Mechanistic Mayhem: Minding our Musings on Mechanisms

Eric Hochstein

Abstract: Mechanistic theories of explanation are widely held in the philosophy of science, especially in philosophy of biology, neuroscience and cognitive science. While such theories remain dominant in the field, there have been an increasing number of challenges raised against them over the past decade. These challenges claim that mechanistic explanations can lead to incoherence, triviality, or deviate too far from how scientists in the life sciences genuinely employ the term "mechanism". In this paper, I argue that these disputes are fueled, in part, by the running together of distinct questions and concerns regarding mechanisms, *representations of* mechanisms, and mechanistic *explanation*. More care and attention to how these are distinct from one another, but also the various ways they might relate, can help to push these disputes in more positive directions.

Keywords: Mechanistic Explanation, Ontic Mechanism, Localization, Decomposition, Dynamical Explanation, Pathway Explanation, Computational Explanation

If the Deductive-Nomological (DN) Model of explanation was the dominant philosophical view of scientific explanation in the early-mid 20th century, then the Mechanistic view of explanation would undoubtedly be one of the dominant views of scientific explanation in the early-mid 21st Century. But just like the DN model before it, serious challenges to the view have emerged over the years, and debates currently rage regarding the ontological nature of mechanisms, and the role of mechanistic explanation within scientific practice. In this paper, I argue that many of these disputes are fueled, in part, by the running together of distinct questions regarding mechanisms, representations of mechanisms, and mechanistic explanation. I intend to disentangle some of these issues and highlight how we might move forward.

In order to do this, I begin in section 1 by providing a brief overview of mechanistic explanation as currently characterized within contemporary philosophy of neuroscience and cognitive science. In section 2, I explore key debates that have emerged surrounding mechanisms and mechanistic explanation in science. In section 3, I attempt to show how these debates have brought with them numerous confusions, especially when it comes to the demarcation between mechanisms as things in the world (hereafter "ontic mechanisms"), representations *of* mechanisms (typically in the form of scientific models), and mechanistic *explanations* (understood as scientifically appropriate answers to why-questions that refer to mechanisms). Lastly, in section 4, I will explore how we might move forward in our understanding of mechanistic explanation.

1. Mechanisms and Mechanistic Explanation

Contemporary theories of mechanistic explanation in science rose to prominence in the aftermath of growing concerns that plagued the DN model of explanation in the second half of the twentieth century. This is well trodden territory at this point, so I will not rehash it here. But problems regarding the asymmetry of explanation (the height of the flagpole explains the length of the shadow, but not vice versa), relevance (a cis-gendered male takes birth control pills and does not get pregnant, but this does not explain *why* he does not get pregnant), and redescription (explaining why opium makes one fall asleep by appealing to the generalization that opium makes one fall asleep merely redescribes what is to be explained instead of explaining it), paved the way for the popularity of mechanistic explanation in the sciences.

Instead of explaining phenomena by way of sound arguments in propositional logic that appeal to antecedent conditions and law-like generalizations, the mechanistic theory of explanation argues that explanations of phenomena in science involve understanding how the phenomenon is being produced and sustained by the organized structures and processes in nature responsible for it. In this respect, scientific explanation often involves a kind of reverse engineering of a phenomenon by understanding the structures that constitute it, the way they are organized together, and the operations/activities/interactions that go on between them. It is this idea that fueled the rise of mechanistic theories of explanation, especially when dealing with the kinds of complex organised systems we study in the life sciences.

Contemporary discussions about mechanisms in philosophy of science center around a very particular account of mechanisms, one associated with what is now called the "New Mechanist" movement[1]. The moniker "New" in "New Mechanist" is intended to emphasize how contemporary scientific accounts of mechanisms differ in important respects from pre-twentieth century notions of mechanisms.

Under previous accounts, mechanisms were thought of as machines with clearly delineated and localized parts that are each designed to play a particular role in contributing to the system's overall functioning. These parts fit together in particular ways so that their interaction produces the system's overall behaviour; the way the gears and springs of a pocket watch are designed to fit together so that the winding of the watch can produce and sustain the movements of the hands in order to keep time. This traditional view of mechanisms brought with it certain assumptions about what was required for a mechanism. First, mechanisms are machines. Second, mechanisms (and the parts that make them up) have a *proper function*. In this regard, mechanisms

[1] Like any major philosophical movement, there is a great deal of variation and disagreement amongst its members, and so I will do my best to identify some general unifying features.

are understood teleologically (they are *supposed to* do things, and they are not doing what they are supposed to do when they break). Third, mechanistic explanation was inherently reductionistic in nature. To understand how the watch works, you need to reduce it to its component parts and the way they are put together.

The sort of mechanistic explanations that began to emerge in the mid-twentieth century differed in some key respects from this picture. While the New Mechanists likewise hold that mechanisms are to be understood as organized systems of parts and operations that together produce or instantiate some complex phenomenon, they do not adopt some of the assumptions that have traditionally been baked into the idea of mechanisms. For instance, they deny that mechanisms are machines (Glennan 2017, p.6-7; Pérez-González 2019, p.79; Bechtel 2020; Levy & Bechtel 2020; Craver, Tabery & Illari 2024; Kalewold 2024, p.29). Machines can be "on" or "off", but mechanisms in nature are always in action and never inert. Machines typically have parts that are clearly localizable and well bounded, but the parts of mechanisms in nature are less cleanly individuated and localized. As Craver, Tabery & Illary note, "components of mechanisms might be distributed and might violate intuitions about the boundaries of objects." (2024). Similarly, having a proper teleological function was no longer considered essential to the very nature of mechanisms. We need not adopt any sort of robust metaphysical view that mechanisms are "supposed to" do anything. All that is required is that there exist complex behaviours or phenomena in nature, and when such behaviour or phenomena are present it is the result of particular sets of structures, organized in particular ways, and interacting in particular ways. But we need not invoke any teleological interpretation of what the mechanism is "supposed to do" (e.g. Darden 2006).

Lastly, the New Mechanist view of mechanisms is not reductionistic in the way traditional accounts were. For the New Mechanists, breaking the system down into organised parts and operations is *part* of the story of how we explain its behaviour, but not the *entire* story. This is for several reasons. First, the way an organized system behaves is different from, and not understandable strictly in terms of, how the parts of the system in isolation behave. Second, biological mechanisms can often themselves be parts of even larger mechanism. When this happens, the way that a mechanism is situated within the larger system can directly change how it behaves and the phenomenon it produces. This is especially true when we consider things like complex feedback loops, as well as inhibitory, and regulatory mechanisms (e.g. Bechtel 2011; Darden 2017; Winning, & Bechtel 2018). In this respect, understanding why a mechanism did what it did involves understanding the context in which the mechanism is situated, and whether it is acting as a part in an even larger mechanistic system. It is for this reason that modern mechanistic explanation is considered "multilevel" and non-reductive (Machamer, Darden, & Craver 2000; Fehr 2004; Craver 2005, 2007, 2013; Bechtel 2007, 2009, 2011, 2022; Thagard 2008; Illari, & Williamson 2010; Glennan 2017; Piccinini 2022; Couch 2023).

