Biol Philos
DOI 10.1007/s10539-017-9595-x

Why one model is never enough: a defense of explanatory holism

Eric Hochstein[1]

Received: 14 July 2016 / Accepted: 14 September 2017

Abstract Traditionally, a scientific model is thought to provide a good scientific explanation to the extent that it satisfies certain scientific goals that are thought to be constitutive of explanation (e.g. generating understanding, identifying mechanisms, making predictions, identifying high-level patterns, allowing us to control and manipulate phenomena). Problems arise when we realize that individual scientific models cannot simultaneously satisfy all the scientific goals typically associated with explanation. A given model's ability to satisfy some goals must always come at the expense of satisfying others. This has resulted in philosophical disputes regarding which of these goals are in fact necessary for explanation, and as such which types of models can and cannot provide explanations (e.g. dynamical models, optimality models, topological models, etc.). Explanatory monists argue that one goal will be explanatory in all contexts, while explanatory pluralists argue that the goal will vary based on pragmatic considerations. In this paper, I argue that such debates are misguided, and that both monists and pluralists are incorrect. Instead of any goal being given explanatory priority over others in a given context, the different goals are all deeply dependent on one another for their explanatory power. Any model that sacrifices some explanatory goals to attain others will always necessarily undermine its own explanatory power in the process. And so when forced to choose between individual scientific models, there can be no explanatory victors. Given that no model can satisfy all the goals typically associated with explanation, no one model in isolation can provide a good scientific explanation. Instead we must appeal to collections of models. Collections of models provide an explanation when they satisfy the web of interconnected goals that justify the explanatory power of one another.

✉ Eric Hochstein
ehochstein@uvic.ca

[1] Department of Philosophy, University of Victoria, Victoria, BC, Canada

Published online: 22 September 2017

Keywords Scientific explanation · Scientific model · Mechanism · Prediction · Understanding · High-level pattern · Regularity · Manipulation · Control · Explanatory interdependence · Explanatory monism · Explanatory pluralism · Explanatory holism

What kind of information must a scientific model convey in order to provide a good scientific explanation? This question has been at the heart of many recent debates within philosophy of science (see, for example: Batterman 2002; Craver 2006; Potochnik 2007, 2010; Weber 2008; Huneman 2010; Kaplan and Craver 2011; Lange 2013; Chirimuuta 2014; Rice 2015; Povich 2016). Traditionally, the explanatory power of a theory or model in science has been thought to relate to its ability to help us satisfy certain kinds of scientific goals. While there are disagreements regarding which goals in particular ought to be considered essential for explanation, a list of frequently defended scientific goals include:

1. Successfully conveying understanding about the target phenomenon, or making it intelligible, to an audience or inquirer (Achinstein 1983; Braverman et al. 2012; Waskan et al. 2014a, b, c).
2. Determining when a given phenomenon is expected to occur, and under what conditions (Hempel and Oppenheim 1948; Hempel 1965; Chemero and Silberstein 2008; Rice 2015).
3. Identifying general principles or patterns that all instances of the explanandum phenomenon adhere to and/or constraints that the phenomenon must conform to (Batterman 2001, 2002; Weber 2008; Matthewson and Weisberg 2009; Lange 2013; Chirimuuta 2014).[1]
4. Identifying the particular physical mechanisms that generate and sustain the target phenomenon (Salmon 1984, 1989; Craver 2006; Bechtel and Abrahamsen 2005; Strevens 2008; Kaplan and Craver 2011).
5. Providing information sufficient to control, manipulate, and reproduce the target phenomenon (Woodward 2000, 2003; Eliasmith 2010)

It is the attainment of these scientific goals that are thought to imbue scientific models with explanatory power.[2] For this reason, I will refer to these as "explanatory goals". There is psychological evidence that each of the above goals is taken to be an essential part of explanation by both scientists and laypeople. For example, some psychological studies have shown that scientists are unlikely to consider a model or theory explanatory unless it provides understanding (Braverman et al. 2012 Waskan et al. 2014a, b, c). Others have shown that both scientists and laypeople tend to view a theory or model as explanatory when it subsumes phenomena under general principles or patterns which can be used for making

[1] These principles can be understood in terms of strict nomological laws, behavioural patterns, broad causal regularities, or true generalizations made about the system.

