

One Mechanism, Many Models: A Distributed Theory of Mechanistic Explanation

Eric Hochstein

In recent years, there has been a surge in support for mechanistic theories of explanation within the philosophy of science, especially in regards to the life and cognitive sciences (see: Machamer, Darden & Craver 2000; Bechtel & Abrahamsen, 2005; Craver 2006, 2007; Piccinini 2006, 2015; Eliasmith 2010, 2013; Zednik 2011; Kaplan & Craver 2011; Milkowski 2011). Despite its many advocates however, no clear consensus exists regarding what sorts of scientific models provide mechanistic explanations, or even what exactly a model must do in order to provide such explanations. This lack of consensus has resulted in disagreements regarding whether or not we should interpret certain kinds of scientific models, such as computational models, dynamical models, and topological models, as providing mechanistic explanations (see, for example: Piccinini 2006, 2015; Rusanen & Lappi 2007; Eliasmith 2010; Huneman 2010; Milkowski 2011, 2013; Zednik 2011; Kaplan & Craver 2011; Jones 2014; Chirimuuta 2014; Ross 2015).

I propose that these disagreements stem in part from the fact that there are two distinct ways of understanding mechanistic explanations in the philosophy of science literature. Under one account, in order to provide a mechanistic explanation, a scientific model must represent the system as being a set of component parts organized such that their interactions bring about the explanandum phenomenon. I refer to this as the “representation-as” account of mechanistic explanation, since it requires that we interpret, or represent, the system *as being* a mechanism. In other word, we interpret the system as being a collection of component parts and operations that together produce a give phenomenon.

Under the second account, a model need not represent the target system *as* a mechanism in order to provide a mechanistic explanation, just so long as it provides essential information *about* the workings of some actual mechanism in the world. I refer to this as the “representation-of” account of

mechanistic explanation, since it requires only that the model in question provides us with information of the ontic mechanism we wish to study, and not that it interprets or represents the system mechanistically.

While both accounts of mechanistic explanation are frequently invoked as a means of making sense of neuroscientific practice, I argue in this paper that neither is in fact sufficient to account for the way in which mechanistic explanations are employed within neuroscience. If we adhere too closely to the first view, then we commit ourselves to the fact that we have virtually no mechanistic explanations in neuroscience. On the other hand, if we adhere too closely to the second view, then we have the opposite problem, and commit ourselves to the idea that virtually any and every model employed within neuroscience is trivially a mechanistic explanation. Both of these views fail to account for mechanistic explanatory practices in neuroscience.

The problem with both accounts is that each is concerned with what a scientific model must do in order to provide a mechanistic explanation. Yet this assumes that a single scientific model or representation is capable of instantiating such an explanation. In contrast, I argue that individual models rarely, if ever, provide mechanistic explanations. Instead, a mechanistic explanation is distributed across sets of scientific models, including some that abstract away from virtually all the structural details of the system, and others that idealize the system in extreme ways. Each of these models contributes limited, but essential, information to the same mechanistic explanation, but none can be considered a mechanistic explanation in isolation of the others.

In section 1 of the paper, I highlight how the two distinct ways of thinking about mechanistic explanations have manifested themselves in current debates within the philosophy of science. In section 2, I focus on the “representation-as” account, and demonstrate why it cannot adequately be used to characterize mechanistic explanations within neuroscience. In section 3, I focus on the “representation-of” account, and how it fails for different reasons. Finally, in section 4, I argue that we should think of

mechanistic explanations as collections of disparate models, and rarely as individual models. To claim that a particular model provides a mechanistic explanation is an elliptical way of saying that this model, *in conjunction with all other background information and pre-existing models of the system together*, provides a mechanistic explanation. This way of thinking about mechanistic explanation allows us to defuse many of the current disagreements within the philosophy of science regarding the mechanistic explanatory status of certain types of scientific models.

1. Mechanistic Explanation and Scientific Models

1.1 Do Computational, Dynamical, and Topological Models Provide Mechanistic Explanations?

To begin, consider an ongoing debate regarding the mechanistic status of computational models within the field of computational neuroscience. Many computational models characterize complex systems in terms of the sorts of computational functions they carry out, but do not identify the causal structure of the system that implements them. With this in mind, some have argued that computational models of this sort do not provide mechanistic explanations given that they are not intended to interpret or describe the system mechanistically; they represent neither the structure, nor the causal dependencies, of the systems under investigation (Sejnowski et al. 1988, p. 1300; Rusanen & Lappi 2007; Shagrir 2006, 2010; Chirumuuta 2014). To illustrate this point, Mazviita Chirumuuta (2014) offers the example of Heeger's contrast normalisation model of the primary visual cortex (Heeger 1992; Carandini & Heeger 1994; Carandini & Heeger 2012). She notes that Heeger considers the model to be a “canonical neural computation” (or “CNC”)¹ and not a mechanistic explanation. More specifically, she claims that:

¹ Carandini & Heeger (2012) define CNCs as “standard computational modules that apply the same fundamental operations in a variety of contexts” (p. 51).

The central point of Carandini and Heeger's (2012) account is that CNC's do not attempt to describe mechanisms—which vary significantly from one instance of a CNC to another, even when described in a very abstract manner—but rather a universal feature of the different systems to which the one CNC model, like normalisation, will apply. This feature is a computation. They write, for example, that “it is unlikely that a single mechanistic explanation [for normalization phenomena] will hold across all systems and species: what seems to be common is not necessarily the biophysical mechanism but rather the computation.” It is worth noting that the distinction between computation and mechanism is rather commonplace in neuroscience. (2014, p.141)

By focusing only on the information processing capacities of systems, these sorts of models make no attempt to describe the way in which these capacities are instantiated in any one system. In this respect, the variables of the computational model will often not correspond to the structural or causal features of any particular neurological system. In other words, it abstracts away from the sort of information that is thought to be required in order to generate a mechanistic explanation.

It is for this reason that Chirimuuta argues that computational models violate one of the central normative guidelines for mechanistic explanation, the “models-to-mechanism mapping” (or “3M”) requirement proposed by David Kaplan (2011; Kaplan and Craver 2011). Kaplan defines the 3M requirement in the following way:

(3M) In successful explanatory models in cognitive and systems neuroscience (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. (2011, p.347)

It is worth noting that the 3M requirement is intended by Kaplan to be what Michael Weisberg (2007) refers to as a “representational ideal”, one which is intended to “guide the direction of theoretical inquiry” and to “regulate which factors are to be included in models” (Kaplan 2011, p. 347). Chirimuuta suggests that the 3M requirement acts as a necessary condition on successful mechanistic explanations, and argues that many computational models (such as Heeger’s model, as well as the Gabor model of V1 receptive fields) violate this requirement. From this, she concludes that these models fail to provide mechanistic explanations (Chirimuuta 2014).