This new picture of mechanistic explanation has the virtue of not only avoiding the traps that snared the DN model of explanation, but also seems to make sense of huge swathes of explanatory practices in scientific fields like physiology, biology, and neuroscience. In this respect, the motivation behind mechanistic explanation is compelling. How we flesh out the details of this story is where controversies start to arise.

2. Controversies Surrounding Mechanistic Explanation

While the idea that many phenomena in nature are scientifically explained by way of mechanistic explanation is an intuitive one, the reach of such explanations and what exactly is required for a mechanistic explanation, have been issues of controversy. Are *all* good explanations in science mechanistic, or only some? Which ones and why? What makes something a mechanism exactly, and what is the relationship between mechanisms and mechanistic explanations? These sorts of issues have raised considerable concerns. Here I will highlight some of these concerns and the debates that have sprung up around them.

2.1. Ontic versus Epistemic Views of Mechanistic Explanation

Consider a common way in which mechanisms are thought to play a role in explanations, embodied by the two following examples (from Craver 2007, p.27, and Craver & Kaplan 2020, p.300, respectively):

1. What explains the car crash? The fact that the brake line was cut, and the driver was drunk.
2. What explains why sea levels are rising? The warming of the planet due to increased carbon dioxide in the atmosphere.

In these cases, it is the mechanisms in the world that seem to be doing the explanatory work. Put another way, my saying the words "global warming" doesn't explain why sea levels are rising, it is the *actual warming of the planet* that does. Likewise, it is the actual cut brake-line and the impaired driving of the victim in the world that explains the accident. According to this way of talking, mechanisms in the world are the explanations themselves. It is in virtue of their structure and interactions that the phenomenon we wish to explain can be accounted for. This view of mechanistic explanation has come to be called the "ontic" view of mechanistic explanation, since it associates the ontic mechanism in the world with the explanation.

However, this way of talking about mechanistic explanation can often be at odds with other ways of talking about mechanistic explanations. Specifically, explanations are supposed to be answers to why-questions. They are the sorts of things that can be “right” or “wrong”, “good” or “bad”. But mechanisms *in the world* are not “good”, “bad”, “right”, or “wrong”. They just are. Explanations are things *we* as inquirers produce in order to answer why-questions. *We* explain the world by *appealing to* mechanisms in our explanations. As William Bechtel notes, “the problem with this ontic view is that mechanisms do not explain themselves” (2008, p.18). This has led some to argue that the ontic view of mechanistic explanation ultimately leads to incoherence by removing the explainer from the act of scientific explanation (Bechtel 2008, p.18; Wright, 2015; Bokulich, 2018). Mechanistic explanations under this alternative account are models or representations *of* mechanisms, not mechanisms themselves. This contrasting view of mechanistic explanation has come to be called the “epistemic” view of mechanistic explanation, since it emphasizes the epistemic nature of explanatory practice.

Advocates of the ontic view have responded to such criticisms by noting that our epistemic practices can *only* be normatively evaluated as good, bad, right, or wrong, in terms of whether they identify or describe the right ontic mechanisms in the world; those that we take to be what causally explains the occurrence of the phenomenon (e.g. Illari 2013; Craver 2014; Povich & Craver 2018). Disputes among defenders of the ontic and epistemic accounts of mechanistic explanation remain ongoing.

2.2. Idealization, abstraction, and the 3M criteria

Another central dispute regarding mechanistic explanation surrounds the so-called “3M” (or “Models-to-Mechanisms-Mapping”) criteria proposed by David Kaplan and Carl Craver (Kaplan 2011; Kaplan & Craver 2011). According to Kaplan and Craver, a scientific model is more explanatory the more the variables in the model can be mapped to structures and causes in nature that constitute and sustain the phenomenon we wish to explain. With this in mind, they propose that, *ceteris paribus*, the more the model describes the relevant underlying structures and causes in nature responsible for the phenomenon, the more explanatory the model becomes.

Yet some have argued that such a view seems to conflict with the role of abstraction and idealization in scientific modeling. Mazvitta Chirimuuta, for instance, characterizes the 3M criteria as a commitment to the idea that "the More Detail the Better" (or MDB). The problem is that it seems that some models in neuroscience are explanatory precisely *because* they violate MDB, and idealize or abstract away from the mechanistic details of the system. Using canonical neural computations as an example, she proposes that some computational models are explanatory not because they are intended to describe the underlying mechanisms of a given neural circuit, but because they explain by showing why a particular kind of computation (one which may be implemented by any number of different mechanisms) is the most efficient way to compute the required function to solve a problem. As she puts it:

> These *efficient coding explanations* account for observed properties of neural circuits in terms of the computational advantages of that particular arrangement of neurons. Note that the appeal to coding principles like redundancy reduction does not involve decomposition of any mechanism thought to underlie the behaviour in question. Rather, it takes an observed behaviour and formulates an explanatory hypothesis about its functional utility. (2014, p.144)

In response, Kaplan & Craver (2020) argue that their commitment to the 3M criteria was meant only to highlight that an explanation which identifies more of the *relevant features* of the mechanism are more explanatory than models that don't, and not that adding true details for the sake of true details will always make the model more explanatory. Moreover, that the use of such computational models still characterizes ontic mechanisms in the world, and in doing so *do* provide us with abstract mechanistic explanations. By leaving out essential details, they are simply limited in their explanatory power. This in turn has led to disputes about whether we should treat things like efficient coding explanations as a species of abstract mechanistic explanation, or as distinct types of explanations with different norms for evaluating their explanatory status.

There are two distinct, but related, points that Chirimuuta is making here, and two distinct, by related, debates about them. The first is that it is not the case that adding more details regarding the mechanisms of a system into our model makes it more explanatory. And second that certain kinds of models provide distinct non-mechanistic kinds of explanations in neuroscience. To avoid confusion, I will address debates about each of these claims separately. My intention in this section is to focus on the debate surrounding whether identifying more relevant details regarding the constitution of the system makes models more explanatory (the 3M criteria).

Similarly to Chirimuuta, Philippe Huneman (2010) argues that graph theoretic models can provide explanatory content regarding the topology of a system, but that such explanatory content is lost when more details regarding the mechanisms that implement the system are included. Anthony Chemero and Michael Silberstein (2008), as well as Lauren Ross (2015) make a similar argument regarding the use of dynamical models in cognitive science. In such cases, trying to include more details about the mechanism into the model makes it *worse* at representing various essential aspects of the phenomenon needed for a good explanation. These cases all seem to be directly at odds with the demands of the 3M criteria. Even the suggestion that we need only add more details of *relevant* features of the mechanism for the model to be more explanatory does not address whether models considered explanatory in neuroscience in fact do represent even these relevant features.