[2] The list above should by no means be interpreted as an exhaustive inventory of the sorts of scientific goals that may be relevant for scientific explanation. Additional goals may well be worth including as well. For the sake of brevity and simplicity, I will focus my attention on these five given that these have all been explicitly defended by philosophers of science in recent years for their explanatory power.

future predictions (Lombrozo and Carey 2006). Meanwhile, others still have shown that identifying mechanisms plays an essential role in the way people explain, and is a result of our need to control and manipulate the world around us (Keil 2006; Gopnik 2000). Studies suggest that the more of these goals a scientific model can satisfy, the more explanatory it is taken to be.

Problems start to arise, however, when we realize that an individual scientific model cannot simultaneously satisfy *all* the scientific goals typically associated with explanation. The ability of a given model to satisfy some goals must come at the expense of satisfying others. This has resulted in numerous philosophical disputes regarding which of these goals are in fact necessary for scientific explanation, and whether a model should count as an explanation given the particular goals that it does and does not satisfy. Those engaged in these disputes typical endorse one of two views: either an explanatory *monism*, or an explanatory *pluralism*. Those who endorse an explanatory monism argue that there is one goal in particular that is essential to scientific explanation in all contexts. A model is explanatory when it satisfies this goal, and fails to explain when it does not. Others endorse an explanatory pluralism. Under this view, satisfying *any* of these goals may be sufficient to count as an explanation under the appropriate pragmatic conditions.

In this paper, I argue that both accounts are incorrect. Specifically, both assume that a model needs to primarily satisfy one of the goals typically associated with explanation in any given context to count as explanatory (they simply disagree on which goal this is, and whether the goal varies based on pragmatic considerations). In contrast, I propose that there is an essential *explanatory interdependence* between the different goals. The explanatory power of each goal stems, at least in part, from the fact that it helps us to more easily satisfy the others. Thus instead of a monism or a pluralism, we have an explanatory *holism*. Given that no one model can satisfy all the goals typically associated with explanation, no one model in isolation can provide a good scientific explanation of a complex phenomenon. Any scientific model which satisfies some goals at the cost of others will, in doing so, undermine the very explanatory power of the model in the process by cutting it off from the other goals. Instead, the explanation is distributed across many different models which altogether satisfy the various goals needed for the explanation.

In order to make this argument, I begin section one of the paper by demonstrating how an individual scientific model can typically only satisfy some explanatory goals at the cost of satisfying others, making it impossible for any individual model to attain them all simultaneously. Using the application of dynamical models in cognitive science and optimality models in evolutionary biology as a guide, I demonstrate in section two how the necessary trade-offs between explanatory goals made by such models have led to numerous philosophical disputes regarding their explanatory status. Finally, in section three, I argue that the interdependent nature of the different explanatory goals means that such debates are largely misguided. No one goal can be shown to have greater explanatory significance or priority over others, and so no one model can be seen as explanatory in isolation of others. The holistic nature of explanation requires that we employ a range of different models which satisfy different explanatory goals in order to grant any of them explanatory

power. In this regard, we can dissolve many current philosopical disputes about the explanatory status of individual scientific models.

Explanatory goals and trade-offs

Suppose we wish to explain why the action potential of a neuron fires the way that it does. A scientific model that provides the best possible explanation (i.e. the ideal explanation) is one that:

1. Will provide us with an *understanding* of how the action potential fires.
2. Will allow us to accurately anticipate when a given action potential will fire given background conditions.
3. Will allow us to identify general patterns or principles that all action potentials adhere to.
4. Will identify the physiological mechanisms that generate the action potential.
5. Will provide information needed to intervene in, inhibit, and reproduce, the firing of the action potential.

While this certainly fits with the sorts of things neuroscientists seek in their explanation of the action potential, problems start to arise when we consider that it is impossible for any individual scientific model to satisfy all of these explanatory goals at the same time. This is because a particular model's ability to satisfy some explanatory goals must come at the expense of satisfying others.

