OXFORD UNIVERSITY PRESS | The British Journal for the Philosophy of Science

Are More Details Better? On the Norms of Completeness for Mechanistic Explanations

Journal:	*The British Journal for the Philosophy of Science*
Manuscript ID	BJPS-2016-364.R3
Manuscript Type:	Article
Keywords:	explanation, abstraction, models, minimal models, mechanism

Are More Details Better? On the Norms of Completeness for Mechanistic Explanations[1]

ABSTRACT:

Completeness is an important but misunderstood norm of explanation. It has recently been argued that mechanistic accounts of scientific explanation are committed to the thesis that models are complete only if they describe everything about a mechanism and, as a corollary, that incomplete models are always improved by adding more details. If so, mechanistic accounts are at odds with the obvious and important role of abstraction in scientific modeling. We respond to this characterization of the mechanist's views about abstraction and articulate norms of completeness for mechanistic explanations that have no such unwanted implications.

[1] Thanks to Julia Staffel for suggesting the form of this essay and feedback on its content. Thanks also to Eric Hochstein, Lena Kastner, Colin Klein, Mark Povich, Graham Priest, Felipe Romero, Jim Woodward, and two or more anonymous referees for comments on earlier drafts.

1. Introduction

Is completeness an important dimension of explanatory progress? If so, what counts as progress along this dimension? These questions have taken center stage in recent philosophical discussions about the norms of mechanistic explanation.[2]

According to a common line of argument, mechanistic accounts of scientific explanation are committed to the thesis that "More Details are Better" (hereafter **MDB**): that models are always improved by adding more details. This commitment, critics claim, is inconsistent with the central role of abstraction in modeling. Specifically, they argue that explanations are often improved by dropping detail to focus on core relevant features; thus, mechanistic completeness is not an ideal of scientific explanation. Worse, they argue, by emphasizing the importance of detail at all costs, mechanists reinforce scientific trends, fueled by high-throughput, large-scale data collection and computational simulations, to emphasize massive data at the expense of analytical insight into the deep structures or patterns that emerge through abstraction.[3] As neurobiologist Eve Marder puts it: "Understanding how the brain works requires a delicate balance between the appreciation of the importance of a multitude of biological details and the ability to see beyond those details to general principles" (Marder [2015], p. 1).[4]

Mechanistic approaches, the critic alleges, upset this balance and reinforce the drive towards maximal detail and away from the theoretical insights required for deep explanations.

---

[2] Several recent papers target similar issues. Baetu [2015] addresses mechanistic completeness, but his primary aim is to show how it can be defined operationally and tested empirically. According to Baetu, showing that a set of parameters in a model or simulation are sufficient to produce the phenomenon provides evidence that it is mechanistically complete. Along similar lines, Boone and Piccinini ([2016a], [2016b] argue that mechanistic explanation is compatible with various kinds of abstraction. They focus specifically on mechanistic abstraction as a way to achieve sufficient generality in both the explanandum and explanans. We argue that this is a matter of how the explanandum is contrastively formulated. Finally, Milkowski [2016] addresses the explanatory role of large-scale brain simulations in neuroscience and is concerned with clarifying how norms of completeness apply to these models. He argues that abstraction is, and has always been, compatible with mechanistic explanation because mechanistic explanations include only causally or constitutively relevant details. We argue for the same point below.

[3] We shall not address whether all scientific explanations describe causes or mechanisms. Rather, we address whether mechanistic explanations are improved by adding details. There may be non-mechanistic forms of explanation (e.g., mathematical, geometrical, or intentional) that have different norms, and for which talk of mechanistic detail may be wholly inappropriate. We do claim, however, that any acceptable explication of these forms of explanation must help us to understand the difference between good and bad explanations of these types.

[4] Some critics also claim that mechanistic views are incompatible with idealization (i.e., deliberate falsification). Many of the arguments below apply, *mutatis mutandis*, to idealization. Here, we focus on abstraction.

This criticism first appeared in philosophical work concerning the Hodgkin-Huxley model of the action potential by Shaffner ([2008]), Weber ([2008]), and Levy ([2013]), but it sharpened considerably in a set of papers in 2013 and 2014. Woodward [2013], for example, contrasts detail-oriented mechanistic models with network models:

> ...it is the generic features of the network topology and not the details of those mechanistic relationships that, so to speak, drive the explanations provided (Woodward [2013], p. 41].

Also that year, we find a similar contrast by Levy and Bechtel ([2013]):

> …abstract models play a role in explaining the particular behavior of particular systems. This seems to be in direct contrast with the remarks from Machamer, Darden and Craver concerning [mechanism] schemata (Levy and Bechtel [2013], p. 259).

Batterman and Rice ([2014]) are more explicit in characterizing the problem and its source:

> [According to mechanists], not only does veridical representation of the causal mechanism make a model explanatory, the more accurate and detailed that representation is, the more explanatory the model will be.... [T]his view mistakenly implies that more accurate detail concerning mechanisms is always better (Batterman and Rice [2014], p. 352).

And Levy ([2013]):

> [W]hat seems to be missing from the mechanistic outlook is an analytical category: a notion that would cover cases in which a model is deliberately 'sketchy'…In other words, the judgment that the H[odgkin and] H[uxley] model [of the action potential] is a sketch stems, I think, from a gap in the mechanistic outlook itself, in which room has not been made for the explanatory fruits of abstracting away from structural detail. (Levy [2014], p. 488)

Chirimuuta ([2014]) brings the argument full circle by connecting the mechanistic preoccupation with details to certain myopic trends in big data science:

> One principled reason why we are likely to need multiple perspectives [beside mechanistic modeling] is, simply put, because the brain is complex…. There are approximately 86 billion neurons in the brain, with trillions of synaptic connections between them.... When faced with such challenges, the standard scientific response is to

simplify the problem space: restrict attention to a limited range of causally significant components and forget about trying to model all of them (Chirimuuta [2014], p. 149).

This sample indicates an emerging consensus that mechanistic views of explanatory completeness are hostile to the role of abstraction in scientific explanation.[5]

We are surprised by this attribution of **MDB** to mechanists and so seek in this paper to root out its sources. We discuss the exegetical sources of this attribution in Section 2. In Section 3, we lay out basic mechanistic commitments about what a philosophical theory of scientific explanation should achieve. In Section 4, we focus specifically on norms of constitutive mechanistic explanation the critics believe imply **MDB**. We show they do not. In Section 5, we introduce the idea of Salmon-completeness and argue for its value in defining the limit of mechanistic completeness. In Section 6, we show how our views on explanatory relevance dispatch many of the critics' central arguments. In Section 7, we distinguish our norms of completeness from the other benefits of abstraction the critics emphasize. In Section 8, we argue, contra the supposition of many of the critics, that models are not the proper focus discussions of the norms of explanatory completeness. This sets the stage for our positive formulation of the norms of mechanistic explanatory completeness, $\mathbf{MC}_{\mathbf{individual}}$ and $\mathbf{MC}_{\mathbf{science}}$, elaborated in Section 9. Along the way, we consider a number of different candidate theses and make our commitments explicit. What emerges is a defensible and consistent view of the norms of mechanistic explanatory completeness that, we think, makes better sense of the practice of science than the norms of completeness the critics appear to endorse.

2. A Balancing Act: When Do Details Matter?

The **MDB** critics focus on two central questions any theory of mechanistic explanation should address: When is an explanation improved by adding details? And when is an explanation improved by dropping details? To answer these questions simultaneously requires a sensitive balance between two competing and uncontroversial intuitions: first, that knowing more about how a mechanism works, on balance, improves one's understanding of how the mechanism works; and second, that unnecessary detail often obscures the abstract causal relationships by

[5] These authors cite one another and clearly are bolstered in their confidence by this mutual agreement. In our view, this mini-consensus rests on a foundation of shared mistakes and misunderstandings. The **MDB** criticism, and the mistakes behind it, have now been repeated elsewhere (e.g., Brigandt [2013]; Rice [2015]; Ross [2015]) and are increasingly accepted without further argument as central problems for mechanistic approaches to explanation.

which the mechanism works.[6] If one emphasizes mechanistic writings on one of these questions at the expense of their writings on the other, the balance is upset, and a lopsided view emerges.

Precursors of the **MDB** criticism first arose in response to Craver's [2006] treatment of the Hodgkin-Huxley (HH) model of the action potential. He argues that Hodgkin and Huxley's mathematical models of how specific ionic conductance across the membrane change during an action potential are merely "phenomenally adequate": The models capture the inputs (voltage) and outputs (conductance changes) with high precision but do not explain why conductance changes with voltage. Hodgkin and Huxley's model of the action potential therefore contains black boxes that, as they emphasize, they could not fill with relevant mechanistic details. Their understanding of the mechanism is only a sketch:

> Models that describe mechanisms can lie anywhere on a continuum between a *mechanism sketch* and an *ideally complete description of the mechanism*. A *mechanism sketch* is an incomplete model of a mechanism. It characterizes some parts, activities, and features of the mechanism's organization, but it has gaps...At the other end of the continuum are *ideally complete descriptions of a mechanism*. Such models include all of the entities, properties, activities, and organizational features that are relevant to every aspect of the phenomenon to be explained. (Craver [2006], p. 360)

Sketches are incomplete in this "ideal"[7] sense because they do not describe all the relevant details. Craver [2006] at one point claims, falsely, that the HH model "does not explain the action potential." This injudicious remark has been taken to suggest that sketches cannot provide even partial explanations.[8] We return to this example below.