Before proceeding further, it should be noted that Chirimuuta’s goal is not merely to show that Heeger’s model fails to provide a mechanistic explanation, but also to argue that it provides a different *type* of explanation altogether (what she calls an “efficient coding explanation”). For the purposes of this paper, the question of what is required in order for a model to provide such non-mechanistic explanations is one that can be left aside. It may well be that computational models can provide both efficient coding explanations *and* mechanistic explanations, or that it only provides one and not the other, or that it fails to explain altogether. My primary interest here is to understand the conditions under which a scientific model can, and does, provide a mechanistic explanation in particular (irrespective of whether the model does or does not provide a different kind of explanation). And for Chirimuuta, computational models that do not identify the structural and causal features of the system, as is the case with Heeger’s model, do not meet the appropriate conditions needed for a *mechanistic* explanation.

In contrast, others have argued that such models in fact can, and do, provide mechanistic explanations, albeit in a limited or abstract form (e.g. Piccinini 2006, 2015; Piccinini & Craver 2011;

Milkowski 2011, 2013). It has been argued that even in contexts where such computational models fail the 3M requirement, they can still provide us with important insights about the mechanism under investigation, often by providing essential constraints on the sorts of structures and processes that are possible for implementing them. In so doing, they act as abstract or limited mechanistic explanations by provided information necessary for the discovery and understanding of such mechanisms. As Piccinini & Craver argue:

A final objection [against the position that computational models provide mechanistic explanations] might be that some computational models focus on the flow of information through a system rather than the mechanisms that process the information (cf. Shagrir 2006, 2010). In such cases, nothing is added to the explanation by fleshing out the details of how the information is represented and processed. Certainly, many computational explanations in psychology and neuroscience have this form. Our point, however, is that such descriptions of a system place direct constraints on any structures that can possibly process such information—on how the different states of the system can be constructed, combined, and manipulated—and are in turn constrained by the structures to be found in the system. It is, after all, an empirical matter whether the brain has structural components that satisfy a given informational description, that is, whether the neuronal structures in question can sustain the information processing that the model posits (under ecologically and physiologically relevant conditions). (2011, p.296)

Not any kind of neurological mechanism will be successfully characterized by a given computational model under existing environmental and physiological conditions. As a result, a computational model allows us to identify what sorts of neurological implementations are likely, and allows us to test

hypotheses about possible mechanisms by seeing whether a postulated mechanism conforms to the computational description provided under known constraints. In this respect, such models do play a role in mechanistically explaining the system.

The fact that such computational models do not directly identify the mechanistic details of the system being modeled just means that we ought to treat them as mechanism *sketches* or *schemas*; incomplete accounts of mechanisms with absent details (for more, see: Machamer, Darden, Craver 2000; Craver 2006; Piccinini & Craver 2011). Despite omitting the structural details of the system, such models are still considered to offer abstract mechanistic explanations in virtue of providing some degree of insight into what the structure and operations of the system are likely to be.

For those who adopt this view of mechanistic explanation, the 3M requirement acts more as a general guideline than a strict necessary condition on mechanistic explanation. This guideline tells us that, in general, the more our model has variables that map to the structure and causes of the system, the better a mechanistic explanation it becomes. However, this does not mean that any model which violates the 3M requirement can never carry explanatory information about the mechanism. They can often provide insight into how the mechanism must be structured, or how it must function, given other known conditions. Such models therefore still provide limited mechanistic explanations in virtue of providing constraints on possible mechanisms, and playing a role in the discovery of those mechanisms (Piccinini & Craver 2011; Zednik 2011; Craver 2014; Povich, forthcoming).²

Similar types of disagreements in the philosophy of science surround the mechanistic status of other types of scientific models. For instance, consider a parallel debate regarding the mechanistic status of dynamical models. Dynamical models employ sets of differential equations to characterize the way in which complex systems change over time, often modeling systems as vectors moving through a state space. These models frequently attempt to characterize the dynamics of complex systems without

² This is arguably still consistent with Kaplan's claim that the 3M requirement is intended as a tool to help "guide the direction of theoretical inquiry" (2011, p. 347)

consideration as to their underlying physical implementation (see: Thelen & Smith 1994; Van Gelder & Port 1995; Chemero & Silberstein 2008; Walmsley 2008). This has resulted in some claiming that such models fail to provide mechanistic explanations (Eliasmith 2010; Stepp, Chemero & Turvey 2011; Ross 2015). Despite this fact, such models can often play an important role in helping to discover the underlying mechanisms of a system, and are often the first step in fleshing out more detailed mechanistic accounts. Based on this, some have defended the claim that such models do provide abstract mechanistic explanations (Zednik 2011).

A similar debate surrounds the application of topological models in neuroscience. Topological models characterize complex systems by representing them in terms of relations between spatial entities, usually in the form of a graph, network, or state space. These spatial entities often do not correspond to the physical components of mechanisms, and are used instead to represent highly abstract, or mathematical, structures. Given this, many topological models can violate the 3M requirement. This has resulted in some claiming that such models do not provide mechanistic explanations (Huneman 2010; Jones 2014). Others, however, have retorted that such models can often play an important role in learning about, and understanding, the structure and operation of mechanistic systems. When they are used in this way, they do provide limited mechanistic explanations (Craver 2015, under review).

These debates all hinge on the same underlying issues: What must a scientific model do in order to provide a mechanistic explanation? Do computational models, dynamical models, and topological models represent systems in the appropriate sort of way to adequately instantiate a mechanistic explanation? I propose that there are in fact two different accounts of mechanistic explanation at work in the philosophy of science literature, and that these accounts differ in their answers to these questions.

1.2 Mechanistic Explanation: “Representation-as” versus “Representation-of”

Before proceeding further, it is worth noting that debates about mechanistic *explanations* should not be confused with debates about what mechanisms are more generally. Many theorists disagree about whether certain types of scientific models provide mechanistic explanations while still being largely in agreement as to what mechanisms are supposed to be (at least very broadly speaking). As Milkowski notes,

While mechanisms are defined variously, the core idea is that they are organized systems, comprising causally relevant component parts and operations (or activities) thereof. Parts of the mechanism interact and their orchestrated operation contributes to the capacity of the mechanism. (2013, p.3050)

This general account of a mechanism has become the standard way of thinking about mechanisms among philosophers of science over the past two decades (for just a small sampling, see: Machamer, Darden & Craver 2000; Piccinini 2006, 2015; Craver 2006, 2007; Rusanen & Lappi 2007; Bechtel 2008; Zednik 2011; Hochstein 2012,2013).