2.3. What makes something a mechanism exactly?

Another controversy surrounding mechanistic explanation is how exactly we ought to define what counts as a mechanism in the world. For example, for something to count as a mechanism, must the parts that make it up all have very particular roles within the system that cannot be accomplished if we swap the parts around with one another? Take the mechanism responsible for chemical neurotransmission. In order for chemical neurotransmission to occur, we cannot simply swap some parts (e.g. the ion channels) for others (e.g. the neurotransmitters) without the phenomenon failing to occur. On the other hand, the behaviours of fluids are thought to be the result of the interactions of its molecular components, yet it seems like we can swap around its constituent molecules without a change in overall flow behaviour of the fluid. Does the fluid case still therefore count as a mechanism? How broadly ought we to define what counts as a mechanism? If we define the term too loosely, then it seems like everything is a mechanism. If we define the term too strictly, then very few things are.

These concerns have led some to argue that mechanistic explanation in biology and neuroscience is far more problematic than we initially thought. Michael Silberstein, for instance, argues that the behaviour of neural systems cannot be understood and explained by merely decomposing the brain into localizable constituent parts and their interactions. This is because the brain is a complex overlay of networks, and system-level features of networks in general influences and constrains what the brain can, and does, in fact do in a range of different contexts and environments. These emergent high-level network properties and dynamics of the entire system cannot be understood by way of merely decomposition and localization of the brain into parts and operations. As Silberstein argues:

> The study of this integrative brain function and connectivity is mostly based in topological features or architecture of the network. Such multiply realized networks are partially insensitive to, decoupled from, and have a one-to many relationship with respect to lower-level neurochemical and wiring details. (2021, p.372)

This shows either that mechanistic explanation does not apply in many cases in neuroscience (since the brain is not understandable by merely breaking it down into the interaction of its constituent components), or else it requires broadening the definition of a mechanism so much that it makes mechanistic explanation trivial. As he puts it:

> Giving up the claim that localization and decomposition are both necessary and sufficient for mechanistic explanation, threatens to make the new mechanist philosophy too broad, non-unique or downright trivial. The essence of mechanistic explanation, what distinguishes it from mere causal or dynamical explanation, is its compositional or constitutive character. If the new mechanists jettison this feature of mechanistic explanation, if they fully acknowledge the essentially dynamical nature of such explanations and systems, it is not clear what if anything is unique about mechanistic explanation. (2021, p.363)

A related worry is that if *everything* is a mechanism, then appealing to mechanisms becomes explanatorily trivial.

2.4. How should we demarcate different kinds of explanations?

Lauren Ross (2021, 2022) has recently argued that to understand mechanistic explanation, we need to focus on how the *concept* of "mechanism" tends to be employed by working scientists within their explanatory practices. She proposes that regardless of whether something turns out to be a mechanism metaphysically, it often isn't *qua* mechanism that we explain its behaviour. For instance, she notes that there are all kinds of causal concepts we invoke in our explanations that are decidedly non-mechanistic in nature. Regarding mechanistic explanation, she tells us:

> In the biological sciences, 'mechanism' is often used to refer to causal systems that have a constitutive character, that are represented in significant, fine-grained detail, and that contain an emphasis on the 'force', 'action', or 'motion' of causal relations. This concept is associated with the causal investigative strategies of decomposition and localization and it is involved in an explanatory pattern where some outcome is explained by appealing to the causal components that produce it. (2021, p.136)

Contra claims made by the New Mechanists that mechanisms are *not* machines, she pushes back by noting that the mechanism concept, as employed by working scientists, is heavily dependent on the machine analogy. More specifically,

> Biologists frequently appeal to mechanisms in their explanations and descriptions of biological phenomena. They discuss mechanisms of gene regulation, DNA synthesis, nerve firing, muscle contraction, visual processing, and so on. When they use the mechanism concept they often suggest that some biological phenomenon can be understood as a kind of machine or mechanical system—such as a car engine or clock—in the sense of having

> particular features. This machine analogy encourages thinking of biological phenomena as having component parts that are spatially organized and that causally interact to produce some behaviour of the system. (2021, p.134)

By keeping her focus on the concept of "mechanism" as employed by scientists in their explanatory practices, she notes that scientists are not always interested in explaining phenomena by treating the system as a kind of machine in this way. Other kinds of causal concepts are invoked in different patterns of explanation. For instance, she contrasts mechanistic explanation with the way in which pathway explanations are invoked in biology…

> The pathway concept is commonly found in the biological sciences. Biologists refer to gene expression pathways, cell-signalling pathways, metabolic pathways, developmental pathways, circulatory pathways, neural pathways, and ecological pathways, just to name a few. In all of these cases the notion of a pathway refers to a sequence of causal steps that string together an upstream cause to a set of causal intermediates to some downstream outcome. For example, gene expression pathways track causal connections from genes, to their intermediate products to a final phenotype of interest. Signal transduction pathways track causal connections from an upstream signal, through intermediate transduction steps, to some final effect. (2021, p.136)

Unlike mechanistic explanation, in which "scientists identify a system and behaviour of interest and then 'drill down' to identify the system's parts, their location, and how they interact to produce the behaviour in question" (2021, p.134), pathway explanations "track the flow of some entity of signal through a system" (2021, p.139), analogizing the brain as a set of pathways or roads (instead of as a machine). Such explanations differ from mechanistic explanations in a number of key ways. First, the mechanism concept is not useful for tracking the flow of some entity or signal through a system, but instead to identify its constitutive structure. Second, mechanistic explanation requires identifying and describing more of the causal mechanism to be more explanatory, while a pathway explanation requires deliberately abstracting away from such details to be a good explanation. Lastly, pathway concepts emphasize connections, and not causal processes that underlie them. Or as she puts it, "Mechanisms involve specifying 'how' X causes Y, while pathways simply capture 'that' X causes Y. Given some set of entities in a system, the goal of the pathway concept is to show what is causally connected to what, as opposed to the fine-grained details of 'how' they are connected." (2021, p.144)

With this in mind, Ross argues that there are good reasons why scientists can and do distinguish "mechanistic explanations" from "pathway explanations". Different types of explanations are used to learn different sorts of things about a given system, and so distinguishing them, as working scientists do, is relevant to understand explanatory practice. This sort of claim captures Chirimuuta's intuition as well that efficient coding explanations are not *attempting* to provide mechanistic explanations, because they are not attempting to analyze the system in mechanistic terms. They are, instead, a distinct type of explanation. Huneman makes the same argument regarding topological explanations, and Chemero & Silberstein regarding dynamical explanations. In contrasts, some mechanists argue that in all these cases of explanation, we are still ultimately providing information about *the ontic mechanistic system in the world*, and thus all the models explain in the same way: by helping us learn about its ontic structure and workings, just in more or less abstract ways (e.g. Craver 2006; Kaplan & Craver 2011; Piccinini & Craver 2011; Zednik 2011; Povich 2018, 2021).

