Models and Mechanisms in Network Neuroscience

Carlos Zednik
carlos.zednik@ovgu.de
Otto-von-Guericke Universität Magdeburg

Accepted Manuscript (12 April 2018) of an article to be published by Taylor & Francis in *Philosophical Psychology*

0. Abstract

This paper considers the way mathematical and computational models are used in network neuroscience to deliver mechanistic explanations. Two case studies are considered: Recent work on klinotaxis by *Caenorhabditis elegans*, and a longstanding research effort on the network basis of schizophrenia in humans. These case studies illustrate the various ways in which network, simulation and dynamical models contribute to the aim of representing and understanding network mechanisms in the brain, and thus, of delivering mechanistic explanations. After outlining this mechanistic construal of network neuroscience, two concerns are addressed. In response to the concern that functional network models are non-explanatory, it is argued that functional network models are in fact explanatory mechanism sketches. In response to the concern that models which emphasize a network's organization over its composition do not explain mechanistically, it is argued that this emphasis is both appropriate and consistent with the principles of mechanistic explanation. What emerges is an improved understanding of the ways in which mathematical and computational models are deployed in network neuroscience, as well as an improved conception of mechanistic explanation in general.

1. Introduction

Network neuroscientists aim to characterize the composition, organization, and activity of networks at several levels of brain organization. To this end, they deploy a variety of mathematical and computational models. As these models become increasingly widespread and sophisticated, questions arise about their explanatory import: To what extent do they do more than merely describe networks in the brain, and explain how these networks give rise to specific behavioral and cognitive phenomena? Insofar as the models deployed in network neuroscience are genuinely explanatory, how exactly do they explain?

Several philosophical commentators have argued that network neuroscience delivers *mechanistic explanations* (see e.g. Bechtel 2015; Craver 2016; Levy and Bechtel 2013; Miłkowski 2016; Zednik 2014, 2015). In so doing, they focus mostly on the contribution of *network models*, graphical representations of networks in biological brains. However, this picture seems incomplete. In addition to network models, network neuroscientists also regularly deploy *simulation models* to animate and intervene on a (biological or artificial) network's topology and behavior, and *dynamical models* to concisely describe its behavior over time. Because of their narrow focus on network models, these philosophical commentators have yet to appreciate the way other kinds of models contribute to the aim of representing and understanding network mechanisms in the brain.

At the same time, other philosophical commentators have denied that some or all of the models deployed in network neuroscience explain mechanistically. Some of these commentators argue that, because many of these models emphasize a network's organizational (or *topological*) properties rather than the properties of individual components, the principles of mechanistic explanation have been abandoned (Rathkopf 2015; Silberstein and Chemero 2013; Woodward 2013).[1] Even Carl Craver (2016), who himself articulates a mechanistic construal of network

1 Negative arguments of this kind are sometimes accompanied by positive arguments in favour of distinctly non-mechanistic construals of network neuroscience. In particular, Silberstein & Chemero (2013) argue that dynamical models in network neuroscience yield covering law explanations rather than mechanistic explanations. Other commentators argue that network models deliver "topological" or other kinds of mathematical explanations (Huneman 2010; Kostić 2016; Rathkopf 2015; Woodward 2013). Although some commentators question the legitimacy of these non-mechanistic explanations (Craver 2016; Kaplan and Craver 2011), the present contribution takes no stance on whether such putative explanations are genuinely explanatory, nor on whether models in

neuroscience, argues that so-called *functional* network models—models that depict statistical correlations between pragmatically-individuated "chunks" of brain tissue—do not deliver explanations at all.

Evidently, there is considerable disagreement about whether and how mathematical and computational models in network neuroscience explain. The aim of the present discussion is to argue that this disagreement can be resolved by recognizing that network, simulation, and dynamical models are only rarely deployed in isolation, but are instead typically used in combination to satisfy the norms of mechanistic explanation. Whereas network models are often used to characterize a network mechanism's composition and organization, simulation models are used to describe the contribution of certain components or organizational properties to the mechanism's overall behavior. Dynamical models, in turn, are often used to describe that behavior in detail, but also to characterize the time-sensitive behavior of individual components. Thus, network, simulation, and dynamical models are used in a complementary fashion, allowing network neuroscientists to represent and understand the composition, organization, and behavior of network mechanisms in the brain.

The discussion is organized as follows. Section 2 will introduce and distinguish between network, simulation, and dynamical models in network neuroscience. In Section 3, two case studies are introduced: one on *C. elegans klinotaxis* and another on *schizophrenia* in humans. These case studies illustrate the combined deployment of network, simulation, and dynamical models. Section 4 will then present a mechanistic construal of network neuroscience; the combined use network, simulation, and dynamical models allows investigators to represent and understand the composition, organization, and behavior of brain network mechanisms. Subsequently, two of the most significant criticisms of this mechanistic construal will be considered. In Section 5, it will be argued that functional network models can in fact be viewed as genuinely explanatory *mechanism sketches*: first approximations of a mechanism's component parts, operations, and/or organizational properties. In Section 6, it will be argued that a model's emphasis of topological organization is by no means antithetical to the principles of mechanistic explanation. Indeed, an emphasis on topology may often be a particularly appropriate way of delivering mechanistic explanations, and philosophers concerned with mechanistic explanation would be well-advised to take a closer look at recent examples from network neuroscience. In summary, what emerges from this discussion is not network neuroscience also contribute thereto. Instead, the focus is on understanding whether and how network, simulation, and dynamical models *also* contribute to mechanistic explanations.

only a mechanistic construal of network neuroscience, but also an improved understanding of the way in which sophisticated mathematical and computational methods can be used to deliver mechanistic explanations quite generally.

Before beginning in earnest, it is worth considering why this debate matters in the first place. Although it may not matter how the modeling practices in network neuroscience are labeled, it is important that their individual contributions be understood, and that their interdependencies not be overlooked. Calling some models "mechanistic" while labeling others "non-mechanistic" or even "non-explanatory" threatens to obscure the fact that network, simulation, and dynamical models are often combined for explanatory gains. The framework of mechanistic explanation has already shown how different modeling practices and experimental techniques are combined in other branches of neuroscience; the present discussion suggests that this framework can do so for network neuroscience as well.

Moreover, insofar as the principles of mechanistic explanation still afford clarification and elaboration, a closer look at new and sophisticated mathematical and computational methods may reveal hitherto unknown ways of discovering, representing and understanding mechanisms in neuroscience and beyond (see also Zednik 2015). In particular, philosophers still have a relatively poor understanding of the way practicing scientists go about discovering, describing, and understanding mechanisms' organization (but cf. Levy and Bechtel 2013). Insofar as many of the models deployed in network neuroscience are purpose-built for characterizing the organization of network mechanisms in the brain, taking a closer look at these models may allow philosophers to better understand the role of organization in mechanistic explanation more generally. To this end, Section 6 proposes a criterion to determine when a topological property of a network is in fact an organizational property of a mechanism. Although more philosophical work is required to develop a comprehensive account of mechanistic organization, the reason to consider whether and how models in network neuroscience explain is not only to subsume one more discipline under the mechanistic banner, but to also better understand the nature of mechanistic explanation itself.

2. Networks and models

2.1. Kinds of networks

Networks can be identified at several levels of brain organization: at the level of neurons within a population; at the level of neural assemblies (e.g. columns or microcircuits) within a region; and at the level of regions within the brain as whole (for a comprehensive review see Sporns 2011). At any single level, different kinds of networks can be distinguished, depending on the way in which network elements are individuated, and on the way in which connections between these elements are defined.

Network elements are individuated by deploying a *parcellation scheme*, a method for analytically breaking apart a particular volume of brain tissue. Some parcellation schemes are grounded in established theoretical principles. For example, the principles of cell biology are often applied to identify network elements with individual nerve cells. Similarly, the cytoarchitectural principles invoked by Brodmann (1909) are still used today to identify network elements with brain regions. In contrast, many other parcellation schemes are distinctly pragmatic. For example, network elements are sometimes identified with the voxels of brain tissue individuated by a particular fMRI scanner, or with the recording sites of EEG electrodes. Although questions arise about the explanatory significance of these pragmatic parcellation schemes (Craver 2016; Wig, Schlaggar, and Petersen 2011; but see Section 5 below), these schemes continue to play a prominent role in contemporary research (see e.g. Bullmore and Sporns 2009).

Connections between elements are defined by applying a *connectivity scheme*: a rule for determining whether and how any two elements are linked. Elements are connected *structurally* if they are linked anatomically, e.g. by synapses, gap junctions, or fiber tracts. In contrast, network elements are said to be connected *functionally* if their activity is correlated statistically over time.[2] Most intriguingly perhaps, network elements are connected *effectively* if they are presumed to interact causally. Notably, connectivity schemes may cross-cut one another. It remains unclear to which extent structural connectivity determines (or is determined by) functional connectivity, and although effective connectivity is typically grounded on measures of functional connectivity (see Section 5 for detail), the former is underdetermined by the latter (for a review see Friston 1994, 2011).

