brain networks at all levels of organization often feature massively reciprocal and non-linear interactions (Varela, Thompson, and Rosch 2001; Sporns 2011). In such systems, the behavior of any individual component is at all times influencing, but also being influenced by, the behavior of the rest of the system. Thus, although it may be possible to physically individuate neurons, neural populations, and cortical regions, it can be difficult (or indeed, computationally intractable) to describe their individual contribution to the behavior of the system as a whole. What researchers in the simulation and complexity traditions of systems neuroscience have shown, in contrast, is that a network's behavior can often be more usefully predicted and understood by focusing on holistic organizational properties instead. Thus, in Silberstein & Chemero's words,

> "rather than viewing the neurons, cell groups or brain regions as the basic unit of explanation, it is brain multiscale networks and their large-scale, distributed and non-local connections or interactions that are the basic unit of explanation." (Silberstein and Chemero 2012, 5)

To summarize: Unlike their colleagues in other branches of neuroscience, systems neuroscientists do not invoke the heuristic strategies of decomposition and localization to discover and describe mechanisms, and indeed, are frequently prevented from doing so due to the characteristic complexity of brain networks. Because their focus is placed squarely on network organization rather than detailed composition, there are good reasons to be skeptical about systems neuroscientists' commitment to mechanistic explanation.

4. Mechanistic Explanation in Systems Neuroscience

The previous section described reasons for believing that systems neuroscientists have abandoned mechanistic explanation. This section argues that these reasons rely on an unnecessarily narrow conception of mechanistic explanation. Mechanistic explanation need not invoke the heuristic strategies of decomposition and localization, and need not be focused on component parts and operations. Indeed, when a particular mechanism's behavior is largely determined by its overall organization, emphasizing this organization is appropriate and consistent with the norms of mechanistic explanation.

Consider first the suggestion that systems neuroscientists have abandoned mechanistic explanation just because they eschew the heuristic strategies of decomposition and localization. Taken at face value, this suggestion concerns the epistemic practices researchers invoke during the process of model development. Thus construed, however, it is wrong to assume that decomposition and localization are "the *sine qua non* of mechanistic explanation" as Silberstein & Chemero suggest. According to the conception of scientific discovery embraced by Bechtel & Richardson (1993), decomposition and localization greatly facilitate mechanism discovery by allowing researchers to quickly (albeit fallibly) traverse a conceptual space of "how-possibly" mechanistic models, with the goal of eventually identifying a "how-actually" model of the mechanism. However, decomposition and localization are not the only heuristic strategies that can be used for this purpose: there are a wide range of strategies to choose from.

Zednik (in press) has recently explored some alternative strategies for mechanism discovery. To cite just one example, consider the way researchers in evolutionary robotics invoke evolutionary algorithms to develop simulated mechanisms for *minimally cognitive* tasks (Harvey et al. 2005; Beer 2003). Because the simulated mechanisms that emerge from such evolutionary algorithms are relatively unconstrained by the ingenuity and design preferences of human researchers, many of them are characterized by features often seen as obstacles to mechanistic explanation: reciprocal non-linear interactions, a close integration of brain, body, and environment, and high sensitivity to temporal detail. Although it remains to be seen to what extent the discovery of such mechanisms in simulation can be used to make concrete inferences about analogous real-world mechanisms (for discussion see Webb (2009) and responses to this target article), the evolution and analysis of these simulated mechanisms can surely be viewed as a heuristic strategy that helps to identify areas of the space of possible mechanistic models that merit further exploration. Thus, the approach adopted by evolutionary roboticists can be understood as a proof of concept that there exist heuristic strategies for mechanism discovery beyond decomposition and localization.

Considered as heuristic strategies, therefore, decomposition and localization are not in fact essential for mechanistic explanation. As a consequence, the mere fact that systems neuroscientists rarely invoke these heuristic strategies does not by itself show that they have abandoned mechanistic explanation. That said, an alternative way of understanding the claim that decomposition and localization are “the *sine qua non* of mechanistic explanation” concerns not the heuristic strategies themselves, but the result of applying these strategies: descriptions of the component parts and operations of mechanisms. Are such componential descriptions, as opposed to descriptions that mainly or exclusively describe a mechanism’s organization, essential for mechanistic explanation?

Recall that mechanisms are organized systems of parts and operations that exhibit a particular phenomenon of explanatory interest, and that mechanistic explanations center on descriptions or models of such mechanisms. Moreover, recall that according to Kaplan & Craver’s 3M constraint, models that figure in mechanistic explanations contain variables that “correspond to components, activities, properties, and organizational features of the target mechanism”, the dependencies between which “correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism” (Kaplan and Craver 2011, 611). Notably, the second part of the 3M constraint is indicative of one of the most important norms of mechanistic explanation, viz. that such explanations should demonstrate how a target phenomenon "is situated in the causal structure of the world" (Craver 2013, 135). Exactly what such a demonstration would amount to is of course controversial, but one common measure of success is that a mechanistic model should render the represented mechanism amenable to interventions of manipulation and control (Woodward 2003; Craver 2007). That is, the model should represent (just) those properties of the mechanism that, when lesioned, activated, or otherwise influenced, effect predictable changes in the mechanism's behavior. Notably, although this norm is usually understood in terms of interventions on individual component parts or operations, it is also satisfied by models that facilitate interventions on the mechanism's overall organization.