To illustrate, consider that a model which includes more details about the structures and causes that generate the action potential of any particular neuron will not be informative as to whether cells with different morphologies and biophysical properties will produce similarly behaving action potentials. If we want to identify general principles shared by action potentials in all kinds of cells, then what we want is a model that abstracts away from the structural and causal differences that exist between the different cells in order to focus on identifying the sorts of principles and patterns that apply to them all. Likewise, the reverse will also be true. Identifying general principles that all action potentials conform to by itself does not tell us how any particular action potential is generated by the mechanisms of an individual cell so as to fit with these general principles.

One of these explanatory goals requires a high degree of specificity (identifying the workings of the particular mechanisms of a particular system), while another requires a high degree of abstraction (identifying general principles that many different mechanisms adhere to). A model which is too detailed will obscure and obfuscate the general principles or patterns we seek to identify, while a model which is too abstract in its characterization of the system will be too course-grained to tell us about the workings of the particular mechanisms we seek (for elaboration, see: Levins 1966; Jackson and Pettit 1992; Matthewson and Weisberg 2009; Batterman 2001, 2002; Potochnik 2010, 2015). Thus in this case, the model must make a trade-off between explanatory goals (3) and (4).

As a second example, consider the conflicts that can arise between models which allow for manipulation and control, and those which allow for accurate predictions of the explanandum phenomenon. Conductance-based models can provide us with a degree of control and manipulation over the workings of the action potential by characterizing the system in terms of measurable quantities needed for such interventions. This allows us to create models of the action potential that allow for direct interventions. However, these measurable quantities can often only be determined by observing the firing rates of many different neurons, and then averaging across them. This leads to idealized models which often cannot accurately predict the behaviour of actual neurons under experimental conditions. As Eugene Izhikevich notes:

> The advantage of using conductance-based models, such as the $I_{Na} + I_K$-model, is that each variable and parameter has a well-defined biophysical meaning. In particular, they can be measured experimentally. The drawback is that […] the parameters are usually measured in different neurons, averaged, and then fine-tuned (a fancy term meaning "to make arbitrary choices"). As a result, the model does not have the behavior that one sees in experiments. (2007, p. 267).

In these cases, in order for a scientific model to quantify over the system in such a way that allows us to more practically intervene in its workings, it must do so by sacrificing the degree to which it can successfully predict the behaviour of any actual real-world system. As a result, these models must make a trade-off between explanatory goals (2) and (5).

Or consider a similar problem facing any model which attempts to identify the causal mechanisms that underlie the firing of the action potential. The underlying mechanisms of a system are often influenced by features outside of the system, as well as by events that happened in the distant past. A model which characterizes the mechanisms of a system often must do so by idealizing away from the many external and historical influences that can disrupt or alter its behaviour. This means that such models are often unable to *predict* the behaviour of real-world systems since such systems are unavoidably influenced by factors not represented in the model (see: Bechtel 2015; Hochstein 2016b). And so there will be a trade-off between explanatory goals (2) and (4).

These trade-offs are due to numerous factors. Some, like the trade-off between specificity and generality, are an unavoidable consequence of the nature of representation (for a detailed explanation of why this is the case, see: Matthewson and Weisberg 2009). Other factors are a result of our psychological or cognitive limitations. The amount of detail needed to effectively control, manipulate, or reproduce sufficiently complex phenomena often comes at the cost of making that phenomenon coherent or understandable to us. Thus a trade-off between these goals is often inevitable (for further details, see: Potochnik 2015). Levins, for example, argues that this is the case with models in population biology (1966, p. 421). In other cases, trade-offs are a result of pragmatic limitations. We have limited computational resources available to us when modeling systems at any given time, and so we often must choose which goals to satisfy in place of others when

computational constraints make satisfying multiple explanatory goals impossible. And this problem is not easily remedied given that when it comes to scientific modeling, "computational constraints will never be removed, as there are always more details that could be simulated" (Eliasmith and Trujilio 2014, p. 4).

What all this means is that a scientific explanation cannot simply be thought of as any model that satisfies *all* our explanatory goals simultaneously, since this is frequently not possible. So what do we do when we have conflicts of this sort? Does sacrificing one explanatory goal (e.g. the ability to predict the phenomenon) in order to attain another (e.g., the ability to manipulate the phenomenon) make our model more explanatory, or less? Which goals are we permitted to sacrifice while still satisfying the conditions needed for a good scientific explanation? In recent years, philosophers have disagreed on exactly this issue, and have emphasized the importance of different sorts of explanatory goals. It is to these conflicts that we now turn.