As a matter of convention, kinds of networks are distinguished by kinds of connectivity:

2 Several commentators (e.g. Craver 2016) have already noted that this use of the term ‘functional’ is misleading, because statistical correlations need not correspond to causally relevant interactions. See Section 5 for further discussion.

structural networks are defined over structural connectivity schemes; *functional networks* are defined over functional connectivity schemes; and *effective networks* are defined over effective connectivity schemes. Networks of each kind may involve elements of any kind, and can be identified at any level of brain organization.

2.2. Kinds of models

Network neuroscientists deploy at least three different kinds of models to describe and understand the composition, organization, and activity of structural, functional, and effective networks in the brain. These models can be distinguished by the involvement of distinct mathematical and/or computational methods, but also by their unique purposes.

Perhaps the most recognizable class of models is the class of *network models.* These consist of *graphs* that depict a particular network in the biological brain: Nodes correspond to network elements, and edges correspond to network connections. Once a graph has been constructed, its local and global organization—its *topology*—can be analyzed using the tools and concepts of *graph theory.* These can be used to identify *hub nodes* (nodes with a relatively large number of edges) and *motifs* (local patterns of connectivity that are repeated throughout the network). They can also be used to determine the degree of *clustering* or *modularity* in a network (the degree to which nodes are arranged into densely interconnected communities of nodes that are sparsely connected to other communities), and for measuring the overall density or degree of randomness of a network's connections (Bullmore and Sporns 2009; Fornito, Zalesky, and Breakspear 2013).

Network neuroscientists strive for more than the representation and analysis of networks in the biological brain, however. For this reason, they also commonly deploy *simulation models,* in which more-or-less realistic networks are simulated on a computer, and their behavior is studied in detail.[3] Although these models may be "neurally inspired" insofar as they incorporate principles of brain function and organization (e.g. neural spiking, spreading activation, and plasticity), their primary purpose is not to represent networks in the biological brain, but rather to explore what different kinds of networks are in principle able to do. In particular, simulation models are often

3 Use of the term 'simulation model' in this restricted sense should not be taken to imply that simulations play no role in other modeling contexts. Indeed, biological networks are also often simulated, and simulation is frequently used to animate dynamical models as well. Thus, in the present context, the term 'simulation' is only meant to highlight the fact that simulation models target artificial networks that may not directly correspond to any network in the biological brain.

used to explore how changes in the composition or organization of a network—considered as an abstract mathematical entity rather than as a concrete biological structure—influences its behavioral dynamics or information-processing capacities (see e.g. Kitano and Fukai 2007; Pérez et al. 2011; Watts and Strogatz 1998).

In addition to network and simulation models, network neuroscientists also often deploy *dynamical models.* Dynamical models are used in many scientific domains to highlight patterns of change over time, as well as these patterns' dependence on certain parameters. In network neuroscience, they are traditionally used to concisely describe a (biological or artificial) network's overall activity over time, as well as to describe the activity of individual units or regions. Thus, some dynamical models specify a small number of variables that provide a low-dimensional projection of a network's total state, allowing researchers to understand how this total state depends on global parameters such as average connection density. In contrast, other dynamical models specify variables that correspond to the states of individual neural units or brain regions, and capture the way these states depend on the concurrent activity of other units (Izhikevich 2007; Izquierdo and Beer 2013). Notably, dynamical models allow investigators to deploy the tools and concepts of *dynamical systems theory* to characterize a particular network's or unit's transient or asymptotic (i.e. long-term) behavior (Zednik 2011).

3. Two case studies

Network neuroscience is characteristically heterogeneous: Network, simulation, and dynamical models are combined within the scope of individual research efforts. Two case studies illustrate this heterogeneity: A recent investigation of *C. elegans klinotaxis*, and a longstanding research effort to uncover the network basis of *schizophrenia*.

3.1. C. elegans klinotaxis

One particularly influential research effort aims to reveal the *connectome* of *Caenorhabditis elegans* (White et al. 1986). The connectome is the structural network that includes every one of the nematode worm's nerve cells, synapses, and electrical gap junctions. Network models of the C. elegans connectome consist of graphs whose nodes correspond to individual neurons, and whose edges correspond to individual synapses or gap junctions. On the basis of these models, graph

theoretic analyses yield significant insights into the topology of the C. elegans nervous system, revealing features such as the presence of hub nodes and network motifs, as well as *rich club* and *small-world* topologies (Towlson et al. 2013; Varshney et al. 2011).

The ability to attain these insights does not imply that network models explain, however. Explanations must be explanations of something; they necessarily refer to an *explanandum phenomenon*. Although network models describe the structural features of the C. elegans nervous system, they do not show how these features contribute to any specific behavioral capacity (Craver 2016; Miłkowski 2016). Consider *klinotaxis*, a form of goal-directed locomotion in which the worm approaches a chemical source by repeatedly sweeping, and over the long run following, the line of steepest ascent along a chemical gradient whose concentration increases with proximity to the source (Figure 1A). Although the time course and spatial trajectory of C. elegans klinotaxis has already been described in detail (Iino and Yoshida 2009), little is known about the way in which this particular behavior emerges from activity in the worm's nervous system.

In a recent series of studies, Eduardo Izquierdo and colleagues aim to "explore in detail the link between neural connectivity and behavior" (Izquierdo and Beer 2013, 1, see also their 2015; Izquierdo and Lockery 2010; Izquierdo, Williams, and Beer 2015). These studies together embody a three-step investigative strategy that deploys each one of the three kinds of models introduced above.

The first step of this strategy is to develop a graphical representation of the *minimal network* —the smallest structural network deemed capable of producing klinotaxis. To this end, Izquierdo & Beer (2013) begin by identifying the connectome elements that are at least potentially relevant to this particular behavioral capacity. These include all 12 chemosensory neurons that are known to detect concentrations of chemical gradients in the environment, 28 head and neck motor neurons that determine the worm's movement, and all 234 interneurons, 6246 chemical contacts and 890 electrical gap junctions that constitute the structural links between them. From this initial selection, the authors remove all structural intermediaries whose contribution may be deemed negligible or redundant. This includes chemosensory and motor neurons that have not previously been associated with klinotaxis in ablation studies (e.g. Bargmann and Horvitz 1991), as well as all weakly connected elements (such as those that have less than two outgoing connections) and long-range pathways (such as those that go through interneurons not immediately adjacent to either a

chemosensory or a motor neuron). The resultant minimal network (Figure 1B) contains only ASE chemosensory neurons, AIZ and AIY ventral and dorsal interneurons, SMB motor neurons, as well as the anatomical links between them.