Recall that many of the artificial and biological networks studied by systems neuroscientists consist of a large number of elements with reciprocal non-linear interactions. In these networks, organizational properties are frequently the primary determinants of overall behavior (Sporns 2011; Varela, Thompson, and Rosch 2001). Indeed, these properties can often be thought of as *order parameters*, the modification of which predicts changes in the network's overall behavioral dynamics and information processing capacities. Thus for example, Tononi & Sporns (2003) measure the effect varying degrees of modularity have on the degree of information integration within a network, presenting their results as a continuous function in which information integration is highest for networks with intermediate degrees of modularity, and lower for completely modular and completely homogeneous networks. Such studies reveal that network models can emphasize organizational properties instead of properties of individual component parts and operations, while still rendering the relevant network mechanism amenable to interventions of manipulation and control. Insofar as this is one of the principal norms of mechanistic explanation, models that describe brain networks in ways that satisfy this norm should be considered mechanistic. Therefore, systems neuroscience explanations that center on such models can after all be viewed as mechanistic explanations.

How far can this line of reasoning be pushed? Although the network models being developed in the network and simulation traditions emphasize topological properties, they still provide some (admittedly not very detailed) insight into the individual component parts and operations of network mechanisms. Therefore, these models satisfy both parts of the 3M constraint, while additionally facilitating interventions of manipulation and control. But now compare these models to the models developed in the complexity tradition, the aim of which is to provide concise mathematical descriptions of brain networks' behavioral dynamics and information-processing capacities. As was discussed previously, such descriptions often amount to phenomenological models, i.e. models that describe the phenomena being explained without representing the causal structures and processes—and a fortiori, the mechanisms—responsible therefore. But not all models in the complexity tradition are pure phenomenological models. Indeed, some of the most interesting models in this tradition link variables that describe a network's global behavior to parameters that represent its topology. Although these models do not represent the relevant network mechanism's individual parts and operations, they do represent its organization in a way that renders it amenable for interventions of manipulation and control. Are these models mechanistic as well?

Whereas the network models developed in the network and simulation traditions are quite clearly mechanistic, and the systems neuroscientists in the network and simulation traditions can therefore be thought to provide mechanistic explanations, it is not so clear what to make of non-phenomenological models in the complexity tradition. On one hand, these models fail to describe what is often viewed as essential for mechanistic explanation: the relevant mechanism's component parts and operations. On the other hand, they do at least describe the mechanism's organization, and researchers often exploit this fact for the purposes of manipulation and control. Therefore, whether or not these models are mechanistic models depends on how these competing considerations are weighed against one another. Perhaps the following additional consideration can be used to break the tie in favor of the latter: Insofar as mechanistic models should *just* represent those properties that significantly contribute to a mechanism's behavior, and insofar as in brain networks these properties are often predominantly organizational, it seems that even those models in the complexity tradition that exclude compositional details entirely should be construed mechanistically.

5. Conclusion: Toward an Account of Mechanistic Organization

The preceding sections aimed to show that systems neuroscience—spanning the network, simulation, and complexity traditions—can often be thought to provide mechanistic explanations of behavioral and neurological phenomena. Although systems neuroscientists only rarely invoke the heuristic strategies of decomposition and localization, these strategies are not essential for mechanistic explanation. Moreover, and most importantly, although systems neuroscientists typically emphasize organizational properties over and above the properties of individual component parts and operations, these properties can often be used for the purposes of manipulation and control. Thus, although they might pay far less attention to individual component parts and operations, systems neuroscientists seem to embrace many of the same explanatory methods and norms as their colleagues in other branches of neuroscience.

In closing, it is worth reflecting on the further philosophical significance of this mechanistic construal of systems neuroscience. Because systems neuroscientists specialize in characterizing the organization of a particular family of mechanisms—brain networks—their work is likely to be particularly beneficial to philosophers of science seeking an improved understanding of mechanistic organization. Although the concept of organization has long been included in philosophical definitions of "mechanism", it remains poorly understood, and indeed, its importance underestimated. Thus for example, one particularly influential treatment of mechanistic explanation presupposes that mechanisms are organized linearly, from "set up to termination conditions" (Machamer, Darden, and Craver 2000). In contrast, Bechtel & Abrahamsen (2010) have demonstrated that many mechanisms in neuroscience and biology are organized cyclically, and Levy & Bechtel (2013) have sought to explore the role of *motif*-like organizational building blocks in mechanisms in the life sciences. Insofar as systems neuroscientists have revealed many canonical forms of mechanistic organization in the brain—as well as many ways of studying this organization—it stands to reason that philosophers of science have much to gain from paying increased attention to this exciting and increasingly influential area of research.