Disagreements and debates

Given that scientific models cannot satisfy all the goals that have typically been associated with explanation, recent philosophical disagreements have emerged regarding which goals should be taken as genuinely explanatory, and which should not, so as to determine which trade-offs are appropriate. In order to illustrate this point, consider the debates that have surrounded the explanatory status of dynamical models in cognitive science, as well as those surrounding optimality models in evolutionary biology.

Dynamical models

Many systems in nature are not merely complex, but are constantly in the process of changing. Effectively studying such systems in science often requires us to identify, track, and predict when and how these changes occur over time. To this end, Dynamical Systems Theory (DST) is a mathematical formalism used to characterize the changing behaviour of complex systems over time by employing sets of differential equations. The application of this formalism allows us to construct scientific models which abstractly represent the system as a vector moving through a multi-dimensional phase space, where the different dimensions of the space represent different variables that are relevant to possible states the system can occupy. Within these state spaces, there will be certain regions that the trajectory of the vector will tend towards or be drawn to. These are often referred to as "attractors". This type of model allows us to map out and predict stable patterns and regularities in the behaviour of the vector moving through the space (and thus to understand the types of possible physical states the actual system is likely to be in, and to move towards). Dynamical models are used in science to study a diverse range of phenomena, from weather patterns, to the behaviour of fluids, to neural activity.

Recently, philosophers of science have debated whether or not these types of models can provide good scientific explanations, and under what conditions. For a concrete example, consider the study of the action potential in the history of neurophysiology. In the 1950s, Alan Lloyd Hodgkin and Andrew Huxley developed a mathematical model of the action potential in the squid giant axon which characterized the ion flow of its sodium and potassium channels (1952). At the time, the model provided previously unknown information regarding electrochemical properties of action potentials. Specifically, that the action potential could be understood in terms of sodium and potassium conductances with specific voltage and time dependencies. While the model was able to mathematically characterize the time and voltage dependencies that were responsible for changes in the electrical potential of the axon's membrane, the model itself remained silent as to the possible mechanisms that might be responsible for producing those dependencies (Hodgkin 1992, p. 291).

In the 1960s, Fitzhugh and Nagumo et al. took the Hodgkin and Huxley model and modified it to create one of the first dynamical models of the action potential (Fitzhugh 1960; Nagumo et al. 1962). In order to do so, they simplified away many of the variables contained within the already abstract Hodgkin and Huxley model to allow for a simpler visualization of the phase space and the trajectory of a vector through that space (Ross 2015, p. 39). This model allowed scientists to identify certain dynamic principles that the action potential adhered to, and allowed scientists to predict the behaviour of many different action potentials in different neurons. The question is: does a dynamical model like the Fitzhugh–Nagumo model provide a good scientific *explanation* of the action potential?

When it comes to models like the Hodgkin and Huxley model and the Fitzhugh–Nagumo model, some have argued that they fail to provide a scientific explanation in virtue of focusing primarily on prediction and failing to identify the causal mechanism responsible for the behaviour of the action potential (Bogen 2005; Craver 2006). James Bogen argues that they "do not purport to explain anything. They are important because investigators can rely on them to suggest facts to be explained and tactics for explaining them" (Bogen 2005, p. 405). In other words, by failing to satisfy explanatory goal (4), they cannot provide a scientific explanation. Others, meanwhile, argue that these types of models fail to explain for different reasons. Specifically, because the abstract and simplified nature of dynamical models means that they cannot provide information that is necessary to intervene in, control, or reproduce the firing of the action potential. Put simply, the model's inability to satisfy explanatory goal (5) is what keeps it from being explanatory (Eliasmith 2010).

Some, however, have argued that such models *do* provide good scientific explanations in virtue of satisfying *other* explanatory goals. Ross argues that dynamical models like the Fitzhugh–Nagumo model are explanatory because they "show how different physical systems display the same universal behavior" (Ross 2015, p. 49). Likewise Weber (2008) notes that the Hodgkin and Huxley model (and by extension the Fitzhugh–Nagumo model) provides an explanation in virtue of identifying general principles or patterns that all action potentials adheres to; in this case, a general causal regularity regarding time/voltage dependencies (2008,

p. 1002). Put simply, such models are taken to be explanatory because they satisfy explanatory goal (3).