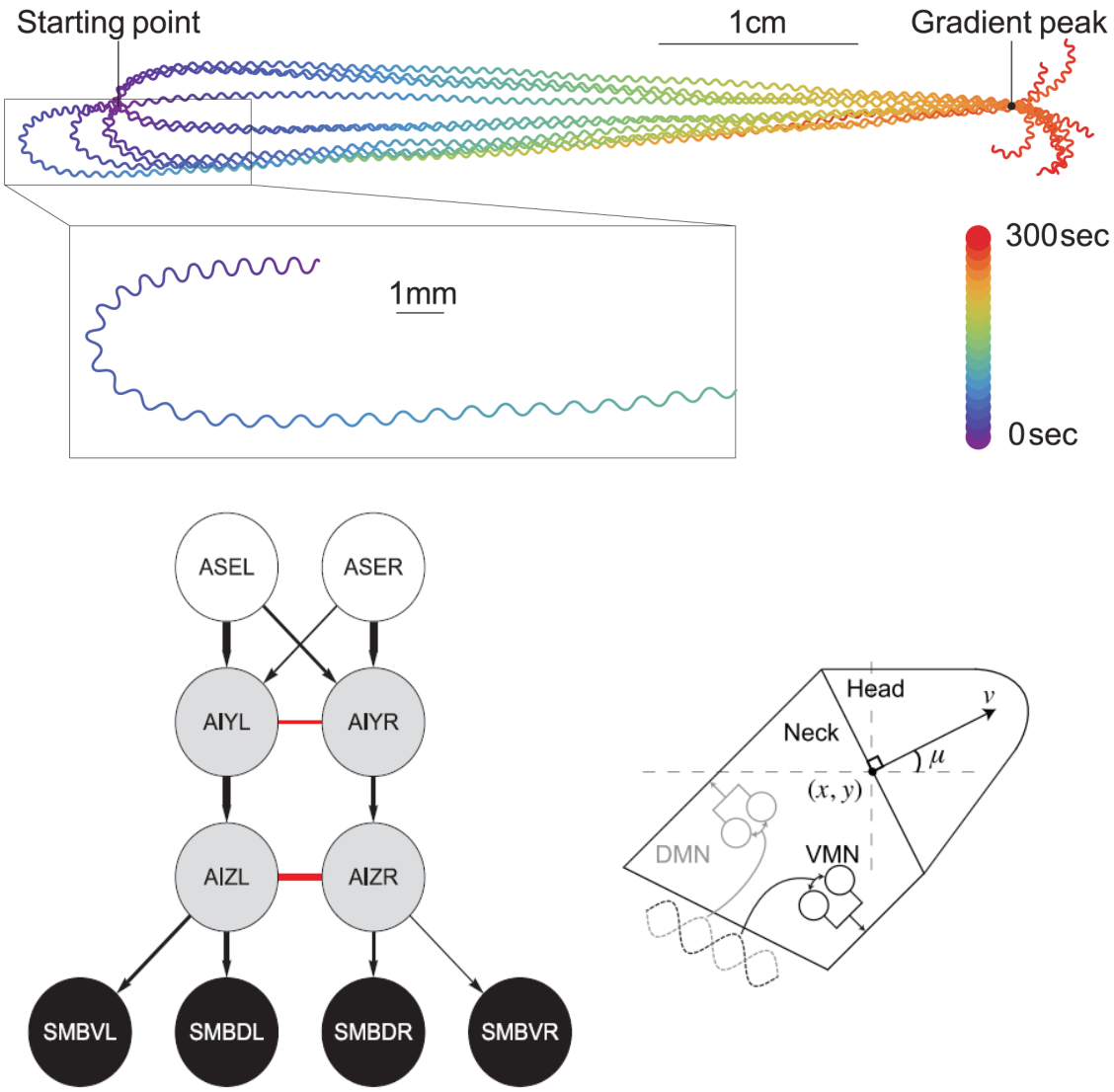

Figure 1: Simulation of C. elegans klinotaxis (Izquierdo & Beer 2013).**A (top):** Klinotaxis in the simulated environment. **B (bottom left):** The minimal network model. **C (bottom right):** The C. elegans body model.

The second step of the investigative strategy is to determine whether, and if so how, the minimal network generates klinotaxis. This is accomplished using a simulation model developed by Izquierdo & Lockery (2010). This model consists of an artificial neural network controller inserted into a C. elegans body model (Figure 2C), and which is situated in a simulated environment (Figure 2A). Izquierdo & Beer (2013, see also their 2015) improve on the original model by replacing the artificial neural network controller with the minimal network derived from the connectome. Using an artificial evolutionary process, they identify circuit parameter values (such as unit biases and connection weights) for producing reliable and efficient klinotaxis in the simulated environment. Notably, the authors observe that simulated klinotaxis bears a close but unexpected qualitative resemblance to klinotaxis in the real world: They both exhibit a sinusoidal relationship between turning bias and bearing, as well as a linear relationship between turning bias and the normal component of the chemical gradient. This unexpected resemblance is viewed as evidence that "the model presented here [may be] especially appropriate for the generation of testable predictions concerning how the biological network functions" (Izquierdo and Beer 2013, 5).

The third and final step of the investigation of C. elegans klinotaxis uses dynamical models to better understand the contribution of individual network elements. In particular, using dynamical models of individual SMB neck motor neurons, Izquierdo & Beer show that these neurons' sensitivity to chemosensory stimulation differs during distinct phases of the worm's oscillatory motion. Because these neurons are arranged symmetrically about the worm's body, the period of highest sensitivity to chemosensory stimulation for the neck motor neuron on one side of the body is also the period of lowest sensitivity for the corresponding neuron on the other side.[4] This ensures that the worm's continuous oscillatory motion will gravitate toward the chemical source along the line of steepest ascent. Thus, it is the dynamics of individual neurons in the minimal network that determines the oscillatory and goal-directed movement characteristic of klinotaxis in simulation. The authors predict that the corresponding network elements in biological C. elegans will exhibit analogous dynamics, and that these dynamics will therefore be similarly important for the production of klinotaxis in the real world.

4 In subsequent work, the investigators also develop an information-theoretic model to show that "each SMB neuron acts as a kind of gate, allowing [information about the change in chemical gradient] from the AIZ layer to pass through at some phases, but blocking or strongly attenuating it at others" (Izquierdo, Williams, and Beer 2015, 14). Intriguingly, the period of highest sensitivity to chemosensory stimulation corresponds to the period in which the informational gate is "open".

3.2. Schizophrenia

It is worth contrasting this recent investigation of C. elegans klinotaxis with past and present work on the network basis of schizophrenia in humans. Although this work also combines network, simulation, and dynamical models, the overarching investigative strategy is different. Most notably, it does not begin with a network model of the human connectome. Although efforts are underway to map the connectome for many other organisms including humans (for a review see Sporns 2012), C. elegans remains the only organism whose connectome has been described in its entirety. Perhaps for this reason, whereas the investigation of C. elegans klinotaxis proceeds "from the bottom up"—beginning with a description of the nervous system—investigations of schizophrenia and many other phenomena instead proceed "from the top down", beginning with descriptions of the phenomena themselves (see also Zednik 2018).

Schizophrenia is characterized by delusions and hallucinations, apathy and loss of affect, and by generally erratic thought and behavior. It has been linked to deficits in the coordination of physically and functionally distributed brain regions (Phillips and Silverstein 2003). One of the most common operationalizations of this coordination is the degree of *neural synchrony* between regions: concurrent, phase-locked oscillatory activity. Dynamical models have long been used to characterize this kind of oscillatory activity in EEG time series data. For example, Uhlhaas et al. (2006) use models of this kind to show that phase-locked oscillatory activity between brain regions is significantly reduced in patients with schizophrenia as compared to healthy individuals; while they are engaged in a variety of perceptual and cognitive tasks, schizophrenics exhibit reduced neural synchrony in the high-frequency beta (13-30Hz) and gamma (30-100Hz) bands (see also Uhlhaas and Singer 2011).

The changing patterns of neural synchrony can be visualized as topological changes in the functional networks whose elements correspond to EEG recording sites, and whose connections are statistical correlations between the time series data recorded at those sites (Figure 2).

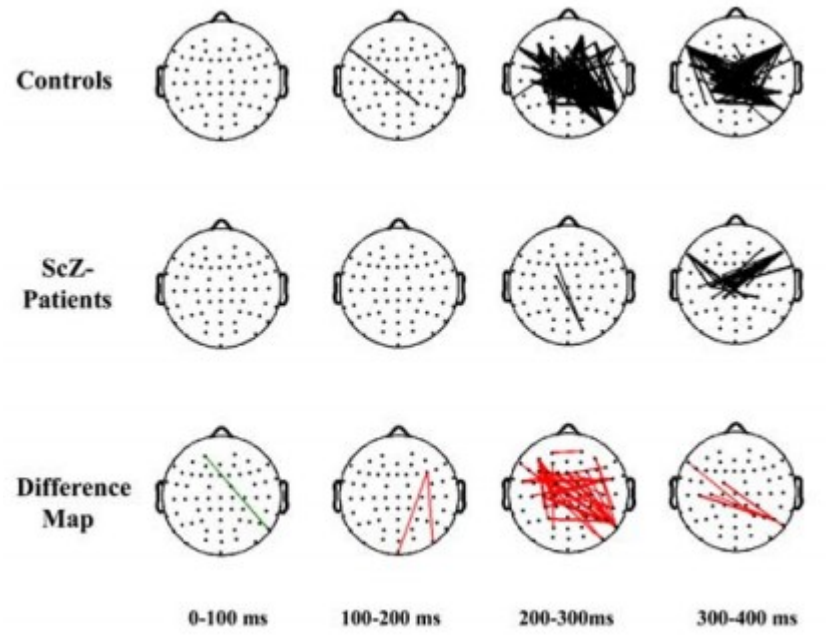

Figure 2: Neural synchrony and schizophrenia (From Uhlhaas & Singer 2006)

The changing topology of such networks can be analyzed using graph-theoretic techniques. Indeed, these techniques have already revealed systematic differences between schizophrenics and healthy individuals: Liu et al. (2008) and Fornito et al. (2011) both identify a reduction in the overall degree of small-world connectivity in between-region functional networks, and Bassett et al. (2008) describe analogous differences in between-region structural networks.

Though these results may be compelling, they still only reveal correlations; it remains unclear whether reduced small-world connectivity is causally relevant to the observed reduction in neural synchrony, and thus, to the emergence of schizophrenia. Because investigators are unable to intervene on brain networks in healthy individuals, many of them instead deploy simulation models in which they are able to intervene on the topology of artificial networks, and measure the resultant behavioral changes. Indeed, Kitano & Fukai (2007) present a simulation model in which small-world networks are more likely to exhibit synchronous activity than other kinds of networks, especially across distant elements (see also Masuda and Aihara 2004). Similarly, Qu et al. (2014)

show that small-world networks with intermediate levels of randomness are more likely to exhibit synchronous activity than regular and highly random networks. By reviewing a series of classic and contemporary simulation models, Friedenberg & Silverman (2011, 222) argue that small-world connectivity is in fact necessary for the emergence of global neural synchrony.

For sure, the epistemic import of these simulation models is bounded by their limited biological plausibility; the networks described often exhibit kinds of small-worldness and synchronicity that bear only a broad qualitative resemblance to the small-worldness and synchronicity that is observed in biological brains. Moreover, there are numerous additional variables (such as firing rates and relative connection strengths) whose contribution to neural synchrony remains unexplored. For this reason, the claim that small-world connectivity is necessary for neural synchrony is far from confirmed.[5] Nevertheless, current simulation models might still be thought to corroborate the claim that reduced small-world connectivity is not only correlated with reduced neural synchrony, but is in fact causally relevant for it.