---

[6] In speaking of "abstract causal relationships" we mean causal relationships that only come into view with abstraction, not causal relationships between abstracta.

[7] "Ideal" here does not mean "best," but rather "complete in the idealized sense of completeness". As noted in Craver [2006], p. 360), the phrase is a nod to Railton's notion of an "ideal explanatory text," introduced to protect the covering-law model from the (irrelevant) charge that scientists rarely spell out the complete argument for a given explanandum.

[8] The quoted claim is false because the HH model, suitably interpreted, describes explanatory causal relationships among conductance, current, and voltage. However, the model does not provide a constitutive, mechanistic explanation for conductance changes. This sentence was corrected in (Craver [2007], [2008]) based on discussions with Jim Bogen [2008], Ken Schaffner [2008], and Marcel Weber [2008].

Critics also ground their attribution of **MDB** to mechanists by appeal to the model-to-mechanism-mapping[9] (**3M**) condition placed on a model's explanatory power:

> We summarize the above considerations in the form of a model-to-mechanism-mapping requirement that makes the commitment to a mechanistic view of explanation explicit: (**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. (Kaplan and Craver [2011], p. 611)

Some take it to follow from **3M** that a model that describes more of the components, activities, properties, and organizational features of a mechanism is more explanatory than one that describes fewer such features (Chirimuuta [2014]).

Yet **3M** is a necessary condition on explanatory mechanistic models; **MDB** is a comparative claim. That said, Kaplan [2011] appears to connect **3M** explicitly to a thesis like **MDB**:

> **3M** aligns with the highly plausible assumption that the more accurate and detailed the model is for a target system or phenomenon the better it explains that phenomenon, all other things being equal (for a contrasting view, see Batterman [2009]). As one incorporates more mechanistically relevant details into the model, for example, by including additional variables to represent additional mechanism components, by changing the relationships between variables to better reflect the causal dependencies among components, or by further adjusting the model parameters to fit more closely what

[9] The term "mapping" is not intended to imply a one-to-one correspondence between each element in the mechanism and each element in the model, and vice versa. This misinterpretation would imply that **3M** is hostile to idealization – an implication we explicitly disavow (Kaplan [2011]). Instead, the intended mapping works only in one direction. **3M** insists that the explanatory force of a model of a mechanism reflects those elements of the model that correspond to the parts, activities, or organization of a mechanism. Thanks to Eric Hochstein for clarifying this point.

is going on in the target mechanism, one correspondingly improves the quality of the explanation. (Kaplan [2011], p. 347).

One might reasonably wonder how one could possibly counterbalance the weight this passage places on the side of accumulating detail.

In the next paragraph, though, Kaplan leans on the opposite side of the fulcrum:

Importantly, **3M** does not entail that only completely detailed, non-idealized models will be explanatorily adequate. **3M** is perfectly compatible with elliptical or incomplete mechanistic explanations, in which some of these mechanistic details are omitted either for reasons of computational tractability or simply because these details remain unknown. Far from requiring a perfect correspondence or isomorphic mapping between model and mechanism, **3M** requires only that some (at least one) of the variables in the model correspond to at least some (at least one) identifiable component parts and causal dependencies among components in the mechanism responsible for producing the target phenomenon. (Kaplan [2011], p. 347).

Kaplan's efforts to restore balance span two pages, in which he notes that, "**3M** allows for explanatory mechanistic models to involve various kinds of idealization or mathematical abstraction" ([2011], p 348) in precisely the sense intended by Batterman [2002] and Weisberg [2007]. He references Machamer, Darden, and Craver's [2000] idea of "mechanism schemas" (contrasted with "sketches"), which are "elliptical mechanistic explanations" in which "known details have been intentionally removed" ([2011], p 348). He closes by noting that "If **3M** entailed that explanatory models could not employ idealizations [or abstractions], very few models (if any) in computational neuroscience would satisfy this overly stringent requirement and count as explanatory" ([2011], p 348).[10]

Looking back, one finds a similar effort to restore balance in the paragraph directly after Craver and Kaplan introduce **3M**:

[10] We define schemas, sketches, and phenomenal models precisely in Section 4.

> [T]he idea of an ideally complete how-actually model, one that includes all of the relevant causes and components in a given mechanism, no matter how remote, negligible, or tiny, without abstraction or idealization, is a philosopher's fiction. Science would be strikingly inefficient and useless both for human understanding and for practical application if it dealt in such painstaking minutiae…The special sciences would be utterly paralyzed if complete how-actually explanations were the guiding objective. Yet these commonplace facts about the structure of science should not lead one to dispense with the idea that models can more or less accurately represent features of the mechanism…and that models that describe more of the relevant features of the mechanism are more complete than those that omit them. (Kaplan and Craver [2011], p. 609-10).

Craver [2006] also weighs in on the import of abstract models:

> Few if any mechanistic models provide ideally complete descriptions of a mechanism. In fact, such descriptions would include so many potential factors that they would be unwieldy for the purposes of prediction and control and utterly unilluminating to human beings. (Craver [2006], p. 360)

Indeed, mechanists have described the import of abstraction in scientific explanation in many places (see, e.g., Craver [2010], p. 588, [2014], p. 49-50, Craver [2010], p. 842; Craver and Darden [2013], p. 32-4, p. 41-50; Piccinini and Craver [2011], fn. p. 22). Most of these passages merely mention in passing the obvious fact that thought, science, modeling, and by extension, explanation require abstraction.

But proponents of mechanistic views of explanation also argue explicitly for the importance of abstraction in mechanistic explanations. For example, Craver [2007] defends the importance of explanations at multiple levels of organization in direct opposition to detail-oriented, reductionist approaches (e.g., Bickle [2003]). The argument (rehearsed below) is grounded in the causal, and so explanatory, relevance of higher-level causes. For example, he uses Marder's seminal work to demonstrate the explanatory relevance of the very network properties to which Woodward appeals in his **MDB** criticisms:

> If Prinz [and Marder] are right, neuroscientists would misunderstand the behavior of the lobster digestive system if they were to suppose that it is regulated in the stomatogastric ganglion by a particular configuration in a network of cells. If one were to focus on a particular organizational structure, one would miss a pattern in the causal structure of the digestive system that is independent of those patterns among the network realizers… ([2007], p. 223)

He concludes:

> A world viewed only at the fundamental level would be a world of gory details unfiltered by higher-level perspective ([2007], p. 259).

If one emphasizes passages in which mechanists discuss the limits of incomplete models without mentioning those in which mechanists argue against detail-oriented reductionism, as the critics have done, one offers only a skewed view of the balance the mechanists hoped to establish.

So let us address the question directly: How is it possible to balance the explanatory value of detail with the explanatory power of abstraction?

3. The Norms of Causal Explanation

The balance achieved in mechanistic views has to be understood in the context of their broader philosophical project to develop an adequate theory of scientific explanation. Two widely held desiderata on any adequate theory of scientific explanation are *demarcation* and *normativity* (see Craver [2007]; Craver and Kaplan [2011]; see also Salmon [1984]; 1989; Woodward [2003]). The theory should demarcate scientific explanation from other kinds of scientific achievement, including mere description, prediction, and confirmation. One can abandon this desideratum only on pain of ceasing to think of explanation as a distinctive kind of scientific achievement. A theory of explanation should also clarify the norms for evaluating explanations. One can relinquish this desideratum only on pain of abolishing any clear idea of what makes an explanation good or bad. These desiderata are related: explanations are distinctive achievements because they satisfy distinctive norms.

At least many mechanists, starting with Coffa [1974], Salmon [1984, 1989], and including Craver [2007], [2014], believe one cannot satisfy these desiderata without taking a stance on the kinds of worldly (i.e., ontic) relations that a putative explanation must reveal to count as explanatory. The discussion of scientific explanation in the wake of Hempel's covering-law model can be read as an effort to establish the appropriate ontic relations that give explanations their force; universal regularities, laws of nature, statistical relations, and causal relations all have been considered (Cartwright 2006). Mechanists have tended to emphasize causal and constitutive relations as the source of explanatory power for a broad class of scientific explanations.

Consider the familiar case of the flagpole and the shadow (Bromberger [1966]; Salmon [1989]). The flagpole's height explains the shadow's length, but the shadow's length does not explain the flagpole's height. Salmon [1984] used this and other examples to argue that such explanations must track the causal structure of a system: the shadow does not cause and so does not explain the flagpole's height. This example illustrates the asymmetry of explanation, a core norm concerning the permissible directions of scientific explanation.

Everywhere we look there are flagpoles and shadows: places where the casual or otherwise ontic structure of a system imposes a directionality on permissible explanations. For example, dynamical models (i.e., models that use differential or difference equations to describe how variables in a system change over time) can be used to describe the interaction of causally relevant components in a mechanism. However, as with the flagpole, we can build accurate dynamical models that fail to describe the causal – and so explanatory – organization of a system (Kaplan and Craver [2011]). For example, suppose a motor turns a gear, *X*, that is connected to three other independent gears, *A*, *B*, and *C*. One might generate differential equations that allow one to predict the rotational acceleration of *A* on the basis of the accelerations of *B* or *C*, but their accelerations are each produced independently by, and so explained by, *X*. Not all dynamical models describe causal relations. Explanatory dynamical models do (Kaplan [2015]).