Next, consider the way that the necessary trade-offs made by dynamical models is relevant to these debates. According to Ross and Weber, the generality of certain models (like the Hodgkin and Huxley model and the Fitzhugh–Nagumo model) is what makes the model explanatory. If this is the case, then in order to satisfy explanatory goal (3), the model must *fail* to satisfy explanatory goal (4) since abstracting away from the underlying mechanisms is what gives the model the generality it needs to identify a regularity or pattern that exists across a range of different cells. In other words, "the superficial model has a nice generality to it, as it applies to very different materials, abstracting from the details of their molecular structure. This is what physical explanations often do" (Weber 2008, p. 1005).

But what if we tried to modify such a model so that it included the mechanistic details, thereby satisfying explanatory goal (4)? After all, while the variables in a dynamical model often represent high-level non-physical parameters or abstract behavioural features of systems (as opposed to components of mechanisms), they *can* be amended to include variables which *do* map directly to particular components of a mechanism. Mechanists have argued that when a dynamical model is able to make such a mapping, it can satisfy explanatory goal (4) and thus successfully provide a good scientific explanation (Kaplan and Craver 2011; Kaplan and Bechtel 2011; Zednik 2011). However, by altering a dynamical model to allow for such a mapping, we change the sorts of trade-offs that the dynamical model must now make.

Take the Fitzhugh–Nagumo model. In order for the variables in the model to map to the structural features of any one mechanism, the model has to give up the essential generality needed to identify features shared by many different neural mechanisms. In this respect, the model must trade off (3) in order to satisfy (4). It is also important to recognize that the creation of such a mechanistic dynamical model, while being able to satisfying explanatory goal (4), would still be unable to satisfy explanatory goal (5). This is because even if the variables in the model map to the physical mechanisms of a system, dynamical descriptions are often too abstract to allow for manipulation and control over the system's workings. As Chris Eliasmith argues:

> A related consequence of DST's treatment of lumped parameters is that the mechanisms described by DST are highly abstract. […]. Hence, methods of interacting with the system are not evident from such models. Without being able to predict the effects of interventions, the models become less useful to the brain sciences. (2010, p. 319)

But suppose we were to modify the model even further so that it not only had variables that mapped to a given mechanism, but also provided the relevant sort of information needed to allow for interventions as well. In doing so, our model would be able to satisfy both goals (4) *and* (5), but by adding additional parameters to our model we complicate it further and in doing so sacrifice our ability to effectively understand the phenomenon. In fact, the very reason that the Fitzhugh–Nagumo model simplifies away many of the details of the Hodgkin-Huxley model (which

was itself already an abstract description of the action potential that did not identify its underlying causal mechanisms) was because such simplifications were essential to provide an *understanding* of the relevant behavioural regularities and patterns (Fitzhugh 1960; Ross 2015). Fitzhugh himself justified the simplifying assumptions of the model on the grounds that such simplifications lead "to a better understanding of the complete system than can be obtained by considering all the variables at once" (Fitzhugh 1960, p. 873). Hoppensteadt and Izhikevich (1997) make a similar argument, noting that models which include additional physiological details needed for manipulation and control can…

> …become a trap, since the more neurophysiological facts are taken into consideration during the construction of the model, the more sophisticated and complex the model becomes. As a result, such a model can quickly come to a point beyond reasonable analysis even with the help of a computer. (1997, p. 5)

This means that while the Fitzhugh–Nagumo model satisfies explanatory goals (1) and (3), it cannot satisfy (4) and (5). By modifying the model to better satisfy (4), it loses the ability to satisfy (3). And by modifying it to better satisfy (5), it loses the ability to satisfy (1). And so the question of whether a dynamical model like the Fitzhugh–Nagumo model provides a good scientific explanation depends on which explanatory goal one chooses to emphasize. Different philosophers emphasize the importance of different goals, and so disagree as to which sorts of dynamical models provide explanations, and under what conditions.