3. Mechanisms, Mechanistic Models, and Mechanistic Explanations

I propose that all the disputes mentioned above tend to frequently get caught up in particular kinds of confusions, conflations, and problematic commitments that have shifted the debates about mechanistic explanation away from what made such explanations philosophically important, as well as away from some of the genuine philosophical problems that still need to be addressed.

A central problem that has permeated these debates is the tendency for theorists to run together talk of mechanisms (as things in the world), *representations of* mechanisms (usually in the form of models or theories), and mechanistic *explanations* (scientifically appropriate answers to why-questions that refer to mechanisms). Of course, these things will all be importantly related, but the ways in which Ontic Mechanisms (OM), Representations *of* Mechanisms (RoM), and Mechanistic *Explanation* (ME) can relate to each other is complex and nuanced, and a failure to keep these things in mind are at the core of many disagreements.

3.1. Ontic vs Epistemic mechanistic explanation

Consider the dispute between ontic and epistemic views of mechanistic explanation. Defenders of the epistemic view criticize the ontic view for ignoring the role of the scientist in the scientific explanation of mechanisms. If explanations are answers to why-questions, then an explanation would seem to require an *explainer* who can provide such an answer. Things in the world aren't "answers" by themselves. Although an explainer might provide an answer to a why-question by *referring to* such things in the world.

Yet I propose that a closer look at what advocates of the ontic view say can help highlight that the dispute here is due more to an unfortunate terminological choice than anything substantive. For instance, Craver says:

> There are mechanisms (the objective explanations) and there are their descriptions (explanatory texts). Objective explanations are not texts; they are full-bodied things. They are facts, not representations. They are the kinds of things that are discovered and described. There is no question of objective explanations being "right'' or "wrong,'' or "good'' or "bad.'' They just are.
>
> Objective explanations, the causes and mechanisms in the world, are the correct starting point in thinking about the criteria for evaluating explanatory texts in neuroscience. [...] Good mechanistic explanatory texts (including prototypes) are good in part because they correctly represent objective explanations. Complete explanatory texts are complete because they represent all and only the relevant portions of the causal structure of the world. Explanatory texts can be accurate enough and complete enough, depending on the pragmatic context in which the explanation is requested and given. Objective explanations are not variable in this way. (2007, p.27)

This way of phrasing things appears at first glance to collapse ME and OM into one another. And this is where epistemic advocates push back. What counts as a mechanistic explanation surely involves learning about some mechanism in the world, but to suggest that mechanisms and explanations are identical seems to be confused. Yet a more careful analysis of what ontic advocates say can highlight why such an accusation is somewhat uncharitable. For instance, Craver is well aware that these things ought to be demarcated, he simply uses different terminology to do so. Craver notes that everyday ways of talking about "explanation" are ambiguous, and are sometimes used to refer to OM, sometimes to RoM, and sometimes to the way one gives an answer to a why-question in order to provide understanding to an audience. To respect this linguistic fact, he demarcates these instead in terms of what he calls *modes of explanation* (Craver 2014). When we talk about the mechanism in the world explaining the phenomenon, we are talking about explanations "under the ontic mode". When we talk about models, theories, or representations *of* mechanisms as explanations, then we are talking about explanations "in the textual mode". And when we talk about explanations as answers to why-questions intended to provide understanding of some mechanism to an audience, we are talking about explanation "in the communicative mode". And so the distinction between mechanisms, representations of mechanisms, and mechanistic explanations are all still very much intact and respected by advocates of the ontic view. They do not collapse these into one another, as their critics claim.

Put another way, one can capture everything the ontic advocate cares about without referring to the ontic mechanism in the world as an "objective explanation", or "an explanation in the ontic mode" if we find the terminology objectionable. And many advocates of the ontic view do exactly this. For instance, Jonathan Waskan characterizes the ontic view in the following way:

> Ontic theories might take many forms, so long as what they propose is that explanations (primarily) reveal something about objective states of affairs. [. . .] Explanations are, on this view, representations—objective facts are not in the business of revealing. Specifically, they are descriptions. (2011, p.4)

Likewise, Stuart Glennan claims that:

> Causal-mechanical explanation exemplifies what Salmon calls the ontic conception of explanation. Explanations are not arguments, but are rather descriptions of features of a mind-independent reality—the causal structure of the world." (Glennan 2002, p.S343)

The emphasis on "ontic" here acts more as a linguistic cue to highlight the fact that we cannot focus exclusively on the structure of our scientific representations at the cost of identifying the relevant causal structures and processes in the world when engaging in a scientific explanation (a mistake that the DN account fell into). The case of the flagpole and shadow, the birth control pill, and the opium causing drowsiness, highlight that ontic mechanisms in the world are essential for us to understand in order to explain the phenomenon. In this respect, the dispute between epistemic and ontic mechanists has hinged on a miscommunication regarding how each linguistically demarcates OM, RoM, and ME.

3.2. 3M or not 3M

Recall that some mechanists argue that dynamical, topological, and efficient coding accounts are all a type of abstract mechanistic explanation in virtue of telling us something about the OM in the world, albeit in a limited fashion (Craver 2006; Kaplan & Craver 2011; Piccinini & Craver 2011; Zednik 2011; Povich 2018, 2021). Now suppose for the sake of argument that we agree with this claim. Suppose we also accept the view that the more information we have about the relevant structures, organization, and activities of the OM, the better the explanation becomes. *Even if* we accept this, such claims do not by themselves provide support for the 3M criteria, since it implicitly conflates representations *of* mechanisms, with mechanistic *explanation*. In other words, believing that the more we describe, identify, or reveal the underlying mechanism of a phenomenon, the better we explain it, is not the same as claiming that the more *a particular model or representation* describes the structures and processes of the mechanism, the better an explanation *that model* becomes. The reason being that radically different kinds of models or representations, each of which idealize and abstract in different ways, are often required to effectively represent the different structural, organizational, or causal features of a mechanism needed to explain it (Hochstein 2016). In other words, the ME may be distributed across representations that *require* deviating from a correct account of the OM in their own ways. Thus, it is simply a mistake to focus on whether a particular model becomes a better mechanistic explanation by representing more of the relevant features of the mechanism simultaneously, since this may be impossible for issues having to do with the limitations of scientific representation. And so debates regarding the 3M criteria, even if we were to accept many of the claims that motivate it, ultimately rests on a confusion between RoM and ME.