According to mechanists, accounts of scientific explanation that emphasize epistemic factors (e.g., description, prediction, and confirmation) at the expense of ontic constraints on acceptable explanations (e.g., Hempel 1965) deliver the wrong verdict in these and other test cases. Such views count manifestly bad explanations as good and, as a result, fail to demarcate explanation from other forms of scientific achievement. They are therefore unhelpful for understanding a distinctive component of how mechanistic explanations work.[11]

4. The Norms of Constitutive Explanation

A central novelty in recent work on mechanistic explanation has been its focus on constitutive mechanistic explanations rather than etiological explanations (the primary focus of, e.g., Salmon [1984]).[12] Etiological explanations reveal the antecedent events that cause the explanandum phenomenon. Constitutive explanations, in contrast, reveal the organized activities of and interactions among parts that underlie, or constitute, the explanandum phenomenon. More specifically, they describe features of the mechanism for a phenomenon, where the mechanism includes the set of all and only the entities, activities and organizational features relevant to that phenomenon.

Perhaps the simplest norm mechanists have embraced concerning constitutive explanations is that a phenomenon cannot be explained by merely redescribing it. Some have called such models (i.e., models that describe but do not explain a phenomenon) "phenomenal models" (Craver [2006]; Kaplan [2011]). Bunge [1964] describes them as "black box" models that describe the behavioral profile of a target system as "units devoid of structure", i.e., without positing intervening variables ([1964], p. 235). Craver [2006] describes phenomenal models as "complete black boxes" that "reveal nothing about the underlying mechanisms and so merely 'save the phenomenon' to be explained" ([2006], p. 360). Mauk [2000], a neuroscientist, defines them as models that produce "the correct input/output behavior with no claims about biological validity" ([2000], p. 649; see also, e.g., Woodward [2003], pp. 5-6). More explicitly, let us say that a model, M, is a phenomenal model of a mechanism if and only if it describes the inputs to, modulators of, and outputs from a mechanism without describing its relevant internal causal structure. All these authors agree that phenomenal models can represent a phenomenon accurately while undertaking no additional commitments about which of the many possible constitutive mechanisms for the phenomenon in fact underlies it.

[11] We know many of our critics question these assumptions about the nature of scientific explanation. Our repeated point has been that our critics have the burden of providing a model of explanation that does not, as Hempel's did, count as explanations things that no scientist would accept as such. We do not intend to be dogmatic or hegemonic, as we are often described. We simply think that an adequate philosophical theory of scientific explanation should not count manifestly bad explanations as good. Such a theory fails the demarcation and normativity desiderata.

[12] For a detailed discussion of differences between etiological and constitutive explanations, see Kaiser and Krickel [2016].

Snell's law is a phenomenal model: it describes a causal, input-output relationship between angles of incidence and refractive indices, on the one hand, and the angle of refraction on the other. Snell's law describes the antecedent causes of the angle of refraction, and so provides information about its etiological explanation. But it does not explain why light bends in these conditions. Snell's law is purely phenomenal with respect to the constitutive explanation. Our first norm of constitutive explanation is that phenomenal models lack explanatory force with respect to the phenomena they describe ("No phenomenal explanations"):

NPE: No purely phenomenal model constitutively explains the phenomenon it describes.

Crucially, **NPE** does not claim that phenomenal models are non-explanatory *simpliciter*. Instead, it claims that a given model is explanatorily vacuous as a constitutive explanation if it merely redescribes the phenomenon in general terms (for defenses of **NPE**, see Craver [2006]; Craver [2007], Ch. 4; Kaplan and Craver [2011]; Kaplan [2011]). The model might convey etiological information about how an effect was caused, but it does not convey constitutive information about how the cause produces the effect.

Considerable confusion in discussions of scientific explanation can result from failing to keep track of the explanandum phenomenon. Snell's law, as a causal law, does express explanatorily relevant information about the angle of light's refraction. But it does not explain why light refracts. To be maximally precise about a given explanandum, mechanists have tended to express it in contrastive terms. If we want to explain why Socrates died (vs. remained alive), then the relevant factor is his consumption of hemlock (vs. some non-poisonous beverage). If we want to explain why he died so quickly (vs. at some other rate), then it will matter how quickly he consumed the hemlock (Dretske [1977]; Hitchcock [1996], [1999]). Even subtle changes in the explanandum contrast can dramatically change which factors are explanatorily relevant. Mechanists often characterize phenomena such as "working memory" or "the action potential" as multifaceted. These construct-terms are shorthand for a host of features, each of which must be explained to explain the multifaceted phenomenon in its entirety. For the action potential, for example, we might try to explain why the membrane depolarizes to +40 mv rather than stopping at 0 mv (as Bernstein hypothesized). The elliptical etiological answer is that this value approaches the equilibrium potential for sodium. Why do neurons exhibit a refractory period in which they are unable to fire right after an action potential? The answer will appeal to the fact that sodium channels remain inactivated and potassium channels remain activated during this interval.[13] Crucially, different aspects of the phenomenon (precisely specified with the aid of contrastive formulation and presumed background) have different relevant parts, activities, and organizational features.

The **3M** requirement complements **NPE**; it states that a constitutive model must describe some relevant details about the mechanism for the phenomenon. Using the contrastive formulation of the explanandum, **3M** can be reformulated:

> **3M***: A constitutive mechanistic model has explanatory force for phenomenon P vs. P' iff (*a*) at least some of its variables refer to internal[14] details relevant to P vs. P', and (*b*) the dependencies posited among the variables refer causal dependencies among those variables (and between them and the inputs and outputs definitive of the phenomenon) relevant to P vs. P'. [15]

3M* clearly implies the thesis that some details (about the features of the constitutive mechanism) are necessary for explaining the phenomenon ("Some Details are Necessary"):

> **SDN**: A putative constitutive explanation for P vs. P' has explanatory force for P vs. P' only if it describes some of the entities, activities and organizational features relevant to P vs. P'.

But **SDN** does not entail **MDB**. **SDN** allows that even models describing only very abstract details (e.g., that interactions are local or obey conservation laws; cf. (Batterman and Rice

[13] Some confusion entered philosophical discussion of the action potential precisely because it is not always clear what the appropriate contrast is: the action potential vs. what? No action potential? An action potential peaking at 0 mV? An action potential with no refractory period? The right answer, absent further specification, appears to be "everything we want to explain about action potentials".
[14] Our reference to internal details is consistent with the obvious fact that often factors outside a mechanism are (etiologically) explanatorily relevant to a phenomenon. Here, we focus on constitutive explanations, in which we are revealing the causal structure between the input and the output. Details are "internal" in this sense.
[15] **3M** does not specify when or how models "refer to" the relevant features. This depends on the intentions of the modeler. Hodgkin and Huxley explicitly deny commitment to the concrete details of their how-possibly models of ionic conductance changes. The conductance equations can be seen with hindsight as satisfying **3M** *for us* because we embrace commitments that Hodgkin and Huxley justifiably eschewed.

[2014])) can count as explanatory. The spirit of **SDN**, **3M** (and **3M***), and **NPE** is to insist that the constitutive explanation for a phenomenon (characterized in terms of the instantiation of an input-output mapping vs. some other input-output mapping) must cite at least one feature of the causal structure between the input and the output (as emphasized in Kaplan [2011], 347-348).[16] Critically, neither **3M*** nor **SDN** specify how to compare two constitutive explanations. So neither entails or implies **MDB**.

As noted above, much of the controversy over **MDB** began in discussions of the Hodgkin and Huxley (HH) model of the action potential (Hodgkin and Huxley [1952]). The HH model is composed of a family of equations. Its centerpiece is the "total current equation." The total current equation describes the total current crossing the membrane as the sum of four component currents: a capacitive current, a potassium current, a sodium current, and a leakage current. These currents change independently as a function of membrane voltage and conductance. The total current is therefore explained by details about the relevant component variables and their causal interactions. But if we ask why the membrane conductance for a particular ion species changes with voltage, Hodgkin and Huxley offer only a phenomenal model, as they themselves emphasize (Hodgkin and Huxley [1952]). The equations for the individual ionic conductance changes in the membrane merely summarize the data obtained in their voltage clamp experiments. In these experiments, Hodgkin and Huxley stepped the membrane voltage to different levels but prevented the cell's natural feedback response by injecting current, thereby "clamping" the voltage in place. From the current flow required to keep the membrane voltage constant, they inferred the conductance of the membrane to that ion species at that voltage. The conductance equations are curves fit to these results. One can simultaneously hold that these empirically fitted conductance equations contain copious information relevant to explaining action potentials and that they are phenomenal with respect to the conductance changes. The arguments in Craver [2006] support only the latter claim. So the considered view (see Craver [2007]) is that the HH model satisfies **NPE** as an explanation of action potentials, but the conductance equations in that model do not satisfy **NPE** as an explanation of conductance changes. Instead, they are phenomenal models, as Snell's law is a model of bending light (see note 6).

[16] As noted above, these requirements allow for etiological and contextual explanations of the phenomenon in terms of its antecedent or environmental causes. But here we are focused on constitutive explanation.