If we accept this general account, then it follows that a mechanistic explanation will be one that explains the occurrence of some phenomenon by showing how it is produced and sustained by the target mechanism. Such an explanation involves *decomposing* a complex system into parts and operations for better understanding. As William Bechtel notes, “a mechanistic explanation is, in an important sense, reductionistic” in so far as it “emphasizes the contributions made by parts of a mechanism to its operations” (2008, p.21). With this in mind, one common way of thinking about

mechanistic explanation is that a model provides such an explanation when it identifies four essential features of the mechanistic system:

- (1) The parts of the system.
- (2) The way in which these parts are spatially and temporally organized within the system.
- (3) The operations that go on between the relevant component parts.
- (4) The resulting phenomenon produced by the system.³

In addition, the characterization of these features cannot be too abstract or idealized. An explanation must be detailed enough to provide us with an understanding of how the mechanism is in fact put together, and how it works. Of course, the issue of what counts as “too abstract” or “too idealized” is a controversial one (see: Eliasmith & Trujillo 2014). However, a general guideline that is commonly adopted is that the model must provide enough detail to allow for intervention, manipulation, and control over the structural workings of the system so as to determine its counterfactual behaviours (see: Woodward 2002; Bechtel 2008; Eliasmith 2010; Raerinne 2011; MacLeod & Nersessian 2015). This means that a model fails to provide a mechanistic explanation if it is so abstract or idealized that it cannot identify the four different aspects of the mechanism, or cannot describe them in sufficient detail needed for manipulation, intervention, and control. As Chris Eliasmith notes:

In the case of cognitive and brain sciences, useful explanations are those that appeal to subpersonal mechanisms. This is because it is precisely such explanations which provide a basis for both intervention in behaviour and the artificial reproduction of those behaviours. These

³ For more on the idea that mechanistic explanation requires the identification of these four features, see: Machamer, Darden & Craver 2000; Bechtel & Abrahamsen 2005; Craver & Bechtel 2006; Bechtel 2008; Colombo, Hartmann, & van Iersel 2015.

mechanisms must be specific enough to allow for intervention. That is, the mechanisms must be specified in a way that relates to the measurable and manipulable properties of the system.

(2010, p.316)

To put this in slightly different terms, a scientific model provides a mechanistic explanation under this account when it describes a complex system *in terms of being* a collection of component parts organized so that their operations generate the target phenomenon. While other sorts of scientific models may tell us important information about the mechanism, a mechanistic explanation involves interpreting and analyzing the system *as* a mechanism (i.e. understanding what the parts and operations of the system are, and how they are organized so as to produce the phenomenon). With this in mind, I will refer to this as the “representation-as” (hereafter “R-A”) view of mechanistic explanation.

Note that according to this view of mechanistic explanation, a model might satisfy the 3M requirement, and yet still *fail* to provide a mechanistic explanation. This is because the model may have variables that map on to components or operations of a mechanism, and yet does not characterize these features in ways that allow for intervention in their workings.⁴

Returning for a moment to the debates surrounding computational, dynamical, and topological models, it becomes apparent why such models would fail to provide mechanistic explanations under the R-A account. If these models do not identify the four aspects of a mechanism in sufficient detail for manipulation and control, then they do not provide the required information for a mechanistic explanation. They might provide information needed for a different kind of explanation, or they might fail to provide explanations altogether, but they do not interpret the system in an appropriate manner to be a *mechanistic* explanation. Chirimuuta’s claim that “there can be principled reasons for analyzing neural systems computationally rather than mechanistically” (2014, p.139) suggests that she has

⁴ Eliasmith, for instances, explicitly argues that most kinds of statistical, dynamical, and computational models fail to provide mechanistic explanations on these grounds (2010).

something like the R-A account in mind; such models are not mechanistic explanations because they do not analyze systems *as* mechanisms, even if they happen to be descriptions *of* mechanisms.

This view of mechanistic explanation can be contrasted with another that has intuitive appeal for very different reasons. Consider that mechanisms are frequently thought to exist out in the world, independent of how we interpret them. Bechtel & Abrahamsen, for instance, claim “mechanisms are real systems in nature” (2005, p.424-425). Likewise, Craver argues that mechanistic explanations “describe real components, activities, and organizational features of the mechanism that in fact produces the phenomenon” (Craver 2006, p.361). Many neuroscientists consider the human brain to be a mechanistic system independent of whether they use a mechanistic model to represent it that way or not.⁵

With this in mind, there is a distinct sense of mechanistic explanation that is also found in the literature: a model provides a mechanistic explanation when it is used to help us learn about, and understand, how some actual mechanism in the world operates, or how it is structured. According to this view, mechanistic explanations need not represent the system *in terms of being* a set of component parts and operations in order for it to provide information needed in the study and understanding *of* those parts and operations. Models can provide abstract or limited mechanistic explanations by providing information needed to test hypotheses about possible mechanisms, by providing constraints on the sorts of mechanisms at work in a given system, and by helping us to discover the causal workings of the system (see: Zednik 2011; Piccinini & Craver 2011; Craver 2014; Povich, forthcoming). Mechanistic explanation is not tied to the way in which we represent or analyze the system, but to whether our model provides important information *about* the mechanism in the world we wish to study.

⁵ Although it is worth noting that this idea is not uncontroversial. There are some who deny that we should always think of mechanisms as out in the world in this manner (e.g. Colombo, Hartmann & van Iersel 2015; Bechtel, forthcoming).

Consider the practice of task-analysis in cognitive science. Task analysis involves understanding the capacity of a system by decomposing it into sub-capacities with the eventual goal of mapping these sub-capacities to structures and operations of the brain. Task analysis is taken by some to be a form of abstract mechanistic explanation in virtue of the fact that by identifying sub-capacities, these models allow us to identify constraints on the sorts of structures and processes that can carry them out. In doing so, they play a direct role in the discovery and understanding of causal mechanisms. As Piccinini & Craver argue,

A task analysis is a mechanism sketch in which the capacity to be explained is articulated into sub-capacities, and most of the information about components is omitted. Nevertheless, the sub-capacities do place direct constraints on which components can engage in those capacities. For each sub-capacity, we expect a structure or configuration of structures that has that capacity. This guiding image underlies the very idea that these are explanations, that they reveal the causal structure of the system. (2011, p.294).