3.3. What makes something a mechanism?

Debates about what makes something in the world an ontic mechanism have similarly been caught up in confusions regarding OM, RoM, and ME. For instance, Silberstein acknowledges this concern when he argues that:

> The problem […] is that the network examples too easily conflate concerns about explanation and abstraction on the one hand, with claims about organizational features of complex biological systems that tell against loc [localization] and decomp [decomposition],

> on the other. Obviously these two concerns are related but it is also important to disentangle them. (2021, p.366)

His claim is intended to highlight that biological systems *in the world* (i.e. the ontic systems) fail to be relevantly mechanistic, regardless of debates regarding abstract or idealized *representations of* such systems, or their relationship to *explanations*. Yet, ironically, Silberstein's claims are often confused in exactly this respect. For instance, as suggested in the quote above, he frequently argues that localization and decomposition are necessary organizational features of ontic mechanisms, and that biological systems often do not have such features. For a second example of this, take his claim that:

> Second, to make clear that the issue here is not primarily about abstraction or idealization in systems neuroscience, but whether or not complex biological systems really do embody loc and decomp as organizational features. (2021, p.367)

What Silberstein has in mind is that the behaviour and functioning of neural systems cannot be metaphysically accounted for purely in terms of the localized parts that make up the system and their interactions (this is because the components of neural systems are not always localizable or stable across contexts, and because the behaviour of the entire system cannot be fully accounted for strictly in terms of the interactions of its constituents in isolation). He argues that there are genuinely emergent properties of the entire system which put direct constraints on, and helps to explain why, the system is structured the way it is and behaves the way it does. He refers to this as *contextual emergence*.

The intuition underlying Silberstein's view is that topological and dynamical characterizations of systems allow us to identify essential constraints that explain *why* the system is organized the way it is, and how it can function. We don't metaphysically "build up" these system level properties merely from the interactions of the component parts by themselves. Instead, essential global features of complex systems that emerge from its contextual embedding explain why the parts and organization of the system must be the way they are. The problem with this criticism of mechanistic explanation in neuroscience is that it runs together different issues regarding OM, RoM*,* and ME.

For instance, he tells us that mechanistic explanation fails when trying to account for neurological systems because such systems do not "embody loc and decomp as organizational features". Yet localization and decomposition are methods by which we *study* and *understand* mechanisms in the world, not organizational features of mechanisms themselves. To claim that complex biological systems metaphysically embody localization and decomposition as organizational features is to collapse mechanisms in the world, and mechanistic explanation, into one another. Mechanistic *explanation* may involve our decomposing the system into what we take to be its localizable parts, but even mechanists themselves treat this as a heuristic for discovering the structure of ontic systems (Bechtel & Richardson 1993). Mechanisms in the world do not decompose themselves, nor is it a requirement of ontic mechanisms that its constituents must be cleanly localizable, well bounded, or stable across all contexts. Many mechanists have argued that a component part of a mechanism may be distributed across the system, or be context-dependent (e.g. Skipper & Millstein 2005; Illari & Williamson 2010; Burnston 2016, 2021; Craver, Tabery & Illari 2024). Thus, there is a conflation here between mechanisms, and explanatory strategies for learning about mechanisms. Or consider the way in which he describes "contextual emergence":

> With contextual emergence, global constraints and other kinds of context sensitivity are fundamentally at play. As Broad puts it, "[A]n emergent quality is roughly a quality which belongs to a complex as a whole and not to its parts" (Broad 1925, 23). According to him, if the properties of an irreducible whole are not given by the properties of the basic parts in isolation, they are emergent (see Humphreys 2016 for more details). For Broad, the global or systemic properties P of a system S are only reducible when the parts in isolation are sufficient to explain the existence of P. That is, there is reducibility when P can be derived or predicted in principle from the parts of S in isolation or when embedded in simpler systems.
>
> Contextual emergence emphasizes the ontological and explanatory fundamentality of multiscale contextual constraints, often operating globally over interconnected, interdependent, and interacting entities and their relations at multiple scales, e.g., topological constraints and organizational constraints in complex biological systems. Contextual emergence focuses on the fact that scientific explanation is often inherently and irreducibly multiscale. (2021, p.379)

But once again, different issues are being run together here. He tells us that systemic properties P of a system S are only reducible when "the parts in isolation" are sufficient to explain the existence of P. And there is "reducibility when P can be derived and predicted in principle from the parts of S in isolation or when embedded in simpler systems". But how are we to make sense of such claims?

Certainly, no mechanist would claim that we can predict the behaviour of the overall mechanism by studying its parts *in isolation from each other*. The way in which the parts of a mechanism are organized together is taken to be essential for something to even be a mechanism. Perhaps he means the organized parts *in isolation from any environmental context* would allow us to derive or predict the behaviour of the whole system? This can't be right either, since ontic mechanisms are always integrated systems, and they always exist in some environment or context. Central to the mechanist position is the idea that mechanisms are *multilevel*. In other words, a mechanism can itself be a part of a larger system. When this happens, the behaviour of that mechanism will change. In this respect, how the system is embedded in different contexts and environments has always been essential to understanding the behaviour of mechanisms for the New Mechanists.

We can, of course, *represent* the parts of mechanisms abstracted away from each other, or from the context in which they are embedded if we wish. And we can ask if such abstract representations of the parts in isolation allow us to fully *explain* the system's behaviour. But if this is what he has in mind, then his insistence that his view does not "conflate concerns about explanation and abstraction on the one hand, with claims about organizational features of complex biological systems" on the other (2021, p.366), and that "the issue here is not primarily about abstraction or idealization in systems neuroscience, but whether or not complex biological systems really do embody loc and decomp as organizational features" (2021, p.367) seems to be confused on exactly these grounds.

In other words, his claim that "the global or systemic properties P of a system S are only reducible when the parts in isolation are sufficient to explain the existence of P" is a claim about whether *abstract* or *idealized* representations of the parts in isolation are sufficient to *explain* the system level properties, and thus is very much a claim about abstraction, idealization, and explanation after all. Thus, he himself conflates concerns about explanation and abstraction on the one hand, with claims about organizational features of complex biological systems on the other.