The conductance equations, as Hodgkin and Huxley insist, are unopened black boxes – phenomenal models – contained within a *mechanism sketch*. A model, M, is a mechanism sketch iff it describes some of the internal details of the mechanism but has black boxes signifying that one or more relevant component parts, activities, and organizational features are unknown. A model, M, is a *mechanism schema* iff an abstract model in which the details are known but omitted, because they are either irrelevant or because they obscure with busy detail the causal relations we take to be central (Machamer, Darden and Craver [2000]). Hodgkin and Huxley total current equation, even supplemented with their copious background knowledge of its conditions of application, has unfilled black boxes standing in for a mechanism to be named later. They explicitly point out, to their frustration, that the conductance equations do not explain the conductance changes. The reason: they are phenomenal models of the conductance changes; i.e., they do not satisfy **3M* or SDN**. Notice, for example, that Hodgkin and Huxley were powerless to explain the refractory period of the action potential. That requires an understanding of the inactivation mechanism in sodium channels, and Hodgkin and Huxley did not know (or even strongly believe) that membranes have ion channels.

Contrary to the critics' assumptions, **3M*** neither entails nor supports **MDB**. **3M*** places no restrictions on the use of abstraction in models. **3M*** and the definition of a phenomenal model entail that phenomenal models lack explanatory power with respect to the phenomenon they describe (**NPE**). **3M*** and the definition of a mechanism sketch jointly entail that sketches have explanatory power. However, because sketches omit features of the mechanism relevant to P vs. P', they are, by definition, incomplete descriptions of the mechanism for P vs. P'. Together, these commitments (**NPE**, **SDN**, and **3M***) express only the uncontroversial thesis that constitutive mechanistic explanations gain their force by revealing features of the underlying mechanism. These commitments are silent about whether adding details to a model always makes it a better explanation.

5. Salmon-Completeness

So we must look elsewhere for a comparative norm of mechanistic explanatory completeness. As a first step, we clarify an absolute notion of "completeness" against which progress toward more complete explanations might be assessed.

Confusion sometimes arises on this matter from a failure to distinguish two ways of using the term "explanation." For many critics of mechanistic theories of explanation (and some of its defenders, e.g., Bechtel [2008]), the term "explanation" refers to a model or representation humans use to convey understanding. This is a common way of speaking. Yet English speakers also use the word "explanation" to refer not to a model but to the thing in the world that "explains" (in an ontic sense) the phenomenon in question. The claim that global warming explains the rise in sea levels is not about a model or a representation; it is about the rise in mean temperatures and its causal relationship to sea levels. From this ontic perspective, models are tools for representing objective explanations (Craver [2014]). The complete constitutive explanation for a given explanandum, in this ontic sense, includes everything relevant (i.e., everything that makes a difference) to the precise phenomenon in question.[17]

We expect this ontic orientation naturally comports with how many scientists think about models and explanations: that there is a world independent of our models, and that the completeness of explanatory models is assessed relative to that. We will not argue for this thesis; we merely acknowledge it as a background assumption and possible locus of confusion and disagreement (see Coffa [1974]). But note: The concept of an ideally complete explanation would appear to be conceptually required to speak meaningfully of abstraction in the first place. After all, abstraction is commonly defined as the intentional dropping of details, and this idea makes sense only if there is something from which details are dropped.

In our view, the causal structure of the world defines the limit of completeness in one's explanatory knowledge. We call this ontic notion of completeness, *Salmon-completeness*:

[17] Notice that this view is consistent with the possibility that higher-level causes can "screen off" lower-level components. Once the higher-level component is fixed, and the relevant background assumptions of the request for explanation are made explicit, differences among the lower level components no longer make an additional difference to the explanandum phenomenon. In Woodward's terms, the lower-level parts can be irrelevant conditional on the behavior of higher-level components (Woodward, under review). For example, once we know the sodium current across the membrane, the precise locations of the individual sodium ions are irrelevant. Likewise, it does not matter which of the thousands upon thousands of sodium channels do and do not open. This is why the total current equation can explain current in terms of conductance changes, while bracketing future knowledge of precisely how these changes are brought about. These low-level differences make no relevant difference once the higher-level behavior is fixed. One consequence of this is that a complete explanation for a properly specified explanandum phenomenon, in other words, need not (and typically does not) end in quarks (for further discussion, see Craver 2007). Which of the various multilevel relevance relationships happens to be screened off depends on the contrastive specification of the explanandum.

SC: The Salmon-complete constitutive mechanism for P vs. P' is the set of all and only the factors constitutively relevant to P vs. P'.

We understand causal relevance precisely as Woodward [2003] does: Causally relevant factors are those that we could intervene upon (ideally) to change the explanandum phenomenon (see also Craver [2007], Ch. 3; Craver and Kaplan [2011]). Mechanists such as Craver have extended Woodward's view of causal relevance to analyze constitutive relevance in terms of "mutual manipulability": A factor is constitutively relevant when (ideal) interventions on putative component parts can be used to change the explanandum phenomenon as a whole and, conversely, interventions on the explanandum phenomenon as a whole can produce changes in the component parts (for details, see Craver [2007], Ch. 4)[18] As noted above, constitutive mechanisms are defined by considerations of relevance: all entities, activities, and features in a mechanism are, by definition, relevant to the mechanism's behavior. That is, no constitutive mechanism for a phenomenon contains entities, activities, or features irrelevant to that phenomenon.

Notice that Salmon-completeness simply defines the worldly limit of completeness. It does not include, entail, or imply the further necessary condition that a model has to describe all the relevant details to have explanatory force. Indeed, we deny the "All Details are Necessary" thesis:

ADN: A putative constitutive explanation for P vs. P' has explanatory force for P vs. P' only if it describes all of the entities, activities and organizational features relevant to P vs. P'.

ADN is ridiculously stringent. **SC**, in contrast, merely provides a clear and objective endgame in terms of which to assess explanatory completeness.

6. From More Details to More Relevant Details

[18] There may be other (perhaps better) ways of understanding interlevel relevance relationships on a manipulationist or some other approach. One version of the manipulationist approach that seeks to reduce constitutive relevance to a set of causal relevance relations can be found in Prychytko (under review). Also see note 17 for a way of bracketing this notion.

Let us now turn to the comparative claim in MDB. Unfortunately, none of the critics defines **MDB** precisely. And no contemporary mechanist has ever articulated a "more details better" thesis or claimed, without qualification, that more detailed models are better than less detailed models. So to build a plausible statement of **MDB** (i.e., one worth taking seriously), we must start from scratch. Gradual refinement of this comparative thesis will reveal and justify norms of completeness that mechanists ought to embrace.

To start with the crudest formulation, one might interpret **MDB** as the thesis that more detailed models are always better than less detailed models for a given explanandum phenomenon, full stop. We call this **unqualified MDB:**

> **$\mathbf{MDB_u}$**: If model M contains more details than model M*, then M has more explanatory force than M* with respect to phenomenon P vs. P'.

Chirimuuta ([2014]) appears to target **$\mathbf{MDB_u}$**. She charges that the mechanistic emphasis on detail expresses and justifies the focus of "big data" projects on compiling all the details about a given system. Such projects are fueled by high-throughput data collection methods, the exponential growth of computational power, and the development of corresponding tools for organizing and analyzing these massive data sets (Sejnowski, Churchland, and Movshon [2014]):

> Some high-profile research programs in neuroscience do conform to the **MDB** methodological prescription. For example, the Human Brain Project is a new Europe-wide flagship which has the stated goal of simulating the entire human brain in a supercomputer...Such ideas are obviously animated by the idea that building computational models which simulate the actual neural mechanisms with as much detail and accuracy as currently possible will be an invaluable means to link neural circuits to behavior…. (Chirimuuta [2014]).

Just as no mechanist has voiced **MDB**, no mechanist has explicitly embraced the Human Brain Project or suggested that such projects are more explanatorily powerful than those that yield less detailed models. And there is good reason for this.

First, the Human Brain Project is a project, not an explanation. The goal of that project is to understand the structural connectivity and organization of the entire brain. The purview of this project is thus the anatomical basis of everything brains do. The information fed into and generated by such projects is undeniably useful for explaining how brains work, but "things brains do" is not a well-defined explanandum.

Second, and relatedly, $\mathbf{MDB_u}$ makes no reference to explanatory relevance, a central feature of mechanistic accounts of constitutive explanation. As noted above, mechanists targeted by **MDB**-based critiques have embraced Woodward's manipulationist view of causal relevance (Craver [2007], Ch. 3; Kaplan [2011]; Kaplan and Craver [2011]) and extended it to provide a condition on constitutive explanatory relevance (Craver [2007]). A reasonable complaint facing the Human Brain Project and other such projects is that they assemble details without any regard to which details are explanatorily relevant to a given explanandum phenomenon and which are not. If one wants to explain the limits of working memory, a map of every synapse in the brain would contain every synapse relevant to working memory, but those synapses will be submerged in an ocean of detail irrelevant to working memory, let alone, its span. Mechanists do not deny the importance of the information compiled by such a project; rather, they deny that it has been filtered by relevance to a given explanandum phenomenon.

Curiously, none of the **MDB** critics discusses the mechanist's views on explanatory relevance. This oversight is largely responsible for their lopsided view of the balance defenders of mechanism hoped to achieve. For if mechanists have embraced anything like **MDB**, it is not $\mathbf{MDB_u}$ but rather something like "More *Relevant* Details Are Better"[19]:

[19] Woodward [In press] acknowledges in a footnote that $\mathbf{MDB_r}$ is a more appropriate expression of the mechanists' view than $\mathbf{MDB_u}$, but the rest of the paper attacks something closer to $\mathbf{MDB_u}$. We suspect some of our critics, including most notably Woodward, assume that mechanists celebrate features of "distinctively mechanistic" explanation beyond that supplied by manipulationist views. This tendency likely has its roots in the historical fact that some mechanists (Machamer [private communication]; Bogen [2004], and sometimes Darden [2006]) have tended to pit ideas about mechanisms against counterfactual theories of causation and explanation. Craver and Kaplan, in contrast, explicitly embrace Woodward's theory of causal relevance as a basis for their view of mechanisms (Craver [2007]; Kaplan [2011]; Kaplan and Craver [2011]). Note again that components of a mechanism are relevant to the explanandum phenomenon *by definition*: the constitutive mechanism for P vs. P' contains all and only the features constitutively relevant to P vs. P'.