4. Network neuroscience: A mechanistic construal

These case studies illustrate the characteristic heterogeneity of network neuroscience: Several different kinds of models are used to investigate different kinds of networks in the brain. At the same time, they can be construed as efforts in representing and understanding the mechanisms of behavior and cognition, and thus, as exercises in *mechanistic explanation*.

At the heart of the mechanistic framework is the definition of 'mechanism' as an organized system of parts and operations, the changing properties of which over time exhibit a phenomenon of explanatory interest (Bechtel and Abrahamsen 2010b; see also Craver 2007). If explanation in network neuroscience is to be construed mechanistically, the networks being investigated must fall under this definition; they must be considered *brain network mechanisms*.

The minimal network for C. elegans klinotaxis is naturally viewed as a brain network mechanism. Its neurons, synapses and gap junctions are prototypical component parts, and the

5 Indeed, some simulation models appear to show quite the opposite: Increased synchrony with decreased small-world connectivity (see e.g. Pérez et al. 2011). For current purposes, however, the empirical question of whether small-world connectivity is in fact necessary for neural synchrony is secondary to the methodological point that simulation models are considered instrumental for answering this question.

activity of individual neurons, as well as the propensity of synapses and gap junctions to mediate between neurons, are plausible component operations. Moreover, the network is organized spatiotemporally within the C. elegans nervous system. That said, the spatial aspects of this organization seem less important to the production of klinotaxis than its temporal aspects—specifically, certain neurons' antiphasic sensitivity to chemosensory stimulation. Insofar as there are reasons to believe that the minimal network for klinotaxis behaves in the real world as it does in simulation, there are reasons to view it as a mechanism for that phenomenon.[6]

The networks discussed in the context of schizophrenia can also be viewed as brain network mechanisms, albeit at a much higher level of brain organization, and with a lesser degree of certainty regarding their detailed composition, organization, and responsibility for the explanandum phenomenon. These networks are situated at the level of brain regions—structures that are commonly viewed as the component parts of mechanisms, and whose interactions are traditionally viewed as component operations (see e.g. Bechtel 2008; Boone and Piccinini 2016). That said, current research has not yet identified any single network of regions as the mechanism for schizophrenia and its characteristic effects. Indeed, it remains unclear which regions are involved, and what these regions actually do. Nevertheless, this kind of uncertainty is no reason to deny that there is in fact such a mechanism, and that the research reviewed here aims to uncover it.

Mechanistic explanations center on *representations* of mechanisms (Bechtel 2016; Glennan 2005; Machamer, Darden, and Craver 2000). What it takes to adequately represent a mechanism is neatly expressed in Kaplan & Craver's *model-to-mechanism-mapping constraint* (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)

Motivated by this constraint, it is tempting to emphasize the explanatory contribution of network models, which represent the nodes and connections of networks in biological brains. Indeed, in the

6 Strictly speaking, the mechanism for klinotaxis is likely to extend beyond the boundaries of the minimal network, and to even include extraneural components such as parts of the worm's body (see the discussions in Izquierdo and Beer 2013, 2015; see also Zednik 2011). These parts seem equally important to the production of klinotaxis, and are difficult to dismiss as mere background conditions. Nevertheless, for convenience, here the minimal network will itself be treated as a mechanism for klinotaxis.

case studies introduced above, network models make an invaluable explanatory contribution: Figure 1B depicts the component parts of the network mechanism for klinotaxis; Figure 2 can be viewed as a depiction of the (possible) functional aspects of a brain network mechanism for schizophrenia and its associated effects (cf. Craver 2016. But see also Section 5); and structural network models such as the one developed by Bassett et al. (2008) can be viewed as representations of this mechanism's (possible) structural aspects.

That said, it is a mistake to assume that the representation of brain network mechanisms only involves network models. Indeed, the network model in Figure 1B only depicts the component *parts* of the minimal network for klinotaxis; component operations such as the activity of individual neurons are instead represented in dynamical models of individual neuronal activity. Thus, this case study supports Hochstein's (2016) claim that the representation of mechanisms is often *distributed*, in the sense that many different models are used to represent the parts, operations, and/or organizational properties of a single mechanism. Hochstein's claim is also supported by the fact that not one, but many network models are used to investigate the network basis of schizophrenia. Although it has yet to be determined whether (and if so how) these different models can be integrated, they are arguably designed to represent different (functional vs. structural) aspects of a single brain network mechanism.

Whereas most philosophical discussions of mechanistic explanation center on the representation of mechanisms, mechanistic explanation also typically involves an *understanding* of the way in which a particular mechanism produces the explanandum phenomenon.[7] William Bechtel has argued that this understanding is often achieved through the practice of *dynamic mechanistic explanation,* in which computational models are used to animate a mechanism's behavior over time, and mathematical models are used to characterize that behavior analytically (Bechtel and Abrahamsen 2010b). Indeed, the case studies introduced in Section 3 show that network neuroscientists are regularly engaged in dynamic mechanistic explanation.

7 There is a wide, albeit relatively inconclusive, literature on the relationship between explanation and understanding. A fairly neutral view has recently been expressed by William Bechtel, who claims that understanding is a "common, although not a necessary, goal of explanation" (Bechtel 2016). The present discussion embraces this view: Although it may not be essential, many network neuroscientists do in fact go beyond the goal of representing brain network mechanisms. Specifically, they make a considered effort to understand how these mechanisms give rise to the phenomena being explained. From a philosophical perspective, it is worth characterizing this kind of understanding.

Regarding klinotaxis, inserting the minimal network to control the behavior of the C. elegans body model in simulation serves not only to discover the network parameter values needed to produce reliable and efficient klinotaxis; it also serves to compare the time course and trajectory of klinotaxis in simulation to that of klinotaxis in the real world, as well as to study the dynamics of individual neural units. Thus, simulation and dynamical models together allow Izquierdo and colleagues to show that the minimal network is in fact capable of producing the explanandum phenomenon, as well as to better understand how this happens through the time-sensitive activity of individual neural units.

Simulation and dynamical models play an equally important role in the investigation of schizophrenia. Recall that Kitano & Fukai (2007) use a simulation model to systematically probe the relationship that obtains between small-world connectivity and neural synchrony, and that dynamical models allow Uhlhaas et al. (2006) to precisely describe the kind of neural synchrony that must be generated so as to establish a link with real-world schizophrenia. Thus, these simulation and dynamical models are crucial for establishing the constitutive relevance of particular organizational properties (see Section 6), but also to show that, as well as to understand how, brain network mechanisms actually produce the behavioral and cognitive capacities being explained.

In this way, the case studies introduced in Section 3 can be construed mechanistically. Insofar as they are representative of the field of network neuroscience as a whole, they support the claim that network neuroscientists aim to deliver mechanistic explanations. Network neuroscientists are concerned with brain network mechanisms, aim to represent these mechanisms, and attempt to understand how these mechanisms give rise the phenomena being explained. Moreover, the case studies introduced above show that it would be a mistake to consider either one of network, simulation, and dynamical models in isolation. Rather, models of all three kinds are combined in the service of mechanistic explanation.

That said, it is also instructive to consider some worries that have recently arisen about the explanatory contribution of network, dynamical, and simulation models in network neuroscience. Addressing these worries will not only bolster the mechanistic construal articulated above, but will also yield a deeper understanding of mechanistic explanation itself.

5. Do functional network models explain?

One particularly nuanced view has been articulated by Carl Craver (2016). Craver acknowledges that structural network models are used to represent a network mechanism's component parts, and that effective network models are well-suited for representing its causally-relevant operations. At the same time, however, Craver deems functional network models to be "non-explanatory" (Craver 2016, 702). Although such models may facilitate the discovery of brain network mechanisms (Craver 2016, 705–6), he argues that they do not themselves represent these mechanisms, and thus, that they fall short as mechanistic explanations. Insofar as the investigation of schizophrenia—like many other research efforts in network neuroscience—leans heavily on models of functional networks, Craver would likely deny that this investigation delivers (or even aims to deliver) mechanistic explanations.

Craver's reasoning can be better understood by considering in more detail the functional networks discussed in the schizophrenia case study. First, consider their connections. Recall from Section 2 that two elements are connected functionally if their activity is correlated statistically. Mere statistical correlations do not contribute to the production of behavioral or cognitive phenomena. Indeed, these correlations "underdetermine the causal and anatomical structures that presumably produce [them]" (Craver 2016, 705). Thus, the connections within a functional network do not correspond to a mechanism's causally-relevant operations.

Next, consider a functional network's elements. The networks explored by Uhlhaas et al. (2006) are composed of elements that correspond to EEG recording sites. The networks discussed by Fornito et al. (2011) are defined over $1mm^3$ voxels of brain tissue. In both cases, elements are individuated using parcellation schemes that are highly pragmatic in the sense of Section 2 above. Craver decries pragmatically-individuated elements as artifacts that cannot generally be viewed as the *working parts* of mechanisms—parts that perform specific component operations (see also Bechtel 2008). Indeed, there is no intuitive reason to believe that the spatial resolution of an fMRI scanner, or the placement of an electrode by a laboratory assistant, reliably tracks the boundaries of a mechanism's causally relevant parts.