Note that the mechanistic status of such models does not stem from the fact that they describe the parts and operations of the system, but from the fact that they help us to *reveal* the causal structure of the system, often by identifying *constraints* on what sorts of mechanisms are possible. More importantly, Piccinini & Craver argue that this can be the case even in situations where the task analysis *does not allow us* to map sub-capacities onto the different parts and operations of the mechanism (thus violating the 3M requirement). As a concrete example of this, consider the problem of task-analysis when dealing with general purpose computers:

General purpose computers can do an indefinite number of things depending on how they are programmed. Thus, one might think, a task analysis of a general purpose computer's capacities places no direct constraints on its structural components and the structural components place no direct constraints on task analysis, because the components are the same regardless of which capacity a computer exhibits. The only constraint is indirect: the computer must be general purpose, so it must have general purpose components. By extension, if the same kind of autonomous task analysis applies to human behavior, then the type of task analysis with which the received view is concerned is not a mechanism sketch.

This objection makes an important point but draws the wrong conclusion. True, general purpose computers are different from most other systems precisely because they can do so many things depending on how they are programmed. But general purpose computers are still mechanisms, and the explanation of their behavior is still mechanistic (Piccinini 2007, 2008). Furthermore, the task analysis of a general purpose computer does place direct constraints on its mechanistic explanation and vice versa; in fact, even the task analysis of a general purpose computer is just an elliptical mechanistic explanation. [...] Whether human brains contain this kind of mechanism is an empirical question, and it can only be resolved by investigating whether brains have this kind of organization. (2011, p.294)

Even in these cases, the task analysis model can still provide constraints on the sorts of mechanisms likely to be involved, and thus can contribute to the discovery and understanding of those mechanisms. The model's status as a mechanistic explanation according to Piccinini & Craver is thus not tied to how it analyzes the system, but to whether it provides information *about* the mechanisms that we can use to better understand its workings.

To further elaborate on this idea, consider Mark Povich's (forthcoming) argument that some cognitive models (such as the JIM, SUSTAIN, and ALCOVE models) count as providing limited mechanistic explanations even if they have variables which fail to map to any actual structures or causes in the brain. He notes that while these variables in isolation do not provide any explanatory information about the structure or causes of the mechanism, the model as a whole carries mechanistic explanatory content in virtue of being used as a means of "suggesting, constraining, and sharpening questions about mechanisms" (p.8), and because they are used to "suggest new lines of investigation" about causal mechanisms (p.10). He concludes from this that we ought to interpret such models as mechanism sketches. Likewise, Carlos Zednik argues that the application of Artificial Neural Networks (ANNs) in cognitive science ought to be interpreted as abstract mechanistic explanations, even if the nodes and connections of the model cannot be mapped to any particular neural structures or organizations of the brain. He argues that ANNs still provide abstract mechanistic explanations in virtue of the fact that such models "capture basic neurobiological principles" (such as Hebbian learning and spreading activation) that can be used to "reveal the mechanism responsible for that phenomenon" (2011, p.241).

With this in mind, I will refer to this alternative way of thinking about mechanistic explanations as the "representation-of" (or "R-O") view of mechanistic explanation, since it requires only that the model in question provides us with explanatory information *of* the ontic mechanism we wish to study, and not that it interprets or analyzes the system mechanistically. While the R-A view ties mechanistic explanation to way in which we represent or analyze a system, the R-O view ties mechanistic explanation to whether our models tell us important information *about* a mechanistic system.

If we return once again to the debates surrounding computational, dynamical, and topological models, we can see a very different answer emerging from the R-O view as to whether such models provide mechanistic explanations. Even if such models do not interpret systems as mechanisms, they can still frequently be used to identify essential constraints on the structure and function of mechanistic

systems, and are often used as a first step in developing more detailed mechanistic accounts. In this respect, they provide limited mechanistic explanations.

So which account of mechanistic explanation should we adopt? Both attempt to justify their position by appealing to the actual practices and commitments of working neuroscientists and biologists. Yet, I propose that if we use neuroscientific practice as our guide, then neither account provides a helpful account of mechanistic explanation in neuroscience. In order to make this point, I will demonstrate in the following sections why both the R-A account and the R-O account run into problems when accounting for neuroscientific practice.

2. The Problems with the “Representation-as” Account

Recall that according to the R-A view, a scientific model provides a mechanistic explanation on the condition that it describes the parts, operations, organization, and resulting phenomenon of a mechanism in sufficient detail for manipulation and control over its structural workings. The problem with the R-A account is that this is often not possible within one single model.

To illustrate how this problem has manifested itself within neuroscientific practice, consider the various models of the action potential that have been developed throughout the history of neuroscience and biology. During the 20th century, various different modeling techniques were used to generate different kinds of models of the action potential, yet each was only partial in its characterization of the underlying mechanism, with no single model able to represent the various aspects of it simultaneously. The Hodgkin & Huxley model (1952) was an early attempt to mathematically characterize electrical features of the action potential (specifically the ion flow of sodium and potassium channels), but was unable to identify the underlying physiological structures and operations that produced them. While Hodgkin and Huxley made conjectures as to the possible underlying causal structures involved, the

model they provided simply lacked the relevant information by itself to confirm any sort of structural story. This was complicated further by the fact that they lacked the ability to measure sodium and potassium conductance in a direct manner.

This changed in the 1970s with the development of single-channel recording. This new electrical recording technique allowed for far more accurate and fine-grained representations of the sodium and potassium channels. Despite such advancements however, the models of the action potential that were developed using this new technique still focused primarily on representing electrical features of the action potential without characterizing the system's underlying causal structure (as an example, see: Aldrich, Corey, & Stevens 1984).

To compensate, other sorts of scientific models from physiology and protein chemistry were used to characterize other aspects of the mechanism. One such model was the Fluid Mosaic Model of cell membranes developed by Singer & Nicholson in 1972. This model focused on characterizing some of the structural aspects of the mechanism responsible for the action potential by representing the general organization and structure of proteins and lipids in the cell's membrane (Singer & Nicholson 1972, p. 720). By focusing on these structural aspects of the system however, the model was unable to represent the more dynamic features and operations of the action potential that electrical models were needed to capture. Biologists and neuroscientists were thus left with a collection of different models from different domains, each providing different kinds of information about the different aspects of the mechanism, but without a single model able to represent them all. As Hans Meves noted in regards to the models being created based on single-channel recording:

Single channel measurements like those reported by Aldrich and colleagues give us a much deeper insight into the function of the sodium channel than previously possible, but undoubtedly many problems still remain to be solved. The essential problem is to combine the

information obtained by the different techniques (single channel recording, total current measurements, fluctuation analysis) into a coherent picture of the sodium channel. (1984, p.425)

With the invention of computer modeling techniques, a new set of tools brought renewed promise for integrating different representations into a single model that could represent the parts, organization, operations, and resulting phenomena of the system in sufficient detail for intervention and control. Yet such attempts met with limited success. Integrating many different representations of a given mechanism into a single computer model still required that some essential features of the mechanism had to be left out in order for the model to successfully represent others.