3.4. How should we demarcate different kinds of explanation?

Let us now return to the question of whether efficient coding explanations, topological explanations, pathway explanations, and dynamical explanations are just a species of abstract *mechanistic* explanation, or whether they are entirely distinct *types* of scientific explanations. Some mechanists argue that these are all species of mechanistic explanation in virtue of providing explanatory content regarding some *ontic mechanism* in the world, or fitting the phenomenon into the "causal structure of the world". Conversely, Ross argues that such a view is at odd with what neuroscientists mean when they invoke "mechanistic explanation" within scientific practice. In such cases, they have something like the "machine analogy" driving their understanding of what a mechanistic explanation is. Conversely, many mechanists explicitly deny we should adopt the machine analogy when interpreting mechanisms in neuroscience. How should we settle these conflicts? I propose that the disagreement here is due to a confusion about what each group is trying to make sense of. The mechanists are interested in whether *ontic mechanisms in the world* are machines, or have the properties we have attributed to human made machines. Ross, Chirimuuta, and others are interested in how working scientists tend to think of *mechanistic explanation* in scientific contexts. In this regard, both are making important points, but about different things.

For the mechanist, the idea that mechanisms in nature are not machines is a metaphysical claim about what our best empirical findings tell us ontic mechanisms in the world are like. This claim is not about how scientists use the term "mechanism" in their explanatory practices, but instead a more general claim about what we have the best reasons to think ontic mechanisms actually are. Ross, on the other hand, isn't focused on the ontological nature of mechanistic systems, but about how scientists use the concept of "mechanism" when explaining phenomena.

According to Ross, mechanistic explanation is invoked when scientists try to "drill down" to identify the parts and organization of a system that *are* sufficiently machine like. Yet not all explanations in neuroscience and biology are interested in doing this, and not all biological systems are machine-like in this way. In this regard, Ross is interested in when and how we classify an explanatory strategy in science as "mechanistic", not whether ontic systems in nature must always meet some machine-like threshold in order to metaphysically count as an ontic mechanism.

Likewise, consider mechanists who view all explanations as showing how things fit into the causal structure of the world. Even if we accept such a claim, it may still be the case that the sorts of models we classify as "mechanistic explanation" in neuroscientific and biological contexts are importantly demarcated from the sorts of models we classify as "pathway explanations", or "efficient coding explanations", even if they all happen to situate things in the causal structure of the world in some way or another. In essence, the question of how models gain explanatory power in some general sense from representing the world (by potentially "situating things in the causal structure of the world" or not), may simply be a different sort of question than how and why we should demarcate different types of models in terms of conveying different types of explanations, or explanatory strategies, for different scientific reasons. And thus Ross's point need not be in tension with the New Mechanists she criticizes.

Of course, it could turn out that how we demarcate our successful explanatory strategies constrains or structures how we understand which systems in the world should count as ontic mechanisms and which should not. Or conversely, that determining what ontic mechanisms in nature are like constrains or structures how we ought to demarcate our explanatory strategies. But this is an open question that must be investigated. We need to shift the debate to *how* our explanatory strategies, and our understanding of the structure of the world, do or do not constrain and influence each other (as opposed to whether one side or the other has mischaracterized mechanistic explanation).

4. Moving Forward

So where do we go from here? With these issues brought to the fore, I want to offer some new insights that may help move things forward. First, I think debates surrounding the 3M criteria have overly focused on individual models and their ability to explain more by representing more, and in doing so have lost an important insight lurking in the background. In other words, I propose that Kaplan & Craver confuse two different points: (1) the more we know about the underlying mechanisms that constitute and realize a phenomenon, the better our ability to explain it. (2) the more a single model has variables that map to, or describe, the parts and operations of the mechanism, the more explanatory that model becomes. I want to suggest that (2) is false, but that there is a way of understanding (1) that is not, and it points to an important lesson.

To understand the sense of (1) that I have in mind, let us consider a few examples. Chirimuuta argues that efficient coding explanations in neuroscience are not mechanistic explanations. This is because such explanations do not try to uncover the underlying mechanisms of the system, but instead explain by identifying what sort of computations will allow the system to solve the relevant problem with minimum resource expenditure and time. In essence, it is the efficiency of a particular computational solution to a problem that explains why some neural circuits behave the way they do. This can explain why neural circuits with very different structures may all implement the same computational function: because that is the most efficient way to solve problems of a particular sort. Likewise, the explanation won’t be mechanistic since different mechanisms can fall under the same efficient coding explanation.

For the sake of argument, let's grant all of this. So there is a class of systems (differently structured neural circuits) that have the same behavioural pattern, which is explained by the efficiency of a particular computation in solving a problem (in virtue of minimizing time and resource costs). Even if we accept this, we must keep several things in mind. First, computational theory tells us that a change in the underlying physical implementation of the system (its computational architecture) will directly affect what *sorts* of computations it will be efficient at carrying out (Eliasmith 2002). And so the underlying mechanisms of the system will directly determine how resource intensive the same computation will turn out to be when carried out by one kind of neural circuit as opposed to another. In effect, the same type of computational solution may be efficient when implemented by one neural circuit, but not when implemented by another (depending on how it's structured). And thus knowing how the circuit is structured will directly determine if a given computational solution *in fact is* the most efficient way for *that particular system* to solve a given problem, and thus whether the efficient coding explanation will work.

In essence, I want to argue that while efficient coding explanations may be best captured by models that deliberately abstract and idealize away from mechanistic details of the system, it is still very much the case that *the more we know about the underlying structures and mechanisms of the system*, the better our ability to use such idealized models to explain, to know when such explanations will be appropriate, and when they will not. By "we" here, I mean the scientific community as a whole, and by "knowing about the underlying structures and mechanisms" I mean the collective information we've gathered that is distributed across many different models and theories from many different domains that highlight the various structural and causal aspects of the system (its parts, organization, activities, etc).

Let us consider some other examples. Silberstein argues that many explanations in biology require paying attention to system level constraints and principles that cannot be inferred from the compositional structure of the system. Using topological explanations as an example, he tells us that "it seems clear that global topological features of network models, help explain, and are in turn explained by global organizational features of complex biological systems such as robustness, plasticity, autonomy and universality." (p.378) But again, there are important things to note here. For instance, whether a global organizing feature like plasticity can explain relevant features of a system's behaviour depends on what kind of system we are talking about. Plasticity comes in degrees, and the brain has a great deal more plasticity than does a hunk of iron. Our understanding of whether neural plasticity can act as an organizing principle for neural and cognitive processes depends on our understanding of the mechanisms underlying those neural and cognitive processes (such as long-term potentiation). This is because those sorts of mechanisms determine the degree and extent of this plasticity, and thus how much that principle will apply to our understanding of the system.

To emphasize this point, consider the long-standing dispute between connectionists and modularists in cognitive science. Those who adhere to a particular brand of modularity argue that the brain is composed of innate evolutionarily specified self-contained modules that evolved in isolation to carry out particular computations. Under this view, the organizing feature of plasticity in the brain is much more constrained, since the innately specified architecture of the brain is set and cannot be radically rewired willy-nilly. Conversely, if we adopt a more connectionist picture, and consider the brain to be a set of overlapping networks that constantly rewire, then plasticity would play a much bigger organizing principle in explaining cognitive and neural behaviour. But understanding if the brain is more modular, or more interconnected, requires understanding more about the underlying mechanisms of the brain. This will likewise apply to robustness (a feature of certain organized networks, but not necessarily modular systems). Likewise, the notion of universality is a case in which a high-level property or pattern is shared by a class of differently structured systems. But not *any* system will fall under a universality class. And knowing more about the underlying structures and processes of those systems will directly inform whether they fall under the university class, and under what conditions.