Because their elements should not be considered component parts, and because their connections should not be considered component operations, Craver concludes that functional networks should not be considered mechanisms. For this reason, whereas research efforts in

network neuroscience that center on the representation of structural or effective networks (like the investigation of C. elegans klinotaxis) can and should be thought to deliver mechanistic explanations, research efforts that depend on the representation of functional networks (like the investigation of schizophrenia) should not.

Craver is right to be weary of the ontological status of functional networks. Nevertheless, his conclusion seems premature: Even if functional networks cannot be viewed as mechanisms, it does not follow that models of functional networks do not describe explanatorily relevant features of such mechanisms. Indeed, models of functional networks may still provide first approximations of a mechanism's component parts and operations, and may therefore be considered *mechanism sketches*: representations of mechanisms that are not perfectly detailed, but that require elaboration and possibly even correction (Craver 2007; Machamer, Darden, and Craver 2000). Importantly, mechanism sketches are themselves genuinely explanatory. Insofar as all models in science are imperfectly detailed, they are "sketchy" to some degree or another. Because it would be unreasonable to conclude that no models in science are genuinely explanatory, mechanism sketches must themselves be deemed explanatory at least to a certain extent. Insofar as functional network models are mechanism sketches, they are genuinely explanatory, and play a more substantial role in mechanistic explanation than the one attributed to them by Craver.

Consider again a functional network's connections. Although Craver is right to observe that statistical correlations underdetermine the presence of anatomical or causal connections, this does not mean that the former contain no information about the latter. To see why, it is important to understand more precisely how network neuroscientists go about uncovering causal interactions in the brain. There is no measurement technique or device that can be used to directly observe these interactions. Indeed, network neuroscientists can only rely on more-or-less reliable indicators, and presumably for this reason prefer to speak of 'effective' rather than 'causal' connectivity.[8] As it happens, one of the most influential indicators of this kind is the information-theoretic measure of *Granger causality*, which quantifies the ability to predict one data series from another (Granger 1969). Defined in Granger-causal terms, effective connectivity is grounded in the same kind of statistical information that is used to define functional connectivity. Indeed, because it depends on "inferences about statistical dependencies over time...Granger causality could be regarded as a measure of lagged functional connectivity" (Friston 2011, 20). Because they are based on the same

8 Curiously, while Craver argues that the term 'functional connectivity' is misleading, he does not recognize that, in this context, the same ought to be said of the term 'causal connectivity'.

kind of statistical information, effective networks, like functional networks, underdetermine true causal interactions in the brain.

This observation might be taken to suggest that effective network models are no more explanatory than their functional counterparts. In line with Craver's suggestion that functional network models contribute to the discovery of mechanisms but not to their representation, it might also be said that effective network models are useful but non-explanatory. But this conclusion seems excessively stringent. At least since Hume, it is clear that there is no way to directly observe causal interactions in the brain or anywhere else. Unless one is prepared to deny the explanatory adequacy of causal models in general, therefore, it seems unfair to criticize network neuroscientists for relying on indirect correlation-based indicators of causal interactions in the brain. Thus, a more permissive conclusion seems warranted: Although functional and effective network models similarly underdetermine causal interactions in the brain, they may nevertheless provide information about these interactions. That is, both kinds of models represent mechanisms, albeit to different degrees of completeness and accuracy. In this sense, functional and effective networks alike are genuinely explanatory mechanism sketches, albeit at different degrees of sketchiness. Although the former may require more elaboration and correction than the latter, they both convey some information about the component operations of mechanisms, and in this sense, contribute to mechanistic explanations.[9]

Now, consider again a functional network's elements. Here, Craver is concerned about the fact that one cannot in general expect pragmatically-individuated elements such as electrode recording sites and fMRI voxels to pick out the working parts of mechanisms. Although Craver does not say so explicitly, his concern is probably motivated by the intuitive idea that the working parts of mechanisms correspond to nature's proverbial joints, and that it is hard to imagine these joints to be reliably picked out by the resolution of an fMRI scanner or by the placement of

9 There may be some residual fuzziness in the distinction between discovery and description (and arguably, confirmation as well): When does a model describe (or represent) a mechanism, as opposed to merely facilitating its discovery (or confirmation)? The present discussion embraces a permissive view, in which a model is descriptive (and thus, genuinely explanatory), when it contains information about the relevant mechanism. This seems in line with the 3M constraint discussed above. Nevertheless, Craver's stance may be more restrictive. More careful philosophical analysis may be required to properly distinguish between discovery, description, and confirmation —or to challenge the assumption that such a distinction can and should be made in the first place.

electrodes. Indeed, they seem more likely to be tracked by principled parcellation schemes that are grounded in e.g. the principles of cell biology.

Although this idea is intuitive, it is important to be cognizant of the practical limitations of scientific research. Whereas principled parcellation schemes are readily available at lower levels of brain organization, it is unclear which principles apply at higher levels. Indeed, there is no general agreement about the functional significance of columns and other kinds of microcircuits (see e.g. Horton and Adams 2005), and little consensus regarding the best way to delineate brain regions.[10] Therefore, if principled parcellation schemes are considered the only reliable means of identifying the working parts of brain network mechanisms, then it is doubtful that mechanisms can (at least presently) be identified at anything higher than cellular levels.

But of course, the philosophical and scientific literature is replete with appeals to mechanisms at medium and high levels of brain organization, including the level of microcircuits and the level of brain regions (see discussion in e.g. Bechtel 2008; Boone and Piccinini 2016). How can these appeals be squared with the difficulties associated with individuating the working parts of high-level brain network mechanisms? One option is to defer to principled schemes that may or may not be developed in the future (see e.g. Glasser et al. 2016 for one recent proposal). In this case, one might accept Craver's rejection of functional network models with pragmatically-individuated elements insofar as they do not (yet) deliver genuinely explanatory information. Another option, however, is to acknowledge that even pragmatically-individuated network elements can (at least approximately) pick out the working parts of mechanisms. In this second case, functional network models are again revealed as being mechanism sketches.

Several considerations speak in favor of the second option. For one, pragmatic parcellation schemes may be thought to offer a unique *perspective* on the problem of carving the brain so as to identify a mechanism's working parts (Craver 2013). Recall that the working parts of mechanisms are just those parts that perform specific component operations (Bechtel 2008). Now, consider that the component operations of mechanisms are generally viewed as those that contribute to the phenomena being explained (see e.g. Craver 2007). This implies that the primary desideratum for designating a chunk of brain tissue to be a working part of a mechanism is not that its boundaries

10 For example, although Brodmann's scheme of individuating brain regions according to cytoarchitectural principles is relatively principled and still widely used today, it is also well-known to cut across functional boundaries.

accord with certain theoretical principles, but just that it contributes to the phenomenon of explanatory interest. Many phenomena—schizophrenia included—are associated with (or even operationalized in terms of) characteristic BOLD signatures and EEG time series data. Because it is easy to measure the contribution of individual voxels and electrode recording sites to such signatures and data, pragmatic parcellation schemes offer a particularly straightforward way of identifying the contribution of individual "chunks" of brain tissue. That is, these schemes offer a uniquely accessible perspective from which to individuate a mechanism's working parts. Of course, it seems likely that this perspective can eventually be improved by supplementing it with other methods—including methods grounded on principled parcellation schemes. Nevertheless, for the purposes of deriving a "first pass" approximation, it seems foolish to dismiss the explanatory import of pragmatic ways of individuating working parts just because they are pragmatic.

There is another way to acknowledge the explanatory import of pragmatically-individuated elements which does not depend on a perspectival approach to the problem of individuating working parts. Because pragmatic parcellation schemes are grounded in measuring techniques such as fMRI and EEG, they can often be used to infer the approximate locations and boundaries of a mechanism's working parts. Although any individual voxel or recording site may not itself be considered a part, its recorded activity in the context of some particular behavioral or cognitive phenomenon suggests that it may *belong* to such a part. Of course, methods such as fMRI have considerable epistemic limitations (Klein 2010). Indeed, Craver denies that BOLD signals constitute the working parts of mechanisms because their "time courses are (presumably) too slow to be part of how the brain performs cognitive tasks" (Craver 2016, 705). Nevertheless, although BOLD signals might not be measures of neural activity *per se*, they might still constitute more-or-less reliable indicators thereof.[11] For this reason, the results of fMRI (or EEG) experiments may be thought to indicate that a particular region of interest is in fact involved in the production of an explanandum phenomenon. Of course, the regional boundaries suggested by such experiments are likely to be rough. However, this again just goes to show that functional network models based on pragmatic parcellation schemes deliver mechanism sketches, rather than ideally complete explanations. This is a much more permissive conclusion than the one originally suggested by Craver, according to which functional network models have no explanatory import at all, and are

11 This argument parallels the argument given in the discussion of functional connectivity above: BOLD signals (like statistical correlations) might not be measures of neural activity (or of causal interactions) *per se*, but might nevertheless be more-or-less reliable indicators thereof.

useful only insofar as they facilitate the process of mechanism-discovery.