To understand why this problem persists, consider the sort of issues facing any model that attempts to provide a mechanistic explanation of complex neurobiological or cognitive mechanisms. It turns out that representing the organization of neurological mechanisms often requires models which deliberately *exclude* details about the structure and interaction of its component parts, making it impossible to use a single model to represent both features at the same time. Arnon Levy and William Bechtel, for example, argue that “to understand organization, one often needs to abstract from the structural specifics of a mechanism and represent it in a skeletal, coarse-grained manner” (2011, p.241). They argue that certain kinds of mechanistic organizations (such as complex feedback loops) require the use of large scale graph-based modeling which, by necessity, abstracts away from the structural and causal features of the mechanism (ibid, p. 246).

In addition to abstractions of this sort, incompatible sets of idealizations are often required to represent different features of the same causal mechanism. These idealizations can range from the use of “point neurons” (simulated neurons with no spatial properties), to the positing of infinite neural populations. Different neuroscientific models will frequently invoke conflicting sets of idealizing

assumptions depending on which causal or structural features of the mechanism they wish to identify and study. This practice, which Weisberg refers to as “Multiple-Models-Idealization”, is commonplace when representing complex mechanisms:

Multiple-models idealization (hereafter MMI) is the practice of building multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to the phenomenon. [...] One most commonly encounters MMI in sciences dealing with highly complex phenomena. (2013, p.103)

In these situations, there is no one single model that can be used to provide a mechanistic explanation, since different models must be employed which adopt conflicting idealizations in order to represent different features of that mechanism needed for the explanation.

Yet another complication is that the way in which a complex mechanism behaves is heavily dependent on context, and understanding how the parts of a system interact to produce the phenomenon often must include relevant details of the surrounding environment. Yet, any individual model of a mechanism will only be informative as to its behaviour under those specific conditions. Thus we often must rely on different models which characterize the mechanism under varying conditions in order to gain a sufficient understanding of its workings (see: Bechtel 2008, p. 22; Bechtel 2009).

But if a single model in neuroscience is rarely, if ever, able to simultaneously represent all four aspects of a neurophysiological mechanism needed for an explanation, or represent them in sufficient detail for intervention and control, then the R-A view of mechanistic explanation becomes much too restrictive. By insisting that a mechanistic explanation must be provided by a *single* scientific model, it sets up standards of explanation that are frequently impossible to meet, and means that the majority of neurological and cognitive phenomena that neuroscientists study cannot be explained mechanistically.

Yet if this is the case, then this cannot be the sense of mechanistic explanation that philosophers of science invoke when they claim that neuroscientists primarily explain phenomena by way of a model of the mechanism.

3. The Problems with the “Representation-Of” Account

The R-O account of mechanistic explanation manages to avoid the overly-restrictive demands on mechanistic explanation that plague the R-A account, and so initially appears to better respect the pragmatics of scientific practice. Under the R-O account, no single model needs to identify all four features of a mechanism in sufficient detail for control and manipulation in order to provide a mechanistic explanation. Instead, it only requires that the model provides us with information *of* or *about* the target mechanism that we can use to better understand it.

The problem with the R-O account is that it overcompensates for the R-A account. It is too *permissive* instead of being too restrictive. Under the R-O account, virtually any and every model that neuroscientists use in the study of the brain will trivially be an abstract mechanistic explanation. In order to see why, consider the arguments for why some computational, dynamical, and topological models are thought to provide abstract mechanistic explanations, even in contexts where they may violate the 3M requirement. According to Piccinini & Craver (2011), such models provide mechanistic explanations because they “place direct constraints on any structures that can possibly process such information—on how the different states of the system can be constructed, combined, and manipulated—and are in turn constrained by the structures to be found in the system” (p.296). This idea fits with Povich’s argument that various biologically implausible cognitive models can still provide limited mechanistic explanations in virtue of “constraining and sharpening questions about mechanisms” (forthcoming, p.8), and Zednik’s

claim that such models provide abstract mechanistic explanations because they help us to “reveal the mechanism responsible for that phenomenon” (2011, p.241).

While it is indeed true that these sorts of models can provide such constraints, and thus help us discover and understand the underlying mechanisms of the system, it is important to note this will also be true of virtually every model that characterizes the input/output relations of the system. As long as we know the environmental conditions under which a behaviour takes place, and the time it takes for the behaviour to be carried out, then *any* account of its inputs and outputs will provide direct constraints on the sorts of causal architectures that are possible.

The reason for this is that changes to the structure of a mechanism will always translate into performance differences in its overall behaviour. These changes will affect how the system performs, under what conditions, and the time frame in which it carries out behaviours (see: Le Cun and Denker 1992; Keely 2000; Eliasmith 2002, 2013; Shapiro 2004; Syropoulos 2008, p.111; Craver 2009, p.586). This means that *any* model that characterizes the inputs and outputs of the system, when combined with information regarding environmental conditions and time constraints, will provide information as to the causal structure of the system, since only very particular causal structures and processes will be capable of carrying out those behaviours under those very particular conditions and temporal constraints (for further discussion, see: Eliasmith 2002). In this respect, any model of the brain that characterizes it in terms of inputs and outputs will trivially count as an abstract mechanistic explanation under the R-O account.

This charge of triviality is made even worse by the fact that even models which are *not* predictively adequate, or which mischaracterize the structure of the system entirely, can still provide constraints on possible mechanisms in this same way. After all, such accounts can tell us something important about the causal structure of the system by telling us conditions under which the phenomenon is *not* generated, by narrowing down lists of possible implementations by eliminating

those which fail to fit with the occurrence of the phenomenon, and by adding further constraints on our understanding of the actual mechanisms and the principles they obey. As such, it now appears that any and all models of the brain, regardless of whether they are predictive or not, whether they correctly describe the structure of the system or not, or whether they decompose the system or not, are all abstract mechanistic explanations in virtue of providing non-trivial information that can be used in the discovery and identification of mechanisms. But if every model of the system is a mechanistic explanation in some limited fashion, then the claim that neuroscientists provide mechanistic explanations amounts to nothing more than the claim that neuroscientists study the brain, and use models to do so. Yet no one would ever deny such a claim. Mechanistic explanation becomes trivial and uninformative under this account.