Optimality models

Let us consider a second example to further illustrate. Consider the use of optimality models in evolutionary biology. Optimality models employ a particular kind of mathematical technique known as Optimality Theory in order to understand and predict the appearance of phenotypic traits in a given population. Put simply, such models treat natural selection as if it were the only causal force in the evolutionary process, and then determine what the most locally optimal trait for a creature to have would be given appropriate biological and environmental trade-offs. The idea is that *locally optimal* results will tend to be produced by natural selection over long enough periods of time, despite there being some obvious exceptions.

For a concrete example, consider the Wang, Dykhuizen, & Slobodkin model (hereafter WDS model). The WDS model was created to help understand and predict the life cycle of a lytic bacteriophage (Wang et al. 1996). A bacteriophage is a virus which infects bacterial cells and reproduces by bursting from the infected cell into the environment in order to infect others. The original host cell is destroyed in this reproductive process. The WDS model characterizes the optimal amount of time for a bacteriophage to incubate within a cell before reproducing to maximize reproductive success (for details and discussion, see: Wang et al. 1996; Bull et al. 2004; Bull 2006; Potochnik 2010). By using the WDS model, biologists are able to accurately predict the lysis time of a bacteriophage, as well as identify general patterns and constraints that influence lysis timing. As a result, the WDS model

satisfies explanatory goal (2) and (3). This has led some, like Angela Potochnik (2015) and Collin Rice (2015), to conclude that optimality models like the WDS model provide a good scientific explanation for why a given trait occurs in a population (i.e. by identifying what is *locally optimal* for the organism).

What is particularly noteworthy about the WDS model, and optimality models more generally, is that in order to satisfy these particular explanatory goals, the model *must* idealize away from most of the known genetic mechanisms actually involved in the evolutionary processes. As Bull et al. note, "optimality models assume that phenotypes evolve by natural selection largely independently of underlying genetic mechanisms" (2004, p. 76). Likewise, Rice argues that "optimality models are typically so idealized that they provide little (if any) accurate information about any of the causes within the model's target system(s)—even if we consider causes at the 'macro' level" (2015, p. 601). So in order to satisfy explanatory goals (2) and (3), the idealized nature of optimality models means they must trade-off their ability to satisfy explanatory goal (4).

However, any attempt to rectify this by including more accurate mechanistic details into the model would actually result in a loss of predictive success for the model (for a detailed explanation, see: Rice 2015), as well as interfere with our ability to identify and understand the phenomenon (Levins 1966; Potochnik 2010; Woods and Rosales 2010). As such, satisfying goals (4) would cost us goals (1) and (2).

Others, meanwhile, have argued that by ignoring the relevant causal mechanisms in the evolutionary process, optimality models like the WDS model *fail* to provide explanations (Gould and Lewontin 1979; Lewontin 1979, 1989; Schwartz 2002). Lewontin (1979), for example, argues that optimality models are a useful heuristic tool, but that in order to provide a true explanation, one needs a model that will "predict the evolutionary trajectory of the community […] on a purely mechanical basis" (1979, p. 6). Similarly, Craver (2006) argues that models which do not identify mechanisms and merely subsume the phenomenon under a set of generalizations or constraints do not satisfy the appropriate sorts of scientific goals needed for a good explanation.[3] Here again we can see that the crux of the dispute regarding the explanatory status of a given model stems from which particular explanatory goals the model is able to satisfy. Theorists disagree as to which goals are essential to the explanation, and thus which trade-offs are acceptable and which are not.

While I have focused my attention on two particular types of models in this section, the fact that models must make trade-offs between different explanatory goals is not a feature unique to dynamical or optimality models, nor are philosophical dispute regarding the explanatory status of such models. The same sorts of debates currently surround the explanatory status of computational models

[3] It should be noted that Craver is not suggesting that a given scientific model will always become better the more mechanistic details it includes (see: Craver and Kaplan, under review). The appropriate amount of mechanistic detail for a model to employ will vary based on our particular needs. Instead, he argues only that a model must always have some variables that map to structural/mechanistic features of the system in order to carry explanatory content (which optimality models do not have). A model which satisfies the other explanatory goals but fails to identify relevant mechanisms cannot be explanatory.

(Chirimuuta 2014), topological models (Huneman 2010), and statistical models (Eliasmith 2010) for the same reason. And so the question becomes, which particular explanatory goals must a model satisfy to provide an explanation when trade-offs are inevitable?