MDB_r: If model M contains more explanatorily relevant details than M* about the SC mechanism for P vs. P', then M has more explanatory force than M* for P vs. P', all things equal.[20]

Bracketing for the moment the vexing "all things equal" clause, **MDB_r** is significantly closer to the norm of completeness mechanist's have intended to express than is **MDB_u**. First, **MDB_r** relativizes the details in the Salmon-complete mechanism to an explanandum phenomenon specified by the contrast, P vs. P'. As noted above, whether something is or is not in the mechanism for a phenomenon depends definitionally on its relevance to that phenomenon.

Relatedly, **MDB_r** refers to a relevance relation. This is presumably why Kaplan [2011], in the apparently damning quote cited in Section 2, claims that models gain explanatory power when they describe more "mechanistically *relevant* details." When Chirimuuta claims that mechanists insist on including irrelevant details, she ignores their views on explanatory relevance. When Levy [2013], p. 479) claims that "[a]n abstraction does not, as such, call for filling in. Often the opposite: scientists abstract because they believe that detail is unnecessary and irrelevant and that including it would impede understanding…", he does not consider whether mechanistic accounts of relevance rule out the irrelevancies he has in mind. And when Woodward [2013] argues that abstract network properties are explanatorily relevant, he does not

[20] Woodward [personal communication] suggests that critics read mechanists as failing to relativize explanations to the explanandum. Yet it has been an explicit component of thinking about mechanisms, starting with Salmon, but also in Kaufman [1974] and Wimsatt [1986], and Machamer, Darden and Craver [2000] that "mechanisms are the mechanisms of the things they do." (Indeed, some mechanists have dubbed this central commitment "Glennan's Law"). The contrastive formulation is explicitly introduced in Craver [2007] to sharpen this view of explanatory relevance.

As noted in footnote 6, some confusion arises from the fact that "the action potential" is not explicitly contrastively formulated, and the intended contrast is likely vague and multifaceted, leading to different ideas of "completeness" for those who focus on different aspects of the same phenomenon. Hodgkin and Huxley's 1952 paper makes considerably more sense as an instance of inference to the best explanation used in defense of the membrane hypothesis (i.e., that action potentials are produced by ionic fluxes across the membrane rather than, for example, chemical reactions or mechanical forces) rather than as an effort to construct an explanatory model. Hodgkin and Huxley used the predictive power of their mathematical model as an argument for the explanatory power of the membrane theory, a theory whose content extends well beyond the content of the total current equation. The phenomena "covered" by the model do not exhaust all aspects of the action potential. They are simply some of the main features that the mechanism of the action potential would have to explain: the rising phase, the peak amplitude, the refractory period, etc. The fact that the total current equation accurately predicts these features when the values for concentration and conductance change are set to empirically determined values argues persuasively that action potentials are produced by the coordinated flux of ions across the membrane. Seen in that light, the total current equation itself systematizes several key commitments of the explanation (as Craver has argued), but only in order to argue for the empirical plausibility of the membrane hypothesis. It would be a mistake to confuse this model, or this persuasive use of the model, for the explanation itself. See our discussion below.

consider mechanist arguments for the explanatory relevance of network properties. The view appears unbalanced precisely because the critics have neglected a central component of mechanistic theories of explanation.

In fact, the same mechanists targeted by the **MDB** critics have argued that explanations are universally made *worse* by adding irrelevant details (e.g., Craver [2007], Ch. 3). This is the point of many classic examples: e.g., hexed salts, and men who take birth control pills. Salmon [1984]; [1989] used such examples to argue for a crucial difference between the norms of explanation and the norms of argument. Adding irrelevant detail to a valid argument leaves a valid argument; irrelevant details always blemish explanations.

Finally, $\mathbf{MDB_r}$ is neutral about levels and abstraction. While $\mathbf{MDB_r}$ claims that a model with more relevant lower-level details is more complete (all things equal) than a mechanism that omits them, it also suggests that a model that leaves out higher-level (i.e., abstract) details is less complete (all things equal) than one that includes them.[21] $\mathbf{MDB_r}$ thus fits well with the mechanists' repeated insistence on *multiple* levels of explanation and their arguments for the relevance of higher-level causes. For example, Craver [2007]; Ch. 6] follows Yablo [1992], Marcus [2001], and Woodward [2003], in arguing that experiments often demand appeal to (i.e., reveal the existence of) higher-level (so necessarily abstract) features that are causally and so explanatorily relevant for a given phenomenon. Again, according to his view, one has correctly partitioned the space of causes when one has identified the (appropriate, proportionate) switch-points corresponding to the explanandum contrast P vs. P'. Different explanandum contrasts are explained by different switch-points. Just as Socrates' drinking hemlock (vs. wine) explains his being dead (vs. alive), being below zero °C (vs. above zero °C) explains why water is frozen. Changing the temperature of the water from -4 to -17 °C does nothing. So the claim that the water is frozen because the room is -17 °C (while not entirely incorrect) fails to identify the relevant switch-point for its freezing (vs. thawing), though it does identify a relevant switch-point for the rate at which it freezes. This method for establishing explanatory relevance is familiar in science and has been a centerpiece of the mechanist's arguments against detail-oriented reductionism. The goal is to identify the "differences that make a difference." In constitutive explanations, the details are by definition lower-level, but exactly how many levels

[21] The reader may find the notion of "higher-level details or "abstract details" oxymoronic, assuming that details are always concrete. If so, simply replace "details" with "facts about how the mechanism works".

there are and which levels are crucial for a given effect contrast, cannot be determined in advance of empirical enquiry.

Sadly, we do not know how best to dispatch the vexing "all things equal" clause in MDB_r. There is no problem if the set of relevant details in M* is a proper subset of the set of relevant details in M, but that is a special case. When models do not stand in this subset relation, any comparative assessment requires a weighting of the relative importance of different, explanatorily relevant details. Both the fuzzy dice and the engine are relevant to the car's acceleration, but the engine seems to be more explanatorily central. Sometimes, as in this case, we assess the magnitude of the difference the causal variable makes to the effect size. Sometimes, we draw a line between foreground and background conditions (though background conditions often make a big difference). Sometimes, we weight a given factor by whether it acts like an on/off switch or like a dimmer switch (i.e., allowing for discrete versus continuous control, respectively). Sometimes we measure the magnitude of the difference a change in the variable makes to the variance observed in a population or to the strength of the correlation between the variable and the phenomenon. Given that these different objective measures yield different weightings, perhaps the best we can hope for is a list of the factors to which one might appeal in making such judgments. Of course, pragmatic factors (i.e., who we are talking to or what we plan to do with the model) often figure in our assessments of which features are central or peripheral and in our assessment of which of the above measures is most appropriate. A pit boss might see the driver's fuzzy dice as a wasteful addition to the mass of the car. An historian following trends in Formula 1 car acceleration strategies can safely neglect the dice. We have little of general interest to say about such considerations, and we welcome progress on how they might be dispensed objectively.

7. Non-Explanatory Virtues of Abstraction

Everyone agrees that irrelevant detail makes explanations worse. But perhaps the **MDB** critics would also object to **MDB_r**. Perhaps they believe explanations are sometimes improved by filtering out even relevant details to identify the core (most highly weighted) details. Concerning the HH model, Levy writes:

> The question is whether there is a distinctive explanatory payoff to a model that omits… mechanistic information. I have argued that there is: abstracting from channel structure allowed HH to depict whole-cell behavior as an aggregate of discrete, independent voltage gates at the molecular level. (Levy [2013], p. 488)

Levy neither explains why he thinks this further filtration constitutes a distinctively *explanatory* payoff nor shows how this weighting among relevant factors is determined.[22]

One plausible justification for Levy's claim is that less detailed models are more intelligible to human cognitive agents than are more detailed models. This psychological hypothesis is perhaps true, but it is irrelevant to our normative question. There are two questions: (1) What is the complete explanation for P vs. P'? and (2) What is the most complete explanation for P vs. P' the unaided human mind can grasp? We address the first.

Another payoff of abstraction is that it makes models more computationally tractable. This is part of what Sejnowski describes in a passage cited by Chirimuuta ([2014], p. 134):

> To make progress in the computational modelling of biological systems, abstraction of the elements of the system will be required. Without abstraction, the resulting models will be too complex to understand and it will not be feasible to carry out the desired computations even using the fastest computers (Sejnowski, [1988], p. 1300).

Abstraction is obviously important for computational modelers (Herz et al. [2006]). But again there are two questions: (1) What is the complete explanation for P vs. P'? and (2) How can a given explanation be coded most efficiently for simulation on a digital computer? We address only the former.