The upshot of this discussion is that it seems wrong to deny the explanatory import of functional network models. Although the connections within functional networks consist of statistical correlations, they nevertheless carry information about a mechanism's causally-relevant component operations. Moreover, although the elements of such networks may be pragmatically individuated chunks of brain tissue, they can still be used to approximate a mechanism's component parts. Thus, functional network models do not merely facilitate the discovery of brain network mechanisms; they are genuine, albeit sketchy, representations thereof. Of course, it seems likely that such sketchy representations can be improved by supplementing functional network models with e.g. information about structural detail. This, however, does not undermine their status as representations; it merely reflects the typical way in which mechanism sketches are gradually transformed into increasingly detailed mechanistic explanations.

6. Toward organization-centric mechanistic explanation

The previous section was concerned with the charge that some models in network neuroscience do not deliver mechanistic explanations because they do not represent mechanisms. A different concern is that, although these models represent mechanisms, they do not represent them *in the right way*. There are at least two variants of this concern. The first is that whereas network, simulation, and dynamical models are used to capture mechanisms' *abstract* (mathematical) properties, mechanistic explanations may be thought to require a representation of concrete (physical) properties. Although the issue of abstraction has been at the center of recent debate (Boone and Piccinini 2016; Chirimuuta 2014; Craver and Kaplan 2018; Levy and Bechtel 2013), the present discussion will focus on a different issue: Whereas many network, simulation, and dynamical models emphasize a mechanism's *organizational* properties, mechanistic explanations are often thought to emphasize the intrinsic properties of its components.

Indeed, network, simulation, and dynamical models appear purpose-built for describing a network mechanism's organization. Although network modeling involves the use of parcellation and connectivity schemes to identify network elements and connections, individual elements and connections are only rarely described in detail. Indeed, network models are commonly treated as

starting points for sophisticated graph-theoretic analyses designed to reveal global topological features such as the overall degree of clustering and modularity, small-worldness and connection density. The development of simulation models is similarly focused on organization. These models are not normally used to understand the contribution of individual elements and connections—Izquierdo and colleagues' investigation of C. elegans klinotaxis notwithstanding—but rather to evaluate the influence of global parameters on a network's overall behavior. Finally, although dynamical models are sometimes used to characterize the dynamics of individual neural units, in the study of large brain network mechanisms they are more commonly used to describe patterns of distributed activity in a network as a whole. Insofar as they are used to highlight organizational properties, the models deployed in network neuroscience are characteristically *organization-centric*.

Several philosophical commentators have taken the organization-centricity of network, simulation, and dynamical models to be antithetical to the principles of mechanistic explanation. Of course, no-one takes a mechanism's organization to be irrelevant to its behavior. Indeed, the notion of organization is explicit in all current definitions of 'mechanism', including the one adopted in Section 4. Moreover, several previous discussions of mechanistic explanation have sought to characterize the modes of organization that biological mechanisms typically exhibit (see e.g. Bechtel and Abrahamsen 2010a; Bechtel and Richardson 1993; Craver 2007; Glennan 2010; Levy and Bechtel 2013). Nevertheless, Silberstein & Chemero (2013), like Woodward (2013), assume that mechanistic explanations characteristically center on a mechanism's composition rather than its organization (see also Rathkopf 2015). In particular, Silberstein & Chemero argue that the "basic units of [mechanistic] explanation" are not organizational properties, but rather "the parts of the mechanism and the operations those parts perform" (Silberstein and Chemero 2013, 961).[12] Woodward invokes the language of difference-making to express a similar view: In mechanistic explanation, he considers the difference-making factors to be "causal relationships between components" (Woodward 2013, 50), and argues that models deliver mechanistic explanations only when they show "how a system behaves by decomposition into components and behaviours intrinsic to these components" (Woodward 2013, 54). In other words, whereas network

12 This view stems in part from the Chemero & Silberstein's focus on decomposition and localization: heuristic strategies for mechanism-discovery that involve isolating individual parts and characterizing their associated operations (Bechtel and Richardson 1993). Indeed, Chemero & Silberstein (2013, 961) think of these strategies as "the sine qua non of mechanistic explanation." That said, it has already been argued elsewhere that it is a mistake to "[raise] a fallible heuristic to the status of normative constraint." (Stinson 2016, 1586; see also Zednik 2014, 2015).

neuroscience is relatively organization-centric, these critics take mechanistic explanations to be characteristically *component-centric*.[13]

It is easy to see why these critics would consider some research efforts in network neuroscience to have abandoned the principles of mechanistic explanation. Whereas the investigation of C. elegans klinotaxis explains by characterizing the individual contribution of nodes and connections—a prototypically mechanistic approach—the intrinsic properties of specific brain areas are of little concern in the investigations of schizophrenia presented in Section 3. These investigations are driven by the suspicion that the emergence of schizophrenia and its associated effects can be explained by understanding the overall degree of small-world connectivity. Therefore, the difference-making "units of explanation" are not the intrinsic activities of individual regions, but rather the small-world properties of a network mechanism's overall organization. In this sense, at least one of the two case studies reviewed above does not reflect the principles of mechanistic explanation as understood by Silberstein & Chemero and Woodward.

The remainder of this section aims to show that these critics' component-centric conception of mechanistic explanation is misleading. Indeed, the organization-centricity of network neuroscience is a red herring: It matters not that network, simulation, and dynamical models emphasize organizational properties over the properties of individual components. What matters for the purposes of mechanistic explanation is just that the properties being emphasized are in fact *properties of mechanisms*.

Before that, however, a concession: Although the idea of organization has always been central to the philosophical conception of mechanistic explanation, philosophers have a relatively limited understanding of the scientific methods and practices that are actually used to uncover, represent, and understand a mechanism's organization. For sure, progress has recently been made by reflecting on the representational capacities of network models (see in particular Bechtel 2015;

13 Note that abstraction may be a particularly effective way of characterizing a mechanism's organization (Levy and Bechtel 2013). For this reason, philosophical discussions of organization are often intertwined with discussions of abstraction—so too the discussions by Silberstein & Chemero (2013) and Woodward (2013). Indeed, it is tempting to view both of these discussions as being primarily or exclusively concerned with abstraction. Nevertheless, the quotations reproduced in this paragraph—which make no mention of abstraction—suggest that issues of organization can be raised quite independently.

Craver 2016; Levy and Bechtel 2013). Nevertheless, it remains quite unclear how practicing scientists go about identifying a mechanism's organization, and how they determine which organizational properties contribute to the production of any particular explanandum phenomenon. Therefore, there is reason to believe that detailed attention to the way in which network, simulation, and dynamical models are used to characterize the organization of network mechanisms in the brain can deliver an improved account of mechanistic explanation more generally.

In order to better understand the way scientists go about identifying and understanding a mechanism's organization, it is worth considering first how they go about uncovering its composition. A good starting point is therefore Carl Craver's influential *mutual manipulability* account of constitutive relevance:

> **MM:** "a part is a component in a mechanism if one can change the behavior of the mechanism as a whole by intervening to change the component and one can change the behavior of the component by intervening to change the behavior of the mechanism as a whole" (Craver 2007, 141).

MM can be used to determine when a particular structure or function is in fact a component part or operation of the mechanism for some explanandum phenomenon. Indeed, it captures many of the "top-down" and "bottom-up" experiments typically used by neuroscientists to determine the composition of mechanisms in the brain, including lesion studies and stimulation experiments, as well as fMRI and single-cell activation studies (see Craver 2007 for an extended discussion).[14] Notice, however, that MM does not say anything about organization. This is particularly noteworthy given that Craver's stated aim is to characterize "the experimental strategies that neuroscientists use to test whether a given entity, activity, property, *or organizational feature* is relevant to the behavior of the mechanism as a whole" (Craver 2007, 140, emphasis added). MM does not fully achieve Craver's aim; it fails to account for the experimental strategies that are used to uncover a mechanism's organization.

Despite falling short, Craver's mutual manipulability account can be modified slightly to accommodate these strategies:

14 Craver's mutual manipulability account of constitutive relevance has recently been the target of criticism (see e.g. Baumgartner and Gebharter 2016). Even if this criticism shows that MM is defective as a metaphysical account of constitutive relevance, however, MM does much to illuminate the use of "top-down" and "bottom-up" experiments for uncovering a mechanism's composition. Thus, even if the extension of MM proposed below turns out to be similarly defective, it will have served an important purpose if it can illuminate the use of analogous "top-down" and "bottom-up" experiments for uncovering a mechanism's organization.

> **MM-O:** A topological feature is an organizational property of a mechanism if one can change the behavior of the mechanism as a whole by intervening to change that topological feature, and one can change the topological feature by intervening to change the behavior of the mechanism as a whole.