Most philosophers tend to fall into one of two camps. Either they argue that one of the explanatory goals will prove essential for all cases of explanation (and thus our ability to explain will require it to be satisfied in all instances), or they argue that satisfying *any* of the explanatory goals may be sufficient to count as an explanation given the appropriate pragmatic context. In other words, they are either *monists* or *pluralists* about explanatory goals. In either case, this provides them a means of determining which individual models provide explanations and which do not in a given situation. I propose that neither monism nor pluralism best fits with the way in which models are used to explain in science.

A defence of explanatory holism

In both examples discussed in the section "Disagreements and debates", it is assumed that because an individual model cannot satisfy all our explanatory goals simultaneously, that some goals *must* trump others for the purposes of explanation so that one model or another can be deemed the explanation for that phenomenon in that context. In other words, it is assumed that there is always some way of determining a clear victor between conflicting explanatory goals so as to grant one model or another the status of an explanation. This is true for both explanatory monists, and for explanatory pluralists.

For the monist, there will be one explanatory goal in particular that must always be satisfied in order to achieve a scientific explanation. Craver, for example, argues that while the goals of prediction and understanding are undoubtedly important to science, they are not constitutive of *explanation*. Instead, only models that satisfy the goal of identifying causal mechanisms provide information sufficient for explanation (Craver 2006; Kaplan and Craver 2011; Piccinini and Craver 2011). The pluralist, on the other hand, argues that satisfying any of the explanatory goals may be sufficient for a scientific explanation under the appropriate pragmatic conditions. Under certain conditions, a model which subsumes the phenomenon under a general pattern will count as explanatory; under different conditions a model which identifies physical mechanisms will count as explanatory, etc.[4]

I propose that both the monist and the pluralist are mistaken. In both cases, it is assumed that there is a particular explanatory goal that a model needs to satisfy in order to count as an explanation in a given context or situation. The disagreement lies in whether this explanatory goal is invariant across all explanatory contexts, or whether it changes depending on pragmatic considerations. The problem with both

[4] It is worth noting that the term "explanatory pluralism" is not always used consistently throughout the philosophy of science literature. As such, this pragmatic contextualist interpretation of explanatory pluralism may not correctly describe all those who self-identify as pluralists. For the sake of clarity, I have in mind here the sort of explanatory pluralism advocated by the likes of Chemero and Silberstein 2008, and Chirimuuta 2014 (among others).

of these options is that they tend to ignore the explanatory interdependence that exists *between* the different scientific goals.

Instead of one goal being given explanatory priority over others in a given context, the different goals are deeply dependent on one another for their explanatory power. In other words, the explanatory power of each goal stems, at least in part, from the fact that it helps us to more easily satisfy the others. Thus any model that sacrifices or trades-off some explanatory goals in favour of others will always necessarily undermine its own explanatory power by doing so.

To illustrate, recall the justification for the claim that dynamical models like the Fitzhugh–Nagumo model provide scientific explanations. Such models are intended to capture general patterns or principles that the explanandum phenomenon adheres to. In other words, it satisfies explanatory goal (3). Yet, the main reason *why* this goal is considered explanatory in the study of the action potential is because, by identifying the dynamic principles that guide the system, we are able to gain a *better understanding* of its overall behaviour than can be gained from models which include additional mechanistic details (Fitzhugh 1960; Ross 2015). Thus, the explanatory justification for satisfying goal (3) is that it is the most effective way to help us satisfy goal (1). But imagine if the Fitzhugh–Nagumo model failed to provide us with any understanding of the action potential's behaviour. In such a case, the claim that the model provides a good scientific explanation loses much of its force. In fact, recent studies have shown that practicing scientists are unlikely to consider such models explanatory if they are unable to provide understanding (Braverman et al. 2012; Waskan et al. 2014a, b; Waskan et al. 2014a, b, c). In a set of experiments conducted by Waskan et al. (2014a, b, c),

> We found that participants were less likely to regard a model as an explanation in the Potentially Intelligible and Never Intelligible conditions than in the Intelligible condition, this despite the fact that the models in question were said to have other major theoretical virtues (e.g., predictive power and fit with surrounding theories). (2014, p. 1025)

Here, satisfying goal (3) is explanatory *because* it helps us to more easily satisfy goal (1). If the model were thus to satisfy explanatory goal (3) at the cost of satisfying explanatory goal (1), then it would undermine its own explanatory power in the process.