Sometimes pragmatic considerations favor abstracting even from relevant details. In different explanatory or practical contexts, different parts of the Salmon-complete mechanism will be especially salient. Communicative explanations must be matched to the background knowledge of an audience, for example. And different amounts of detail might be required depending on whether one is teaching undergraduates or performing brain surgery. Again there

[22] He also does not explain why attention to the gating mechanisms in different ion channels obscures our understanding of their "discreteness" and "independence". Yet, surely, exactly the opposite is true.

are two questions: (1) What is the complete explanation for P vs. P'? and (2) How should the explanation be modeled to best achieve one's communicative or pragmatic ends? For reasons discussed above, we doubt there is a general answer to the latter question, and so have focused on the first.

Finally, some of the **MDB** critics emphasize that abstract models have broader scope than detailed models, on the assumption that dropping details increases the range of cases to which the model might usefully be applied. However, decreasing detail (increasing the degree of abstraction) of a model does not always increase its scope; the extent to which detail and scope correspond depends on the nature of the target system being modeled (see Matthewson and Weisberg 2009). The assumption behind this argument for the explanatory value of abstract models appears to be that the more abstract a model is (i.e., the more generally it applies to a range of apparently disparate cases), the more explanatory it is.

We reject this "More General is Better" thesis implicit in much of the critics' work: i.e., the thesis that explanations are always improved by expanding their scope. An explanatory model that applies only to a single, rare strain of fly is not less explanatory than a model that applies to a ubiquitous strain simply in virtue of the fly's rarity (for a similar view, see Hitchcock and Woodward 2003; Woodward 2003). The model explains fewer tokens, but it explains each of them, we suppose, perfectly well. To come at it from the other side, any generality a model has beyond the intended scope is explanatorily gratuitous: it carries no explanatory force beyond the domain of application. If one is trying to explain odor discrimination in the fruit fly, one will want a model that applies to the fly. Should it turn out that the same model also works for primates, that will be a further interesting and important fact, but it will not improve on the explanation we had for the fly (see Kaplan [In press]). Similarly, our understanding of fluid dynamics in water (Batterman and Rice [2015]) is not diminished one bit by the fact that the Navier-Stokes equation fails to apply to anisotopic liquid crystals (their shape foils one of the main assumptions of the model) (Povich [2015]). Nor is our appreciation of water flow necessarily improved by building a more general model of fluid that covers anisotopic liquid crystals as well. Our question is not "Which model applies to more phenomena?" but rather, "What is a complete explanation for a phenomenon?"

Biology and neuroscience textbooks typically contain families of models with varying degrees of scope describing different biological processes at different levels of organization

(Schaffner 1993). One can have highly general models of molecular interactions and highly specific models of the how particular allosteric interactions takes place. These models are "tuned to" different explanatory and practical needs. Students master overlapping models of "wild type" organisms and several families of "mutants". They learn about kidneys generally, and about the idiosyncrasies of frog kidneys, or shark kidneys, or monkey kidneys. The truly knowledgeable scientist, one with deep explanatory knowledge of the field, knows when the general model works and when the "mutant" models have to be called into play.

The appropriate grain of the explanans depends on how the explanandum is specified and on the relevant natural (causal and constitutive) dependencies: i.e., which changes to which factors account for the explanandum. The explanandum might be formulated abstractly or in fine detail (action potentials, or action potentials in the squid giant axon); it might range over very general types, tokens, and any range of scope between these extremes. As the explanandum phenomenon becomes more abstract, it necessarily drops details to smooth over variation in the scope of the explanandum. This can make the difference to which variables turn out, on manipulationist grounds, to be relevant to the phenomenon thus specified.

To sum up, everyone is (or should be) in agreement that abstraction serves psychological, computational, communicative, and other practical functions. Everyone recognizes the utility of highly general models. But there does appear to be some disagreement about whether these virtues are explanatory virtues. To one who simply insists that a model satisfying these virtues is, ipso facto, explanatory, we can only point out that there are models that satisfy these norms but fail to count as good scientific explanations. The norms of explanation are more restrictive than those that must be satisfied in psychological, computational, communicative, practical, and projectional uses of models. Finally, and most importantly, everyone in this discussion agrees (or should) that higher-level and abstract features of a mechanism can be causally and so explanatorily relevant. There are abstract patterns[23] in the causal structure of the world. Abstraction is not always used to mask our ignorance (as in unacknowledged sketches). If we agree on all this, perhaps we are on the road to a new consensus.

8. From Explanatory Models to Explanatory Knowledge

[23] See Dennett [1991] and Haugeland [1998].

Return now to $\mathbf{MDB_r}$. Is $\mathbf{MDB_r}$, with further allowance for the benefits of abstraction discussed in section 7, an adequate account of the norms of comparative completeness for constitutive mechanistic explanations? Despite its advantages over $\mathbf{MDB_u}$, $\mathbf{MDB_r}$ nonetheless errs (in our view) in treating models as the proper focus of norms of explanatory completeness. Models often serve their diverse ends in science without including all the relevant details about a mechanism; they are not the appropriate focus for discussions of norms of *explanatory* completeness.

Some **MDB** critics reasonably start from the assumption that all explanations are models or representations (Churchland [1989]; Bechtel and Abrahamson [2005]; Potochnik [2015]). If knowledge requires representation, and all representations are models, then all explanatory knowledge requires models. Furthermore, all actual scientific explanations seem to come in the form of models (Bechtel [2008]). In our view, however, the critics are guilty of equivocating between two senses of explanation introduced at the start of Section 5.

We distinguish models from explanations, first, because there are models of phenomena that are in no way intended to explain them. Phenomenal models do not explain the phenomena they describe. Furthermore, models can be used to design experiments, to describe and manipulate data, and to assist in prediction (Bogen [2005]; Downes [2011]). Even if all explanations are models, only a subset of the models explain anything at all; explanations are distinguished in this set by the norms they must satisfy to count as such.[24] In our view, those norms can be satisfied only by looking out to the world and expressing commitments about the kinds of worldly relations that count as explanatory.

We distinguish models and explanations, second, because typically many models are required to explain even minimally complex phenomena (Hochstein 2015; Morgan and Morrison 1999; Cartwright 1999). Hodgkin and Huxley's total current equation is one of many models they used to explain action potentials. They also modeled cells as isolated fluid compartments, membranes as electrical circuits, the intracellular fluid as a homogeneously distributed mixture, diffusion as a movement of charged particles, and so on. One cannot fully understand the action potential (in its multifaceted complexity) by attending to only one of these models (see, e.g., Trumpler [1997]). Furthermore, one cannot apply any of these models to real systems without

[24] We take this as the primary lesson of decades of discussion of scientific explanation (Salmon [1989]). Some **MDB** critics appear to embrace that lesson (e.g., Woodward [2003]; Batterman and Rice [2014]).

drawing on background knowledge about what the model describes and how the model applies to the world.

It would in many cases be unwieldy to create a single "über-model" that conjoins all these individual models and the relevant background knowledge. And even if we wanted to build such a thing, the component models typically make such different and mutually incompatible idealizing assumptions that their conjunction is logically incoherent (Cartwright [1999]; Hochstein [2015]). Of course, some logically coherent über-model must exist for every mechanism (lest the world be logically incoherent; Cartwright [1999]), and if one could build a useful model of this sort, it would surely be a momentous achievement (as was the HH model). But if one aims to reveal something about the nature of scientific practice (as critics, such as Bechtel, often claim to do), then perhaps it is worth emphasizing the somewhat obvious fact that scientists typically use many different models to describe basic constraints on the working of a single mechanism (cf. Craver [2007], Ch. 7) and that many models are involved, explicitly or implicitly, in most explanations. This is why discussions of the norms of explanatory completeness should focus on stores of explanatory knowledge rather than on individual models. Individual models invariably express this knowledge about the causal structure of the world only in a piecemeal, patchwork fashion.[25] A model might contain explanatory information sufficient for a given explanatory contrast, or a complete explanation might involve applying many different models at once. We use models to gesture toward explanations that invariably have more content about the causal structure of the world than any single, useful model can express.

Thus, we hesitate to embrace $\mathbf{MDB_r}$. Single explanatory models often express only a fraction of the explanatory store. Models never (or at least rarely) describe the Salmon-complete constitutive mechanism; they are always abstract, and they play their crucial roles in science because they are abstract. The SC mechanism is the limit of completeness toward a terminal point along one dimension of explanatory progress. Scientists need not reach that limit or even strive to reach that limit for it to define a spectrum of progress one might make in giving a mechanistic form of scientific explanation.

While someone might be interested in assessing how much of the Salmon-complete mechanism a given model describes, we focus instead on what it means for a science as a whole

[25] The views of Craver [2006] make more sense if one thinks of his use of "model" as referring to this über-model. Perhaps Craver [2006] was struggling to fit his ontic view into the more common, model-based way of describing explanations. We claim here that struggle was misguided.

or an individual scientist to have and work toward a complete explanatory grasp of a given phenomenon (either a given phenomenon contrast or many such contrasts conjoined, as in the case of "the action potential"). Again, for us, "complete" means *Salmon-complete*, even if we usually deal with models that are only "complete enough" (see also Craver and Darden [2013]). The idea that one can measure explanatory completeness in terms of how much a given science or scientist knows about the relevant features of a mechanism does not entail that any single model would be improved by adding more detail about that causal structure. In Section 9, we sketch a positive proposal along these lines.

9. Mechanistic Completeness Reconsidered

SC is the idea that the complete constitutive mechanism for P vs. P' includes all and only the factors constitutively relevant to P vs. P'. **SC** provides an objective benchmark against which the completeness of explanatory knowledge can be measured. Our explanatory knowledge is enriched, in this sense, when we learn more explanatorily relevant facts. Why is this a better characterization of the ideal of mechanistic completeness than those considered above?