4. A Distributed Account of Mechanistic Explanation

It seems that our intuitions pull us in two directions. On the one hand, the R-A view avoids the triviality of the R-O view by insisting that an explanation of a mechanism involves more than just providing informative *about* the mechanism. It must provide information sufficient to understand what the mechanism is, and how it operates to produce the phenomenon. On the other hand, the R-O view avoids the impossibly high standards of the R-A view, and correctly notes that as a matter of descriptive fact, neuroscientists often do not describe all four essential features of a mechanism within a given model when they consider themselves to be engaging in mechanistic explanation. So how then do we proceed?

The real confusion in this debate centers on the assumption that a mechanistic explanation can be adequately conveyed by any single scientific model or representation. The debate about what a model must do in order to provide a mechanistic explanation assumes that such an explanation can be

adequately captured by such a model. This is an idea that we must abandon. While it is undeniably *through the use of models* that we develop mechanistic explanations, we need to keep in mind that an individual model is rarely applied in isolation, and is often used to complement a huge body of background information and pre-existing models about the target system. As Colombo, Hartmann, & van Iersel note,

This modelling process relies on relevant background knowledge, well-established empirical results about the target mechanism, inferential methods connecting background knowledge to such results, and on practical ends and epistemic interests. (2015, p.196)

With this in mind, I propose that we should think of mechanistic explanations as distributed across *sets* of scientific models, with each model in the set contributing a piece to the same overall explanatory whole. Some of these models will be extremely abstract in their characterization of the mechanism, while others will invoke numerous idealizing assumptions. Each provides insights into the mechanism being studied, and in doing so becomes part of one and the same overarching distributed explanation.

To illustrate, let us return to our example of the action potential. Instead of appealing to any one model of the action potential as a mechanistic explanation, scientists have historically used a collection of different models and representations to instantiate such an explanation.⁶ Many physiology textbooks, for instance, have used different kinds of representations of the action potential side-by-side in order to provide a more comprehensive account of the mechanism, instead of attempting to integrate them into a single representation or model (e.g. Ruch & Fulton 1960, p.49; Hille 1984, p.15; Kandel, Schwartz & Jessell 1991, p.70. See Trumpler 1997 for discussion). Even the more integrative models of

⁶ Another historical example is the study of the mechanism for protein synthesis. The mechanism for protein synthesis was not explained by any single representation, but by the piecemeal collection of various different models generated within different fields (Machamer, Darden & Craver 2000, p.18-21).

the action potential always implicitly invoke other models and theories as background assumptions, and frequently use them as a means of filling in details of the mechanism that they leave out. Colombo, Hartmann, & van Iersel, for example, argue that:

The parts comprised by the mechanism of the action potential will always be defined against a background of relevant accepted knowledge with which they must cohere, as well as of relevant evidence, with which they must fit. This background knowledge will be reflected in various modelling rules, which should be taken into consideration both in the definition of mechanistic parts and in the assessment of the resulting mechanistic model. (2015, pp.198-199)

Some, like Maria Trumpler, have concluded from this that a truly integrative account of the mechanism of the action potential exists only in the minds of working scientists (1997). I propose that this idea is not quite correct. Instead, I suggest that scientists *do not need* an integrated model of the mechanism in order to provide a mechanistic explanation. Instead, as long as our collection of models together allows us to identify the relevant features of the mechanism needed to explain it, then we can appeal to individual models within the collection as the need arises to gain an understanding of some restricted aspect of it that we wish to focus on.

An interesting consequence of this account is that mechanistic explanations will often be fractured or fragmented in nature. In other words, the models that constitute an explanation will sometimes partially overlap with one another, and sometimes conflict with one another (in virtue of adopting incompatible sets of idealizing assumptions). In this respect, these models should not be thought of as seamlessly fitting together like puzzle pieces combining to form a unified picture. Instead, we come to conclusions about the unified mechanism in the world by drawing relevant inferences from the different and sometimes contradictory models in our collection used to represent the different

facets of the system. We move between the different models in our collection as the need arises, drawing information from each when appropriate (see, for example: Longino 2006, 2013; Potochnik, draft manuscript).

This account of mechanistic explanation allows us to salvage the virtues of both the R-A and R-O accounts of mechanistic explanation, while largely avoiding their pitfalls. The R-A account claims that a mechanistic explanation must identify the four features of the mechanism in sufficient detail for manipulation, intervention, and control. The account I propose likewise adopts this condition on mechanistic explanation, and thus avoids the triviality of the R-O account. Not any model, or even any collection of models, will be sufficient to constitute a mechanistic explanation unless it can provide this information about the mechanism. Where the R-A account goes wrong is its insistence that all of this must be captured within the same model. By adopting a distributed account, we can avoid this by noting that this information is provided by the multitude of different models at our disposal, even if we cannot integrate them into a single account. It is the *collection* of models that provide the mechanistic explanation together, and not any individual model in isolation.

When neuroscientists refer to a scientific model as a mechanistic explanation, this is an elliptical way of saying that the model, when embedded within a collection of pre-existing models about the system, provides a mechanistic explanation. Therefore, whether we treat a model as a mechanistic explanation or not can often depend entirely on what other background information, and pre-existing models of the system, we have available.

One potential worry with this account is that we need some way to determine which sorts of models we are permitted to include in our distributed set, and which we are not. For instance, we may have many different conflicting models from different research traditions at our disposal, each of which may be informative of the mechanism in some ways, but may have serious deficits in others. How do we determine which of these models to include? To make this problem more concrete: imagine some

model M is to be assessed as a mechanistic explanation based on the collection of background models we bring to bear on its interpretation. How can we tell which models should go into this background collection? Suppose, for instance, some of these models (from one research tradition) are predictive of certain behaviours in certain contexts, but are biologically implausible. Others (from different research traditions) are biologically plausible in different ways, but fail to predict. The researchers behind M do not specify their own research tradition at all. How should we then assess M's status as a mechanistic explanation? It seems that one could arbitrarily pick some set of models as the background, but they will all clearly be deficient in different ways. How then do we proceed?⁷

The solution is to realize that we do not pick our background models arbitrarily, or restrict ourselves to a single research tradition, but justify the inclusion of models into our collection based on their ability to meet the constraints of other models that we know empirically can successfully represent different restricted facets of the system. For instance, despite being biologically implausible, the predictive models from one research tradition will provide behavioural data that the more biologically plausible models from the other research tradition will have to conform to. Not any account of the structure of the system will be capable of producing the behavioural regularities identified by the predictive model given known environmental conditions, and this gives us a means of deciding between different biologically plausible models. Likewise, knowing more details about the structures and causes of the system from our biologically plausible models will allow us to better refine and alter the predictive models from the other research tradition by providing new boundary conditions based on the known structural limitations of the system.