MM-O can be used to determine when a particular topological feature—such as a network's degree of small-worldness—is in fact an organizational property of the mechanism for some explanandum phenomenon. Notably, MM-O captures many of the organization-centric modeling practices deployed in network neuroscience, and reveals them to be platforms on which to conduct close analogues of the "top-down" and "bottom-up" experiments mentioned previously. For example, in the schizophrenia case study, simulation models are used to show that changes in network topology lead to the kinds of changes in neural synchrony that are characteristic of the difference between schizophrenic and healthy individuals. This use of simulation models is analogous to "bottom-up" interference and stimulation studies discussed in the original context of MM. Moreover, by comparing the large-scale topology of brain networks in schizophrenics with the topology of brain networks in healthy individuals, investigators are able to determine whether the observed behavioral differences go hand in hand with topological differences in the brain. Thus, network models are used in the context of "top-down" experiments in which cognitive or behavioral changes bring about changes in the organization of the underlying mechanism.[15]

In this way, MM-O can be used to better understand the way network, simulation, and dynamical models are used to uncover the organization of network mechanisms in the brain. But to what extent is MM-O a legitimate extension of the mechanistic framework? Silberstein & Chemero, as well as Woodward, have already worried that accommodating organization-centric modeling practices under the mechanistic banner would require this banner to be illegitimately "enlarged" (Woodward 2013) or "stretched" (Silberstein and Chemero 2013) to a point where almost "all explanations in biology automatically turn out to be 'mechanistic'" (Woodward 2013, 64; see also Silberstein and Chemero 2013, 958).

These worries are unfounded, however. First, recall that it is within the stated ambitions of

15 It would be wrong to think that the "top-down" and "bottom-up" strategies for revealing a mechanism's organization can be reduced to the corresponding strategies for uncovering its component parts and operations. Although a mechanism's organizational properties supervene on the properties of individual components, the former may be difference-makers while the latter are not. As has already been discussed by e.g. Woodward (2013), a network's behavior may be robust with respect to changes in any particular node or connection, as long as the overall organization (e.g. the degree of small-worldness) remains unchanged.

Craver's original work to also account for experimental strategies for uncovering a mechanism's organization. For this reason, MM-O should not be viewed as an illegitimate "enlargement" or "stretch", but only as a way of tying up loose ends. Second, MM-O still leaves ample room for non-mechanistic explanations in neuroscience and beyond. This second point bears elaboration. Silberstein & Chemero, like Woodward, accept that manipulationist principles are instrumental in contemporary network neuroscience. Thus, they may acknowledge that network, simulation, and dynamical models can be used to perform the kinds of "top-down" and "bottom-up" experiments described above. Nevertheless, Silberstein & Chemero in particular deny that the use of these experiments signifies a commitment to mechanistic explanation:

> "explanations in [network] neuroscience are consistent with manipulationist or interventionist theories of explanation in general. Indeed, not just structural decompositions but also dynamical and graphical explanations can be and often are interventionist explanations. Mechanistic accounts of explanation that focus on localization and decomposition have no monopoly on interventionist explanation. There is nothing that says the knobs being tweaked must be structural components; they can also be nomological and graphical features." (Silberstein and Chemero 2013, 969; see also Woodward 2013, 63)

This statement is correct insofar as not every use of interventionist techniques is tantamount to a commitment to the norms and practices of mechanistic explanation, and moreover, that not all graphical and dynamical modeling methods are used to deliver mechanistic explanations. Indeed, when Silberstein & Chemero write about intervening on 'nomological features,' they are presumably referring to the kinds of interventions that Craver (2007) discusses in the context of *etiological* relevance: Interventions on the antecedent conditions or causes of the explanandum phenomenon. These interventions do not in fact indicate a commitment to mechanistic explanation —though they are not, of course, incompatible therewith.

Nevertheless, these kinds of interventions must be distinguished from interventions that are used to determine *constitutive* relevance, i.e. interventions on the composition and organization of the mechanism responsible for the phenomenon being explained. Systematic use of these interventions does in fact indicate a commitment to the norms and practices of mechanistic explanation, and establishing this fact is one of principal burdens that Craver's philosophical work tries to bear. Notably, the kinds of experiments that fall under MM-O are interventions of exactly this kind: They are used to reveal that certain topological features are in fact the organizational properties of mechanisms. Thus, explanations in network neuroscience are mechanistic not because they invoke interventions, but because they invoke interventions to uncover the composition and/or

organization of network mechanisms in the brain.

In summary, MM-O shows how network, simulation, and dynamical models contribute to mechanistic explanations in network neuroscience. Far from being antithetical to the principles of mechanistic explanation, these models allow investigators to achieve the important but ill-understood goal of uncovering, representing, and understanding a particular mechanism's organization. Rather than worry that an inclusion of these models threatens to illegitimately "enlarge" or "stretch" the mechanistic framework, philosophers should consider their use in network neuroscience so as to further develop that framework.

7. Conclusion

The preceding discussion sought to construe network neuroscience in mechanistic terms. Two case studies were used to illustrate the use of network, simulation, and dynamical models, but also to show how these different kinds of models are combined for the purposes of representing and understanding brain network mechanisms. By considering two prominent concerns, it was also possible to attain a more nuanced understanding of the particular way in which network, simulation, and dynamical models contribute to mechanistic explanation. Although it is tempting to dismiss the explanatory import of functional network models that describe statistical correlations between pragmatically-individuated "chunks" of brain tissue, a closer look reveals that these models constitute genuinely explanatory mechanism sketches. Moreover, although the use of organization-centric models in network neuroscience may differ in many ways from the canonical examples of mechanistic explanation in other disciplines, greater attention to these models may reveal how neuroscientists are not only able to uncover, represent, and understand a mechanism's composition, but also its organization.

8. References

Bargmann, C. I., and H. R. Horvitz. 1991. “Chemosensory Neurons with Overlapping Functions Direct Chemotaxis to Multiple Chemicals in C. Elegans.” *Neuron* 7 (5): 729–42.

Bassett, Danielle S., E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyer-Lindenberg. 2008. “Hierarchical Organization of Human Cortical Networks in Health and Schizophrenia.” *Journal of Neuroscience* 28 (37): 9239–48. https://doi.org/10.1523/JNEUROSCI.1929-08.2008.

Baumgartner, Michael, and Alexander Gebharter. 2016. “Constitutive Relevance, Mutual Manipulability, and Fat-Handedness.” *The British Journal for the Philosophy of Science* 67 (3): 731–56.

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

———. 2015. “Generalizing Mechanistic Explanations Using Graph-Theoretic Representations.” In *Explanation in Biology*, edited by Pierre-Alain Braillard and Christophe Malaterre, 199–225. Springer.

———. 2016. “Using Computational Models to Discover and Understand Mechanisms.” *Studies in History and Philosophy of Science Part A* 56: 113–21.

Bechtel, William, and Adele Abrahamsen. 2010a. “Complex Biological Mechanisms: Cyclic, Oscillatory, and Autonomous.” In *Philosophy of Complex Systems*, edited by C. A. Hooker. Elsevier.

———. 2010b. “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): 321–33. https://doi.org/10.1016/j.shpsa.2010.07.003.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT Press ed. Cambridge, Mass: MIT Press.

Boone, Worth, and Gualtiero Piccinini. 2016. “The Cognitive Neuroscience Revolution.” *Synthese* 193 (5): 1509–34. https://doi.org/10.1007/s11229-015-0783-4.

Brodmann, Korbinian. 1909. *Vergleichende Lokalisationslehre Der Grosshirnrinde*. Leipzig: Johann Ambrosius Barth.

Bullmore, Ed, and Olaf Sporns. 2009. “Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems.” *Nature Reviews Neuroscience* 10 (3): 186–98. https://doi.org/10.1038/nrn2575.

Chirimuuta, M. 2014. “Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience.” *Synthese* 191 (2): 127–53. https://doi.org/10.1007/s11229-013-0369-y.

Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

———. 2013. “Functions and Mechanisms: A Perspectivalist View.” In *Functions: Selection and Mechanisms*, 133–158. Springer. http://link.springer.com/chapter/10.1007/978-94-007-5304-4_8.

———. 2016. “The Explanatory Power of Network Models.” *Philosophy of Science* 83 (5): 698–709.

Craver, Carl F., and David M. Kaplan. 2018. “Are More Details Better? On the Norms of Completeness for Mechanistic Explanations.” *The British Journal for the Philosophy of Science*.

Fornito, Alex, Jong Yoon, Andrew Zalesky, Edward T. Bullmore, and Cameron S. Carter. 2011. “General and Specific Functional Connectivity Disturbances in First-Episode Schizophrenia During Cognitive Control Performance.” *Biological Psychiatry* 70 (1): 64–72. https://doi.org/10.1016/j.biopsych.2011.02.019.

Fornito, Alex, Andrew Zalesky, and Michael Breakspear. 2013. “Graph Analysis of the Human Connectome: Promise, Progress, and Pitfalls.” *NeuroImage* 80 (October): 426–44. https://doi.org/10.1016/j.neuroimage.2013.04.087.

Friedenberg, Jay D., and Gordon W. Silverman. 2011. *Cognitive Science: An Introduction to the Study of Mind*. 2nd ed. SAGE Publications.