Lastly, consider Lauren Ross's example of pathway explanations, which she contrasts with mechanistic explanations. When talking about metabolic pathways, for instance, the way in which the pathway is mechanistically instantiated is not relevant to the pathway explanation, since *how* the metabolic processes allow a signal to be carried through the system is not what's relevant, only that it *is* being carried through the system. However, the conditions under which the pathway will allow this information to be carried throughout the systems, and the conditions under which it will not, typically require understanding *how* the pathway is implemented, since in different organisms the functioning of such pathways will work or breakdown in very different ways, and under very different conditions. Thus the more we know about the mechanisms that can implement metabolic pathways in different organisms, the more it directly influences how and when we use pathway explanations effectively.

So let's bring this back to the 3M criteria. There is an important intuition that Kaplan and Craver are getting at, which is that idealized models are often considered more explanatory the more the scientific community has accurate information regarding how the system is implemented mechanistically *even though the idealized model itself does not include such information in its explanation*. In other words, if we have no information whatsoever on how metabolic pathways work, how they are structured differently in different organisms, when they work, and when they don't, then even if the pathway model answers the relevant why-question, we are less likely to consider it a good explanation than if we use the identical model to answer the identical question *when such mechanistic information is available for us to consult, and informs our understanding of the idealized model*. Thus, there's a difference between my not needing to directly refer to such information when employing a pathway explanation, and the scientific community not *having any implementational information whatsoever* when employing a pathway explanation. In this regard, there is an important point underlying Kaplan and Craver's claim here, but framing it in terms of what information must be included in a particular model for it to be explanatory was simply the wrong lesson to learn from it.

Next, what are we to make of the claim that if mechanists broaden their definition of what a mechanism is too much, then mechanistic explanation becomes trivial? I think such worries are largely overblown. Recall Silberstein's insistence that:

> Giving up the claim that localization and decomposition are both necessary and sufficient for mechanistic explanation threatens to make the new mechanist philosophy too broad, non-unique or downright trivial. The essence of mechanistic explanation, what distinguishes it from mere causal or dynamical explanation, is its compositional or constitutive character. If the new mechanists jettison this feature of mechanistic explanation, if they fully acknowledge the essentially dynamical nature of such explanations and systems, it is not clear what if anything is unique about mechanistic explanation. (2021, p.363)

Is this true? Let's suppose he is correct that localization and decomposition are insufficient to fully account for the behaviour of biological systems due to system-level emergent properties. If we broaden the definition of "mechanism" to include systems of this sort, does this mean there is no longer anything that distinguishes mechanistic explanation from mere causal or dynamical explanation? It's not clear why he would think so.

Consider once again the infamous opium example. What explains why one falls asleep after taking opium? We can provide a causal explanation of this phenomenon by insisting that the opium *causes one* to fall asleep (identifying a causal regularity). This is surely a causal explanation, but not a particularly informative one. The explanation seems to require understanding *how the chemical composition of opium interacts with the physiological structures and processes within the subject to induce unconsciousness.* While both explanations are causal, the second explanation is relevantly different from the first precisely in virtue of decomposing the system into structures and processes.

Now suppose we accept Silberstein's claim that the body is *not* a mechanistic system under a strict definition of "mechanism", given that it has system-level contextually emergent properties not fully accounted for by localization and decomposition. If we widen our definition of a mechanism to *include* systems like the human body, does this now mean there is no difference between the two explanations of opium just given? The adoption of a more liberal definition of a mechanism that includes dynamic systems with contextually emergent properties does not appear to change the fact that to explain why opium makes one fall asleep still involves understanding the physiological structures and processes that interact with opium to bring about unconsciousness. In other words, if we broaden our definition of a mechanism, it does not undermine the idea that decomposition and localization *are* still essential to how we explain such systems, and that this is still a defining feature of mechanistic explanation. At most, it suggests that such implementational accounts abstracted away from context and environment are often not *sufficient by themselves* to account for *all* of the system's behaviours. But this is in no way makes mechanistic explanation trivial, or no different from other kinds of causal or dynamical explanations.

What about the more general criticism that if *everything* turns out to be a mechanism, then mechanistic explanation becomes vacuous? This too seems false. Consider a different kind of example: dynamical explanations. Virtually every system in nature is a *dynamical* system that changes over time. Does the fact that *everything* is a dynamical system thereby trivialize dynamical descriptions in science, or make dynamical explanations vacuous? Not at all. Why? Because not *any* set of differential equations will accurately describe the dynamics of *any* system. Similarly, we can use probability theory to create a statistical model of any system in nature we want. Does this suggest that statistics, or statistical models in science, are vacuous or trivial? Again, no. Because not *any* statistical description accounts for the behaviour of any system. If everything turns out to be an ontic mechanism, then the relevant question is not *if* the system is mechanistic, or dynamical, or statistical. It's what is the *correct*, or *appropriate*, or *useful* characterization *of* the mechanism, or the dynamics, or the probabilistic behaviour of the system. In our parlance, there is a distinction between mechanisms in the world, and mechanistic explanations. Even if everything turns out to be an ontic mechanism, the question of what the appropriate mechanistic explanation of that system is, and how best to represent it, is still very relevant.

Put simply, whether we choose a more liberal, or a more stringent, account of what makes something an ontic mechanism is far less important than whether we are consistent in our usage of the term "mechanism". The lack of standardization means disputes have emerged in part because people are not consistent in what they have in mind by a mechanism in the world. But I don't see why either a liberal or stringent account of what an ontic mechanism is would be problematic *a priori* for mechanistic theories of explanation. It is something we must investigate.

Conclusion

It is important to stress that the goal of this paper has not been to suggest that our philosophical understanding of mechanisms or mechanistic explanation does not face metaphysical, conceptual, or scientific problems. How we ought to understand explanation (mechanistic or otherwise) remains an important issue, as does the relation between models and explanations, and the relation between explanation and the structures in the world. My intention instead is to highlight that many current disputes surrounding mechanisms and mechanistic explanation have persisted in part due to a number of confusions, miscommunications, and problematic assumptions. Getting clear on such issues can help refocus debates onto the more substantive issues that need to be addressed. With this in mind, I hope to have pushed us towards more solid ground for the debates to come.