Recall the earlier point that the way in which a mechanism is structured effects the sorts of behaviours it can carry out, the time they can be carried out in, and the contexts in which they can occur. Different types of models will provide different information about these features, and so can be

⁷ I would like to offer special thanks to a blind referee for highlighting this worry.

used to inform one another. Statistical models, dynamical models, topological models, control theoretic models, computational models, and others all provide different constraints on what the structure and function of the mechanism is by characterizing different causal, organizational, or behavioural properties that will allow us to rule out some models, and refine others. The more constraints we identify, the more we can narrow the list of models suitable for inclusion into our distributed explanatory set.

To illustrate, consider once again the different scientific models of the action potential. The Hodgkin & Huxley model developed in the 1950s was able to accurately represent some aspects of the mechanism under investigation, but not others. In particular, it correctly noted that sodium and potassium conductances had specific voltage and time dependencies. Once this was discovered, these dependencies became one of the primary constraints on developing new scientific models of the action potential after that point (Trumpler 1997, p.63). Models which attempted to represent other features of the mechanism needed to provide data that fit with the electrical constraints identified by the Hodgkin & Huxley model.

This fact, however, did not mean that the Hodgkin & Huxley model itself was immune from revision. With the invention of single channel recording, far more detailed accounts of these dependencies were made possible. This, in addition to new insights provided by other models from other domains regarding structural features of the system (such as the Fluid Mosaic Model), allowed scientists to produce electrical representations which were much more consistent with the biological constraints identified by those models. This resulted in the replacement of the Hodgkin & Huxley model with these more refined representations (Trumpler 1997, p.68). Thus the collection of models used to constitute the mechanistic explanation of the action potential were not taken at random, but were chosen based on their ability to conform to the parameters and boundary conditions identified by other models from different domains and research traditions.

This means that the process of mechanistic explanation will often be a long drawn out process where models are continually altered and refined over years before we can develop an appropriate distributed collection of models needed for the explanation. This is why mechanistic explanations are often only ever clear in hindsight, after decades of refinements and interactions between models of different types.

This framework for selecting models can be further supplemented with a criterion of non-redundancy. Note that with the invention of single channel recording, we were able to create models that captured the dependencies of the Hodgkin & Huxley model, while also refining and improving upon them. As a result, these newer models “marked the replacement of the series of Hodgkin-Huxley conductance curves with the single-channel trace as the established representation of sodium conductance” (Trumpler 1997, p.68). While the Hodgkin & Huxley model remains important as a historical example, it is no longer included within the set of models that constitute our best mechanistic explanation of the action potential, since the content it provides is redundant with the information provided by the more accurate models generated through single channel recording. In this respect, having a mechanistic explanation that is composed of hundreds of models that provide virtually the same information is unhelpful. Thus, a useful guide is that the models in our collection must provide at least *some* information about the parts, operations, organization, and resulting phenomenon of the mechanism that is not already provided by other models in the set (i.e. they must provide information that is *relevant and non-redundant*).

Conclusion

In this paper, I have argued that we should not think of mechanistic explanations as being captured by any one model or representation. Mechanistic explanations span *sets* of different models which all

contribute to the same overall explanation. These models can often invoke different idealizing assumptions, and describe the system at different grains of abstraction.

With this in mind, let us revisit one last time the ongoing debates about whether computational, topological, and dynamical models provide mechanistic explanations. Those who argue that they do not, do so on the grounds that such models do not identify the relevant structural and causal features of a mechanism needed to be a mechanistic explanation. And indeed, if we consider such models in isolation, *this is true*. A computational model by itself does not provide the appropriate information needed for a mechanistic explanation. On the other hand, if we think of the computational model as *adding* to our pre-existing set of models about the system, then it can become an essential part of our mechanistic explanation (see: Piccinini 2015; Eliasmith & Trujillo 2014, p.4). Even though the computational model by itself does not describe the essential parts and operations of the system, it can *defer* to other pre-existing models for this information, and use them as constraints on its own account of the mechanism. When this happens, the different models inform one another, and contribute to the same overall mechanistic explanation.

When scientists claim that a model is a mechanistic explanation, they are implicitly embedding the model within a collection of pre-existing models, and judge that this collection provides the mechanistic explanation they seek. Thus, under this account, we can claim both that a model *is* and *is not* a mechanistic explanation depending on what sort of additional background information we currently have available, and whether we bring this information to bear on our interpretation of the model. Once we recognize this, many of these disputes within the philosophy of science dissolve.

References:

Aldrich, R., Corey, D., and Stevens, C. (1984) A Reinterpretation of Mammalian Sodium Channel Gating Based on Single Channel Recording. *Nature* 308: 436–441.

- Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Lawrence Erlbaum Associates.
- Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology* 22 (5): 543-564.
- Bechtel, W. (forthcoming). Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A Mechanistic Alternative. *Studies in the History and Philosophy of Biomedical Sciences* 36: 421-441.
- Carandini, M., and Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264, 1333–1336.
- Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience* 13: 51–62.
- Chemero, A., and Silberstein, M. (2008). After the Philosophy of Mind: Replacing Scholasticism with Science. *Philosophy of Science* 75:1–27.
- Chirimuuta, M. (2014). Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience. *Synthese* 191(2): 127-154.
- Colombo, M., Hartmann, S., and van Iersel, R. (2015). Models, Mechanisms, and Coherence. *The British Journal for Philosophy of Science* 66: 181-212.
- Craver, C. (2006). When Mechanistic Models Explain. *Synthese* 153 (3): 355-376.
- Craver, C. (2007). *Explaining the Brain*. Oxford: Oxford University Press.
- Craver, C. (2009). Mechanisms and Natural Kinds. *Philosophical Psychology* 22 (5): 575-594.
- Craver, C. (2014). The Ontic Account of Scientific Explanation. In M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*. Synthese Library. 27-52.
- Craver, C. (2015) Graphing the Brain's Dark Energy: Network Analysis in the Search for Neural Mechanisms. Paper presented at annual meeting of the *International Society for the History, Philosophy, and Social Studies in Biology*. Montreal, QC, Canada.
- Craver, C. (Under Review). Network Models and Explanations.
- Craver, C., & Bechtel, W. (2006). Mechanism. In S. Sarkar and J. Pfeifer (Eds.), *Philosophy of Science: An Encyclopedia*. New York: Routledge.
- Eliasmith, C. (2002). The Myth of the Turing Machine: The Failing of Functionalism and Related Theses. *Journal of Experimental & Theoretical Artificial Intelligence* 14 (1): 1-8.