Friston, Karl J. 1994. “Functional and Effective Connectivity in Neuroimaging: A Synthesis.” *Human Brain Mapping* 2 (1–2): 56–78.
———. 2011. “Functional and Effective Connectivity: A Review.” *Brain Connectivity* 1 (1): 13–36. https://doi.org/10.1089/brain.2011.0008.
Glasser, Matthew F., Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, et al. 2016. “A Multi-Modal Parcellation of Human Cerebral Cortex.” *Nature* 536 (7615): 171–78. https://doi.org/10.1038/nature18933.
Glennan, Stuart. 2005. “Modeling Mechanisms.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 443–64. https://doi.org/10.1016/j.shpsc.2005.03.011.
———. 2010. “Mechanisms, Causes, and the Layered Model of the World.” *Philosophy and Phenomenological Research* 81 (2): 362–381.
Granger, C. W. J. 1969. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica* 37 (3): 424–38.
Hochstein, Eric. 2016. “One Mechanism, Many Models: A Distributed Theory of Mechanistic Explanation.” *Synthese* 193 (5): 1387–1407. https://doi.org/10.1007/s11229-015-0844-8.
Horton, J. C., and D. L. Adams. 2005. “The Cortical Column: A Structure without a Function.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1456): 837–62. https://doi.org/10.1098/rstb.2005.1623.
Huneman, Philippe. 2010. “Topological Explanations and Robustness in Biological Sciences.” *Synthese* 177 (2): 213–45. https://doi.org/10.1007/s11229-010-9842-z.
Iino, Yuichi, and Kazushi Yoshida. 2009. “Parallel Use of Two Behavioral Mechanisms for Chemotaxis in Caenorhabditis Elegans.” *Journal of Neuroscience* 29 (17): 5370–80.
Izhikevich, Eugene M. 2007. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Computational Neuroscience. Cambridge, Mass: MIT Press.
Izquierdo, Eduardo J., and Randall D. Beer. 2013. “Connecting a Connectome to Behavior: An Ensemble of Neuroanatomical Models of C. Elegans Klinotaxis.” Edited by Lyle J. Graham. *PLoS Computational Biology* 9 (2): e1002890. https://doi.org/10.1371/journal.pcbi.1002890.
———. 2015. “An Integrated Neuromechanical Model of Steering in C. Elegans.” In *Proceedings of the European Conference on Artificial Life*, 199–206. https://doi.org/10.7551/978-0-262-33027-5-ch040.
Izquierdo, Eduardo J., and S. R. Lockery. 2010. “Evolution and Analysis of Minimal Neural Circuits for Klinotaxis in Caenorhabditis Elegans.” *Journal of Neuroscience* 30 (39): 12908–17. https://doi.org/10.1523/JNEUROSCI.2606-10.2010.
Izquierdo, Eduardo J., Paul L. Williams, and Randall D. Beer. 2015. “Information Flow through a Model of the C. Elegans Klinotaxis Circuit.” Edited by Gennady Cymbalyuk. *PLoS One* 10 (10): e0140397. https://doi.org/10.1371/journal.pone.0140397.
Kaplan, David Michael, and Carl F. Craver. 2011. “The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective.” *Philosophy of Science* 78 (4): 601–627.
Kitano, Katsunori, and Tomoki Fukai. 2007. “Variability V.s. Synchronicity of Neuronal Activity in Local Cortical Network Models with Different Wiring Topologies.” *Journal of Computational Neuroscience* 23 (2): 237–50. https://doi.org/10.1007/s10827-007-0030-1.
Klein, Colin. 2010. “Images Are Not the Evidence in Neuroimaging.” *The British Journal for the Philosophy of Science* 61 (2): 265–278.
Kostić, Daniel. 2016. “The Topological Realization.” *Synthese*, 1–20.
Levy, Arnon, and William Bechtel. 2013. “Abstraction and the Organization of Mechanisms.” *Philosophy of Science* 80 (2): 241–61. https://doi.org/10.1086/670300.
Liu, Y., M. Liang, Y. Zhou, Y. He, Y. Hao, M. Song, C. Yu, H. Liu, Z. Liu, and T. Jiang. 2008. “Disrupted Small-World Networks in Schizophrenia.” *Brain* 131 (4): 945–61. https://doi.org/10.1093/brain/awn018.
Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. “Thinking about Mechanisms.” *Philosophy of Science* 67 (1): 1–25.
Masuda, Naoki, and Kazuyuki Aihara. 2004. “Global and Local Synchrony of Coupled Neurons in Small-

World Networks." *Biological Cybernetics* 90 (4): 302–9. https://doi.org/10.1007/s00422-004-0471-9.
Miłkowski, Marcin. 2016. "Explanatory Completeness and Idealization in Large Brain Simulations: A Mechanistic Perspective." *Synthese* 193 (5): 1457–78. https://doi.org/10.1007/s11229-015-0731-3.
Pérez, Toni, Guadalupe C. Garcia, Víctor M. Eguíluz, Raúl Vicente, Gordon Pipa, and Claudio Mirasso. 2011. "Effect of the Topology and Delayed Interactions in Neuronal Networks Synchronization." Edited by Matjaz Perc. *PLoS One* 6 (5): e19900. https://doi.org/10.1371/journal.pone.0019900.
Phillips, W. A., and S. M. Silverstein. 2003. "Convergence of Biological and Psychological Perspectives on Cognitive Coordination in Schizophrenia." *Behavioral and Brain Sciences* 26 (1): 65–82.
Qu, Jingyi, Rubin Wang, Chuankui Yan, and Ying Du. 2014. "Oscillations and Synchrony in a Cortical Neural Network." *Cognitive Neurodynamics* 8 (2): 157–66. https://doi.org/10.1007/s11571-013-9268-7.
Rathkopf, Charles. 2015. "Network Representation and Complex Systems." *Synthese*, 1–24.
Silberstein, Michael, and Anthony Chemero. 2013. "Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences." *Philosophy of Science* 80 (5): 958–970.
Sporns, Olaf. 2011. *Networks of the Brain*. Cambridge, MA: MIT Press.
———. 2012. *Discovering the Human Connectome*. Cambridge, MA: MIT Press.
Stinson, Catherine. 2016. "Mechanisms in Psychology: Ripping Nature at Its Seams." *Synthese* 193 (5): 1585–1614. https://doi.org/10.1007/s11229-015-0871-5.
Towlson, E. K., P. E. Vertes, S. E. Ahnert, W. R. Schafer, and E. T. Bullmore. 2013. "The Rich Club of the C. Elegans Neuronal Connectome." *Journal of Neuroscience* 33 (15): 6380–87. https://doi.org/10.1523/JNEUROSCI.3784-12.2013.
Uhlhaas, Peter J., D. E. J. Linden, Wolf Singer, C. Haenschel, M. Lindner, K. Maurer, and E. Rodriguez. 2006. "Dysfunctional Long-Range Coordination of Neural Activity during Gestalt Perception in Schizophrenia." *Journal of Neuroscience* 26 (31): 8168–75. https://doi.org/10.1523/JNEUROSCI.2002-06.2006.
Uhlhaas, Peter J., and Wolf Singer. 2011. "The Development of Neural Synchrony and Large-Scale Cortical Networks During Adolescence: Relevance for the Pathophysiology of Schizophrenia and Neurodevelopmental Hypothesis." *Schizophrenia Bulletin* 37 (3): 514–23. https://doi.org/10.1093/schbul/sbr034.
Varshney, Lav R., Beth L. Chen, Eric Paniagua, David H. Hall, and Dmitri B. Chklovskii. 2011. "Structural Properties of the Caenorhabditis Elegans Neuronal Network." Edited by Olaf Sporns. *PLoS Computational Biology* 7 (2): e1001066. https://doi.org/10.1371/journal.pcbi.1001066.
Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393: 440–42.
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 London* 314: 1–340.
Wig, Gagan S., Bradley L. Schlaggar, and Steven E. Petersen. 2011. "Concepts and Principles in the Analysis of Brain Networks: Brain Networks." *Annals of the New York Academy of Sciences* 1224 (1): 126–46. https://doi.org/10.1111/j.1749-6632.2010.05947.x.
Woodward, James. 2013. "Mechanistic Explanation: Its Scope and Limits." *Proceedings of the Aristotelian Society Supplementary Volume* 87 (1): 39–65. https://doi.org/10.1111/j.1467-8349.2013.00219.x.
Zednik, Carlos. 2011. "The Nature of Dynamical Explanation." *Philosophy of Science* 78 (2): 238–263.
———. 2014. "Are Systems Neuroscience Explanations Mechanistic?" In *Preprint Volume for Philosophy Science Association 24th Biennial Meeting*, 954–75. Chicago, IL: Philosophy of Science Association.
———. 2015. "Heuristics, Descriptions, and the Scope of Mechanistic Explanation." In *Explanation in Biology*, 295–318. Springer.
———. 2018. "Computational Cognitive Neuroscience." In *The Routledge Handbook of the Computational Mind*, edited by M. Sprevak and Matteo Colombo. London: Routledge.