Consider three tests for explanatory completeness. One test, which Woodward [2003] articulates, involves gauging the ability to answer more what-if-things-had-been-different (or *w*-) questions: correctly answering more questions about how the phenomenon would be altered under a variety of conditions, e.g., how the action potential would change with different sodium concentrations or at different temperatures. A crucial determinant of the number of w-questions, of course, is the precision with which the intended explanandum contrast is formulated. It is also determined by how the causal and constitutive relationships in the world turn out to be arranged: Whether the underlying details do or do not make a relevant difference to the contrastively specified phenomenon you are trying to explain, whether higher-level causes screen off lower level differences. Not all scientists care equally about all the *w*-questions one might ask about a given phenomenon, of course, but a scientist who can answer more of them (for a given explanandum contrast) knows more about how the mechanism works than a scientist who can answer fewer of them (*ceteris paribus*). The same idea scales up for science as a whole.

A second test for detecting limits in one's explanatory knowledge focuses on the depth of one's understanding of a hierarchy of mechanisms. Here we focus on how-does-that-work (*h*-) questions. Though Woodward does not discuss them specifically, *h*-questions are a subset of *w*-

questions of particular relevance for constitutive explanations (where Woodward's primary focus has been etiological explanation). Hodgkin and Huxley could only go so deep in their answers to *h*-questions. Specifically, they did not understand how membranes change conductances. Such knowledge is relevant to understanding why tetrodotoxin can eliminate action potentials and why the neuron is refractory during the after-hyperpolarization phase of the action potential. If these features are included in one's multifaceted characterization of the action potential phenomenon (see Craver [2007], Ch. 4), then a body of knowledge that fails to explain this leaves something out. We suspect no electrophysiologist would deny that explaining the effect of tetrodotoxin is an important constraint on constitutive explanations of the action potential precisely because it is taken to be a core, defining feature of the action potential phenomenon. If scientists do not understand that, there is more to be learned. Such knowledge also helps to understand why mutations in certain genes produce the symptoms of the family of sodium channel disorders. The more *h*-questions one can answer about a given explanandum contrast (specific or multifaceted), in our view, the richer is one's explanatory grasp of that explanandum (*ceteris paribus*).

We emphasize that these lower-level details make sense only if one knows the higher-level causal patterns they constitute. The flip side of the *h*-question is thus the *r*-question: What role does this item play in higher-level mechanisms? A scientist who knows many facts about action potentials but does not know that action potentials play a role in neural signaling is missing an important filter on which details are relevant. The components of the action potential only hang together as such because we appreciate that the action potential is a significant feature of neuronal behavior. No description of the molecular details could be organized into a causally productive unit without taking this higher-level perspective on the bewildering causal mayhem of the neuron and its components. This is presumably why mechanists like Craver insist, contra Bickle [2003], that a list of gory details unfiltered by higher-level perspective is a bad explanation.

Our norms of mechanistic explanatory completeness make sense of these tests. If one knows less than the Salmon-complete mechanism for a phenomenon, then there must exist certain *w*-questions, *h*-questions, and *r*-questions one cannot answer. If these tests are good indicators of explanatory completeness, we are justified in measuring progress relative to **SC**.

We formulate two versions of *mechanistic completeness*: one for science as a whole, and one for individual scientists. For science as a whole:

> **$MC_{science}$**: If a store of explanatory knowledge *K* contains more relevant details about the Salmon-complete cause or constitutive mechanism for P vs. P' than a store of explanatory knowledge *K**, then *K* has more explanatory power with respect to P vs. P' than does *K**, all things equal.

$MC_{science}$ is an overall measure of what we do and do not know collectively. **$MC_{individual}$**, in contrast, is intended as a measure of the completeness of an individual scientist's explanatory grasp:

> **$MC_{individual}$**: If individual scientist A knows more relevant details about the Salmon-complete cause or constitutive mechanism for P vs. P' than individual scientist B, then A has a better explanatory grasp of (i.e., wields more explanatory power over) P vs. P' than does B, all things equal.

If scientist A knows more of the relevant details that make a difference to P vs. P' than scientist B, all things equal, there are some *w*-questions, *h*-questions, and *r*-questions that A can answer and B cannot.

These formulations, unfortunately, still contain the vexing "all things equal" clause that prevents one from saying anything definite about cases in which K* is not a proper subset of K or where individual scientist A does not know every relevant thing B knows and more. If one desires a more general formulation than these provide, one should be very clear about the weighting metric. As we noted in Section 6, the different available objective weighting metrics (such as absolute contribution to the magnitude of the effect, contribution to variance, strength of correlation coefficient, switch-like versus dial-like control, foreground versus background, etc.), each of which is reasonable in some contexts, yield incompatible judgments. None of these options provides the uniquely correct metric, and each potentially has its place depending on what one aims to do with one's explanatory knowledge. In our view, appeal to pragmatics at this point is less a philosophical theory and more a naked statement that "it depends." We therefore keep the "all things equal" clause in the formulation because it emphasizes that whatever weighting might be developed, our stated standards of completeness do not recommend a foolish rush toward the multiplication of tiny details and do not entail that all details must be equally weighted. More to the point, it is an acknowledged black box in our collective understanding of scientific explanation.

Keeping that caveat in mind, we emphasize that these norms of completeness also make sense of the instrumental value of explanatory knowledge. A science or scientist that knows more relevant details will know more of the buttons and levers in a system that might be used to make it work for us. In our view, this is why explanatory knowledge is important, why the mechanistic norms of explanation are justified, and why explanatory knowledge is rightly distinguished from other forms of scientific achievement.

Before closing, we must address a very reasonable complaint often voiced in reaction to the ideas presented here: That the idea of Salmon-completeness is a rarefied, "philosopher's" ideal, beyond our epistemic reach, and so has no bearing on the actual practice of science. We disagree on all counts. Depending on one's characterization of the phenomenon (and what one choses to hold fixed in the context of the request for explanation) the relevant features in the explanation might be very few in number. For multifaceted phenomena, such as the action potential or working memory, the task is considerably more complicated: there are many aspects of the phenomenon that have to be explained, and the explanations have to be compatible with one another (even if the models are not; Hochstein [2015]; Longino [2012]). In those cases, it is all the more important to see Salmon-completeness as an epistemic ideal that drives scientific progress. We favor setting a high bar and then deciding when we are tired of jumping rather than setting a low bar and declaring an early victory. And for this very reason, we think these ideals do have bearing on the actual practice of science precisely because they drive the curious to find more and wonderful variety in the world and to develop new and powerful tools for detecting and describing it. The quest for deeper explanations, answering more *w*-, *h*- and *r*-questions fuels progress in mechanistic sciences. And for us, one dimension of explanatory depth is the completeness with which one can describe the relevant mechanistic details.

10. Conclusion

MDB in its simple form is untenable, both as an expression of the mechanistic view and as an adequate norm of explanatory completeness. Yet the **MDB** criticism has been a useful stimulus for thinking more precisely about the appropriate norms of completeness. Specifically, it has forced us to unearth and clarify the assumptions driving these discussions, and in so doing, to reveal some widespread confusions driving this rapidly expanding philosophical literature.

In an effort to clarify the philosophical landscape, we have argued in favor of **NPE**, **SDN**, **3M***, **SC**, **MC$_{science}$**, and **MC$_{individual}$**. It follows from these commitments that phenomenal models lack explanatory import for the phenomena they describe, that models gain their constitutive explanatory force by describing relevant features of the mechanism, and that if we cannot open black boxes to reveal the relevant factors for explaining P vs. P', we have a less-than-complete understanding of P vs. P'. It also follows that our explanatory knowledge of P vs. P' is enriched as we learn more about the factors (abstract and concrete) explanatorily relevant to P vs. P'. Further, it follows that if (contrary to standard scientific practice) we could combine all and only the relevant features in our explanatory store for a given explanandum phenomenon into an über-model that contains more relevant details about the causes or constituents of P vs. P' than its rivals (all things equal), then its explanatory power would be greater than that of its rivals. But the über-model is a philosopher's contrivance. Scientists deal with families of models, with different degrees of abstraction and different idealizing assumptions, each of which contributes collaboratively to the growth of our explanatory knowledge. Consequently, both **MC$_{science}$**, and **MC$_{individual}$** apply to stores of knowledge, not models.

These commitments do not entail that only complete models have explanatory force (**ADN**) or that a model with more details necessarily has more explanatory force than a model with fewer details (**MDB$_u$**). Nor do our commitments entail that abstract, detail-omitting models lack explanatory value or necessarily explain less than detailed models. In fact, they entail the negation of these claims.

Some of the confusion concerning these commitments can be blamed on the mechanist's early formulations of these ideas. In particular, defenders of mechanistic theories have tended to express the norms of mechanistic completeness in terms of models rather than stores of knowledge. Some of it arises from a general failure to keep track of the explanandum phenomenon and its implied contrasts. Some of the confusion also results from the critics' unbalanced presentation of the mechanists' views, which we have tried to correct. Finally, and perhaps most importantly, some of it results from disagreement about the philosophical problem of scientific explanation. On this matter we have simply presented the mechanist's general orientation. We hope these efforts have identified points of agreement and disagreement that will lead to more productive discussions of the norms of explanatory completeness, mechanistic or otherwise.