- Eliasmith, C. (2010). How we ought to describe computation in the brain. *Studies in History and Philosophy of Science Part A*, 41: 313-320.
- Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.
- Eliasmith, C. and Trujillo, O. (2014). The use and abuse of large-scale brain models. *Current Opinion in Neurobiology* 25:1–6.
- Heeger, D. J. (1992). Normalization of cell responses in the cat striate cortex. *Visual Neuroscience* 9: 181–197.
- Hille, B. (1984). *Ionic Channels of Excitable Membranes*. Sutherland, Mass: Sinauer.
- Hochstein, E. (2012). Minds, Models, and Mechanisms: A New Perspective on Intentional Psychology. *Journal of Experimental & Theoretical Artificial Intelligence* 24 (4): 547-557.
- Hochstein, E. (2013). Intentional Models as Essential Scientific Tools. *International Studies in the Philosophy of Science* 27 (2): 199-217.
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology* 117: 500–44.
- Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese* 177 (2): 213-245.
- Jones, N. (2014). Bowtie Structures, Pathway Diagrams, and Topological Explanation. *Erkenntnis* 79 (5):1135-1155.
- Kandel, E.R., Schwartz, J.H., and Jessell T.M. (1991). *Principles of Neural Science*, 3rd Edition. New York: Elsevier.
- Kaplan, D.M. (2011). Explanation and description in computational neuroscience. *Synthese* 183: 339–373.
- Kaplan, D.M., & Craver, C.F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science* 78: 601-627.
- Keeley, B. (2000). Shocking Lessons from Electric Fish: The Theory and Practice of Multiple Realizability. *Philosophy of Science* 67: 444-465.
- Le Cun, Y., and Denker J. S., (1992). Natural versus universal probability, complexity, and entropy. *IEEE Workshop on the Physics of Computation*.
- Levy, A. and Bechtel, W. (2013). Abstraction and the Organization of Mechanisms. *Philosophy of Science* 80 (2): 241-261.

- Longino, H. (2006). Theoretical Pluralism and the Scientific Study of Behavior. In S. Kellert, H. Longino, and C.K. Waters (Eds.), *Scientific Pluralism*. Minneapolis: University of Minnesota Press. 102-132.
- Longino, H. (2013). *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. Chicago: University of Chicago Press.
- Machamer, P., Darden, L. & Craver, C. (2000). Thinking About Mechanisms. *Philosophy of Science* 67 (1): 1-25.
- MacLeod M. & Nersessian N.J. (2015). Modeling systems-level dynamics: Understanding without mechanistic explanation in integrative systems biology. *Stud Hist Philos Biol Biomed Sci*. 49: 1-11.
- Meves, H. (1983). A New View of the Sodium Channel. In "News and Views," *Nature* 306: 425.
- Milkowski, M. (2011). Beyond Formal Structure: A Mechanistic Perspective on Computation and Implementation. *Journal of Cognitive Science* 12 (4): 359-379.
- Milkowski, M. (2013). A Mechanistic Account of Computational Explanation in Cognitive Science. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Piccinini, G. (2006). Computational Explanation and Mechanistic Explanation of Mind. In M. DeCaro, F. Ferretti & M. Marraffa (Eds.), *Cartographies of the Mind: The Interface Between Philosophy and Cognitive Science*. Dordrecht: Kluwer.
- Piccinini, G. (2015). *Physical Computation: A Mechanist Account*. United Kingdom: Oxford University Press.
- Piccinini, G. and Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese* 183: 283-311.
- Potochnik, A. (unpublished). Idealization and the Aims of Science.
- Povich, M. (forthcoming). Mechanisms and Model-based fMRI. *Philosophy of Science*.
- Raerinne, J. (2011). Causal and Mechanistic Explanations in Ecology. *Acta Biotheoretica* 59: 251-271.
- Ross, L. (2015). Dynamical Models and Explanation in Neuroscience. *Philosophy of Science* 82 (1): 32-54.
- Ruch, T. and Fulton, J. (1960). *Medical Physiology and Biophysics*, 18th Edition. Philadelphia: Saunders.
- Rusanen, A. M., & Lappi, O. (2007). The limits of mechanistic explanation in neurocognitive sciences. In *Proceedings of the European Cognitive Science Conference*.
- Sejnowski, T. J., Churchland, P. S., & Koch, C. (1988). Computational neuroscience. *Science* 241: 1299–1306.
- Shapiro, L. A. (2004). *The Mind Incarnate*. Cambridge, MA, MIT Press.

- Shagrir, O. (2006). Why we view the brain as a computer. *Synthese* 153(3): 393–416.
- Shagrir, O. (2010). Brains as analog-model computers. *Studies in the History and Philosophy of Science* 41(3): 271–279.
- Singer, S.J., and Nicolson, G. L. (1972). The Fluid Mosaic Model of the Structure of Cell Membranes. *Science* 175 (4023): 720-731.
- Stepp, N., Chemero, A., & Turvey, M. (2011). Philosophy for the Rest of Cognitive Science. *Topics in Cognitive Science* 3 (2): 425-437.
- Syropoulos, A. (2008). *Hypercomputation: Computing Beyond the Church-Turing Barrier*. Springer.
- Thelen, E., and Smith, B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.
- Trumpler, M. (1997). Techniques of Intervention and Forms of Representation of Sodium-Channel Proteins in Nerve Cell Membranes. *Journal of History of Biology* 30 (1): 55-89.
- Van Gelder, T., and Port, R. (1995). It's About Time: An Overview of the Dynamical Approach to Cognition. In R. Port and T. Van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press. 1-43.
- Walmsley, J. (2008). Explanation in Dynamical Cognitive Science. *Minds and Machines* 18 (3): 331–48.
- Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy* 104 (12): 639–659.
- Weisberg, M. (2013). *Simulation and similarity: using models to understand the world*. New York: Oxford University Press.
- Woodward J (2002) What is a mechanism? A counterfactual account. *Philosophy of Science* 69: 366–S377
- Zednik, C. (2011). The Nature of Dynamical Explanation. *Philosophy of Science* 78 (2): 238-263.