References:

Bechtel, W. (2007). Reducing Psychology While Maintaining its Autonomy Via Mechanistic Explanations. In Schouten, M. and De Joong, H.L. (eds.), The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience and Reduction. Blackwell Publishing.

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. (2011). Mechanism and biological explanation. Philosophy of science, 78(4), 533-557.

Bechtel, W. (2020). Living machines: the extent and limits of the machine metaphor. In Philosophical Perspectives on the Engineering Approach in Biology (pp. 79-96). Routledge.

Bechtel, W. (2022). Levels in biological organisms: Hierarchy of production mechanisms, heterarchy of control mechanisms. The Monist, 105(2), 156-174.

Bechtel, W., & Richardson, R.C. (1993). Discovering complexity: Decomposition and localization as strategies in scientific research. Princeton, NJ: Princeton University Press.

Bokulich, A., (2018). Representing and explaining: The eikonic conception of scientific explanation. Philosophy of Science, 85(5), pp.793–805. https://doi.org/10.1086/699693.

Burnston, D. (2016). A Contextualist Approach to Functional Localization in the Brain. Biology & Philosophy 31:527–50. https://doi.org/10.1007/s10539-016-9526-2.

Burnston, D. C. (2021). Getting over atomism: Functional decomposition in complex neural systems. The British journal for the philosophy of science.

Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. Philosophy of Science, 75, 1–27.

Couch, M. (2023). “Clarifying the Relation Between Mechanistic Explanations and Reductionism”, Frontiers in Psychology, 14: 984949. doi:10.3389/fpsyg.2023.984949

Craver, C. (2005). “Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience”, Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36(2): 373–395. doi:10.1016/j.shpsc.2005.03.00

Craver, C. (2006). When mechanistic models explain. Synthese, 153(3), 355-376.

Craver, C. (2007). Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience, Oxford: Clarendon Press. doi:10.1093/acprof:oso/9780199299317.001.0001

Craver, C. (2013). Functions and Mechanisms: A Perspectivalist View. In Philippe Huneman (ed.), Functions: Selection and Mechanisms (Synthese Library 363). Dordrecht: Springer, 133–158. doi:10.1007/978-94-007-5304-4_8

Craver, C. (2014). The Ontic Account of Scientific Explanation. In: M.I. Kaiser, O.R. Scholz, D. Plenge and A. Hüttemann, eds. Explanation in the Special Sciences: The Case of Biology and History. Vol. 367, Synthese Library. Dordrecht: Springer Netherlands, pp.27–52. https://doi.org/10.1007/978-94-007-7563-3_2.

Craver, Carl, James Tabery, and Phyllis Illari, "Mechanisms in Science", The Stanford Encyclopedia of Philosophy (Fall 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2024/entries/science-mechanisms/>.

Craver, C. and Kaplan, D. (2020). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. The British Journal for the Philosophy of Science, 71(1): 287–319. doi:10.1093/bjps/axy015

Darden, L. (2006) Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution (Cambridge Studies in Philosophy and Biology), New York: Cambridge University Press. doi:10.1017/CBO9780511498442

Darden, L. (2017). Strategies for discovering mechanisms 1. In The Routledge handbook of mechanisms and mechanical philosophy (pp. 255-266). 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. http://dx.doi.org/10.1080/09528130210153514.

Fehr, C. (2004). Feminism and science: Mechanism without reductionism. NWSA Journal, 136-156.

Glennan, S. (2017). *The New Mechanical Philosophy*. Oxford: Oxford University Press.

Hochstein, E. (2016). One Mechanism, Many Models: A Distributed Theory of Mechanistic Explanation. Synthese 193 (5): 1387-1407. DOI: 10.1007/s11229-015-0844-8

Huneman, P. (2010). Topological explanations and robustness in biological sciences. Synthese, 177(2), 213-245.

Illari, P., (2013). Mechanistic explanation: Integrating the ontic and epistemic. Erkenntnis, 78(S2), pp.237–255. https://doi.org/10.1007/s10670-013-9511-y.

Illari, P., & Williamson, J. (2010). Function and organization: comparing the mechanisms of protein synthesis and natural selection. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 41(3), 279-291.

Kalewold, K. H. (2024). The hybrid account of activities. Synthese, 203(1), 29.

Kaplan, D. (2011). Explanation and Description in Computational Neuroscience. Synthese, 183(3): 339–373. doi:10.1007/s11229-011-9970-0

Kaplan, D., and Craver, C. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. Philosophy of Science 78(4): 601–27. doi:10.1086/661755

Levy, A., & Bechtel, W. (2020). Beyond machine-like mechanisms. In Philosophical perspectives on the engineering approach in biology (pp. 99-122). Routledge.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about Mechanisms. Philosophy of Science, 67(1): 1–25. doi:10.1086/392759

Pérez-González, S. (2019). The search for generality in the notion of mechanism. Teorema: Revista internacional de filosofía, 38(3), 77-94.

Piccinini, G. (2022). Neurocognitive mechanisms a situated, multilevel, mechanistic, neurocomputational, representational framework for biological cognition. Journal of Consciousness Studies, 29(7-8), 167-174.

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. Synthese, 183, 283-311.

Povich, M. (2018). Minimal Models and the Generalized Ontic Conception of Scientific Explanation. The British Journal for the Philosophy of Science 69 (1): 117–137.

Povich, M. (2021). The Narrow Ontic Counterfactual Account of Distinctively Mathematical Explanation. The British Journal for the Philosophy of Science 72 (2): 511–543.

Povich, M. and Craver, C. (2018). Because without cause: Non-causal explanations in science and mathematics (review). Philosophical Review, 127(3), pp.422–426. https://doi.org/10.1215/00318108-6718870.

Ross, L. N. (2015). Dynamical models and explanation in neuroscience. Philosophy of Science, 82(1), 32-54.

Ross, L. N. (2021). Causal concepts in biology: How pathways differ from mechanisms and why it matters. The British Journal for the Philosophy of Science.

Ross, L. N. (2022). Cascade versus mechanism: The diversity of causal structure in science.

Skipper, R., and Millstein, R. (2005). “Thinking about Evolutionary Mechanisms: Natural Selection”, Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36(2): 327–347.

Thagard, P. (2008). Hot thought: Mechanisms and applications of emotional cognition. MIT press.

Waskan, J., (2011). Intelligibility and the CAPE: Combatting Anti-Psychologism about Explanation. Available at: <http://philsci-archive. pitt.edu/8530/> [visited on 4 October 2023].

Winning, J., & Bechtel, W. (2018). Rethinking causality in biological and neural mechanisms: Constraints and control. Minds and Machines, 28, 287-310.

Wright, C. (2015). The ontic conception of scientific explanation. Studies in History and Philosophy of Science Part A, 54, pp.20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Zednik, C. (2011). The Nature of Dynamical Explanation. Philosophy of Science 78 (2): 238-263.