References

Baetu, T. M. [2015]: “The Completeness of Mechanistic Explanations.” *Philosophy of Science* **82** (5), p. 775–786.
Batterman, R. W. [2002]: Asymptotics and the role of minimal models. *British Journal for the Philosophy of Science* **53**, p. 21-38.
Batterman, R. W . [2009]: “Idealization and Modelling” *Synthese* **169**, p. 427-446.
Batterman, R. W., and Rice, C. [2014]: “Minimal Model Explanations.” *Philosophy of Science* **81** (3), p. 349–76.
Bechtel, W. [2008]: *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.
Bechtel, W. and Abrahamsen, A. [2005] “Explanation: A Mechanistic Alternative.” *Studies in the History and Philosophy of Biology and Biomedical Sciences.* 36: 421-41.
Bickle, J. [2003]: *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht, Netherlands; Springer.
Bogen, J. [2004]: “Analysing Causation: The opposite of counterfactual is factual.” *International Studies in the Philosophy of Science.* 18: p. 3-26.
Bogen, J. [2005]: “Regularities and Causality; Generalizations and Causal Explanations.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **36** (2), p. 397–420.
Boone, T. W. and Piccinini, G. [2016a]: “The Cognitive Neuroscience Revolution.” *Synthese* **193** (5), p. 1509–34.
Boone, T. W. and Piccinini, G. [2016b]: “Mechanistic Abstraction.” *Philosophy of Science*. <http://www.journals.uchicago.edu/doi/pdfplus/10.1086/687855>.
Brigandt, Ingo. [2013]: Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences* **44** (4), p. 477-492.
Bromberger, S., [1966]: “Why Questions”, in R. G. Colodney (ed.), *Mind and Cosmos*, Pittsburgh, PA: University of Pittsburgh Press.
Cartwright, N. [1999] *The Dappled World: A Study at the Boundaries of Science.* Cambridge: Cambridge University Press.
Chirimuuta, M. [2014]: “Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience.” *Synthese* **191** (2), p. 127–53.
Churchland, P.M. [1989]: *From A Neurocomputational Perspective.* Cambridge, MA: MIT Press.
Churchland, P.S. and Movshon, J. A. [2014]: “Putting Big Data to Good Use in Neuroscience.” *Nature Neuroscience* **17** (11), p.1440–41.
Coffa, J.A. [1974]: “Hempel’s Ambiguity” *Synthese* **28**, p. 141-63.
Craver, C. F. [2006]: “When Mechanistic Models Explain.” *Synthese* **153** (3), p. 355–76.
———. [2007]: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford, UK: Oxford University Press.

———. [2008]: “Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential.” *Philosophy of Science* **75** (5), p. 1022–33.

———. [2010]: “Prosthetic Models.” *Philosophy of Science* **77**, p. 840-851.

———. [2014]: “The Ontic Account of Scientific Explanation”. In Kaiser, M.I. et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*. Synthese Library, p. 27-52.

Craver, C. F., and Lindley Darden. [2013]: *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago: University of Chicago Press.

Dennett, D. C. [1991]: “Real Patterns” *Journal of Philosophy* **88**(1), p. 27-51.

Downes, S. [2011]: “Scientific Models” *Philosophy Compass* **10**, p. 757-764.

Dretske, F. I. [1977]: “Referring to Events.” *Midwest Studies in Philosophy* **2** (1), p. 90-99.

Haugeland, J. [2000]: “Pattern and Being” In Haugeland, J. *Having Thought: Essays in the Metaphysics of Mind.* Cambridge, MA: Harvard University Press.

Hempel, C. G. [1965]: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: Free Press.

Herz, A. V. M., Gollisch, Y., Machens, C. K.and Jaeger, D. [2006]: “Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction.” *Science (New York, N.Y.)* **314** (5796), p. 80–85. <doi:10.1126/science.1127240>.

Hitchcock, Christopher [1999]: Contrastive Explanation and the Demons of Determinism. *British Journal for the Philosophy of Science* **50** (4), p. 585-612.

Hitchcock, C. R. [1996]: “The Role of Contrast in Causal and Explanatory Claims.” *Synthese* **107** (3), p. 395 - 419.

Hochstein, E. [2015]: “One Mechanism, Many Models: A Distributed Theory of Mechanistic Explanation.” *Synthese* **193** (5), p. 1387–1407.

Hodgkin, A. L., and Huxley, A. F. [1952]: “A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve.” *The Journal of Physiology* **117** (4), p. 500.

Kaiser, M., Krickel, B. (2016). "The Metaphysics of Constitutive Mechanistic Phenomena," The British Journal for the Philosophy of Science, doi: 10.1093/bjps/axv058.

Kaplan, D. M. [2011]: “Explanation and Description in Computational Neuroscience.” *Synthese* **183** (3), p. 339–73.

———. [2015]: “Moving Parts: The Natural Alliance between Dynamical and Mechanistic Modeling Approaches.” *Biology & Philosophy* **30** (6), p. 757–86.

———. [In press]: “Neural Computation, multiple realizability, and the prospects for mechanistic explanation”. In Kaplan, D. M. (ed.), *Explanation and Integration in Mind and Brain Science*. Oxford, UK: Oxford University Press.

Kaplan, D. M. and Craver, C. F. [2011]: “The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective.” *Philosophy of Science* **78** (4), p. 601–27.

Levy, A. [2013]: “What Was Hodgkin and Huxley’s Achievement?” *The British Journal for the Philosophy of Science* **65** (3), p. 469-492.

Levy, A. and Bechtel, W. [2013]: “Abstraction and the Organization of Mechanisms.” *Philosophy of Science* **80** (2), p. 241–61.

Longino, H. [2012]: *Studying Human Behavior.* University of Chicago.

Machamer, P.K., Darden, L. and Craver, C.F. [2000]: Thinking about Mechanisms. *Philosophy of Science.* **67**, p. 1-25.

Marcus, E.A. [2001]: "Mental Causation: Unnaturalized but not Unnatural" *Philosophy and Phenomenological Research.* **63**, p. 57-83.
Marder, E. [2015]: "Understanding Brains: Details, intuition, and big data." *PloS Biology* **13**, p. e79306.
Matthewson, J. and Weisberg, M. [2009]: "The Structure of Tradeoffs in Model Building" *Synthese* 170 (1):169-190.
Milkowski, M. [2016]: "Explanatory Completeness and Idealization in Large Brain Simulations: A Mechanistic Perspective." *Synthese* **193** (5), p. 1457-1478.
Morgan, M.S., and Morrison, M. [1999]: Models as Mediating Instruments. *Ideas in Context*. **52**, p. 10-37.
Piccinini, G. and Craver, C. F. [2011]: "Integrating psychology and neuroscience: Functional analyses as mechanism sketches." *Synthese* **183**, p. 283-311.
Potochnik, A. [2015] "Causal Patterns and Adequate Explanations" *Philosophical Studies* 172(5): 1163-1182.
Rice, C. [2015]: Moving Beyond Causes: Optimality Models and Scientific Explanation. *Noûs* **49**(3), p. 589-615.
Ross, L. N. [2015]: "Dynamical Models and Explanation in Neuroscience." *Philosophy of Science* **82** (1), p. 32–54.
Salmon, W. C. [1984]: *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
Salmon, W. C. [1989]: "Four Decades of Scientific Explanation." *Minnesota Studies in the Philosophy of Science* **13**, p. 3-219.
Schaffner, K. F. [2008]: "Theories, Models, and Equations in Biology: The Heuristic Search for Emergent Simplifications in Neurobiology" *Philosophy of Science* 75: 1008-1025.
Sejnowski, T. J., Koch, C. and Churchland. P. S. [1988]: "Computational Neuroscience." *Science (New York, N.Y.)* **241** (4871), p. 1299–1306.
Strevens, M. [2008]: *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.
Trumpler, M. [1997]: "Converging Images: Techniques of Intervention and Forms of Representation of Sodium Channel Proteins in Nerve Cell Membranes." *Journal of the History of Biology* **30**, p. 55-89.
Van Frassen, B. [1989]: Laws and Symmetry, Oxford, UK: Oxford University Press.
Waters, K. [2007]: "Causes that Make a Difference." *Journal of Philosophy* **CIV**, p. 551–579
Weber, M. [2005]: *Philosophy of Experimental Biology*. Cambridge, UK: Cambridge University Press.
Weber, M. [2008]: "Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation." *Philosophy of Science* **75** (5), p. 995–1007.
Weisberg, M. [2007]: *Simulation and Similarity*: *Using models to understand the world.* Oxford, UK; Oxford University Press.
Woodward, J. [2003]: Making Things Happen: A Theory of Causal Explanation. Oxford, UK: Oxford University Press.
Woodward, J. [2010]: "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology and Philosophy* **25** (3), p. 287-318.
Woodward, J. [2013]: "II—James Woodward: Mechanistic Explanation: Its Scope and Limits." In *Aristotelian Society Supplementary Volume*, **87**, p. 39–65.

Woodward, J. [2016]: Explanatory Autonomy: The Role of Proportionality, Stability, and Conditional Irrelevance.
Woodward, J. [In press]. “Explanation in Neurobiology: An Interventionist Perspective”. In Kaplan, D. M. (ed.), *Explanation and Integration in Mind and Brain Science*. Oxford, UK: Oxford University Press.
Yablo, S. [1992]: “Mental causation.” *Philosophical Review* **101**, p. 245–280