References

Bassett, Danielle Smith, and Edward T. Bullmore. 2006. “Small-World Brain Networks.” *The Neuroscientist* 12 (6) (December): 512–523.

Bechtel, William. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.

Bechtel, William, and Adele Abrahamsen. 2010. “Dynamic Mechanistic Explanation: Computational Modeling of Circadian Rhythms as an Exemplar for Cognitive Science.” *Studies in History and Philosophy of Science Part A* 41 (3) (September): 321–333.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

Beer, Randall D. 2003. “The Dynamics of Active Categorical Perception in an Evolved Model Agent.” *Adaptive Behavior* 11 (4) (December 1): 209–243; discussion 244–305.

Bullmore, Edward T., and Olaf Sporns. 2009. “Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems.” *Neuroscience* 10 (March): 186–198.

Craver, Carl F. 2007. *Explaining the Brain*. Oxford: Oxford University Press.

———. 2013. “Functions and Mechanisms: A Perspectivalist View.” In *Functions: Selection and Mechanisms*, edited by Phillipe Huneman, 133–158. Dordrecht: Springer.

Harvey, Inman, Ezequiel A. di Paolo, Elio Tuci, Rachel Wood, and Matt Quinn. 2005. “Evolutionary Robotics: A New Scientific Tool for Studying Cognition.” *Artificial Life* 11: 79–98.

Izquierdo, Eduardo J., and Randall D. Beer. 2013. “Connecting a Connectome to Behavior: An Ensemble of Neuroanatomical Models of C. Elegans Klinotaxis.” *PLoS Computational Biology* 9 (2).

Kaplan, David M. 2011. “Explanation and Description in Computational Neuroscience.” *Synthese* 183: 339–373.

Kaplan, David M., and Carl F. Craver. 2011. “The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective.” *Philosophy of Science* 78 (October): 601–627.

Levy, Arnon, and William Bechtel. 2013. “Abstraction and the Organization of Mechanisms.” *Philosophy of Science* 80 (2) (April): 241–261.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. “Thinking about Mechanisms.” *Philosophy of Science* 67 (1) (April): 1–25.

Mauk, Michael D. 2000. “The Potential Effectiveness of Simulations versus Phenomenological Models.” *Nature Neuroscience* 3 (7) (July): 649–651.

Meunier, David, Renaud Lambiotte, and Edward T. Bullmore. 2010. “Modular and Hierarchically Modular Organization of Brain Networks.” *Frontiers in Neuroscience* 4 (December).

Perez, Toni, Guadalupe C. Garcia, Victor M. Eguiluz, Raul Vicente, Gordon Pipa, and Claudio Mirasso. 2011. “Effect of the Topology and Delayed Interactions in Neuronal Networks Synchronization.” *PLoS One* 6 (5).

Silberstein, Michael, and Anthony Chemero. 2012. “Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences.” In *Philosophy of Science Association 23rd Biennial Meeting*. San Diego, CA.

Sporns, Olaf. 2011. *Networks of the Brain*. Cambridge, MA: MIT Press.

Sporns, Olaf, Christopher J. Honey, and Rolf Kötter. 2007. “Identification and Classification of Hubs in Brain Networks.” *PLoS One* 2 (10).

Sporns, Olaf, and Rolf Kötter. 2004. “Motifs in Brain Networks.” *PLoS Biology* 2 (11) (November): e369.

Sporns, Olaf, Giulio Tononi, and Rolf Kötter. 2005. “The Human Connectome: A Structural Description of the Human Brain.” *PLoS Computational Biology* 1 (4) (September): e42.

Tononi, Giulio, and Olaf Sporns. 2003. “Measuring Information Integration.” *BMC Neuroscience* 4 (31) (December 2).

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 2001. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Webb, Barbara. 2009. “Animals Versus Animats: Or Why Not Model the Real Iguana?” *Adaptive Behavior* 17 (4) (July 28): 269–286.

Weiskopf, Daniel A. 2011. “Models and Mechanisms of Psychological Explanation.” *Synthese*.

White, J. G., E. Southgate, J. N. Thomson, and S. Brenner. 1986. “The Structure of the Nervous System of the Nematode Caenorhabditis Elegans.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 314: 1–340.

Woodward, Jim. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Zednik, Carlos. in Press. “Heuristics, Descriptions, and the Scope of Mechanistic Explanation.” In *How Does Biology Explain? An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, edited by Christophe Malaterre and Piere-Alain Braillard. Dordrecht: Springer.