Others, meanwhile, have argued that models which satisfy goal (3) are explanatory because they help us to *predict* the explanandum phenomenon, thereby satisfying explanatory goal (2) (Chemero and Silberstein 2008. See also: Chirimuuta 2014, p. 140). Subsuming the phenomenon under a general regularity or pattern, or identifying principles that govern the system's behaviour, is informative precisely because it allows us to *predict* how the system is likely to behave in different situations. In fact, psychological studies have shown that the very reason we psychologically subsume phenomena under generalizations when providing explanations is because it allows us to extrapolate from those generalizations to solve problems in novel situations. Lombrozo and Carey, for instance, conducted a number of experiments showing that "a psychological function of explanation [in the sense of conforming to a predictable pattern] is to highlight information likely to

subserve future prediction and intervention" (2006, p. 167). Consider: if the Fitzhugh–Nagumo model claimed to identify time–voltage dependencies of the action potential, but failed to accurately predict any such dependencies in all the neurons studied, then the explanatory force of the model becomes undone. Thus the model would undermine its own explanatory force by satisfying (3) at the cost of (2).

Interestingly, the goal of prediction is also thought to be explanatory by many scientists because it is an essential tool in helping us to satisfy the explanatory goal of identifying mechanisms. Models which predict the occurrence of the phenomenon are often treated as explanatory in part because they put essential constraints on the sorts of mechanisms that are capable of fitting with those predictions, and thus play an essential role in their discovery (see: Piccinini 2015; Piccinini and Craver 2011; Hochstein 2016a, b). Yet, ironically, it has also been argued that part of what makes the identification of mechanisms explanatory is the very fact that knowing the mechanisms of the system allows us to better predict what the occurrence of the phenomenon will be under various interventions, and in counter-factual situations (see: Woodward 2000, 2003; Craver 2006). In this respect, a model that identifies mechanisms is partially explanatory because it allows us to better predict the behaviour of the system, and a model which is predictive is partially explanatory because it allows us to better identify the mechanisms of the system.

Likewise, the identification of mechanisms is frequently considered to be explanatory because it helps us to satisfy the goal of intervention, control, and reproduction of the phenomenon. As Eliasmith notes:

> In the case of cognitive and brain sciences, useful explanations are those that appeal to subpersonal mechanisms. This is because it is precisely such explanations which provide a basis for both intervention in behaviour and the artificial reproduction of those behaviours. These mechanisms must be specific enough to allow for intervention. That is, the mechanisms must be specified in a way that relates to the measurable and manipulable properties of the system. (2010, p. 316)

Craver makes a similar argument, noting that models which identify mechanisms have greater explanatory power than other models because they "are much more useful than merely phenomenal models for the purposes of control and manipulation" (2006, p. 358). In other words, the explanatory force of goal (4) comes from the fact that satisfying it helps us to more easily satisfy explanatory goal (5). This idea is likewise supported by evidence from numerous studies in developmental psychology and comparative psychology that have shown that we view the identification of causal mechanisms as explanatory because "once we represent the causal relations among ourselves, our conspecifics and objects, we can intervene in a much wider variety of ways to get a particular result." (Gopnik 2000, p. 303). And so if a model allowed us to identify mechanisms, but failed to give us information needed for such interventions, then its explanatory force is likewise undercut.

This, however, does not mean that explanatory goal (5) is intrinsically explanatory by itself either. Our ability to manipulate, control, and reproduce

phenomena is frequently thought to be explanatory because it helps us to *understand* the explanandum phenomenon (see: Dretske 1994). In other words, explanatory goal (5) is explanatory in part because satisfying it helps to satisfy explanatory goal (1). Cases where such models fail to provide understanding are typically *not* considered explanatory by practicing scientists (Braverman et al. 2012).

To complicate matters even further, the explanatory goal of understanding is itself often only considered explanatory because it is a sign that we have satisfied many of our other explanatory goals. Hempel, for instance, thought that "understanding why an outcome occurs is a matter of seeing that it was to be expected on the basis of a law." (Woodward 2017). Wesley Salmon argued that "to understand why certain things happen, we need to see how they are produced by [causal] mechanisms" (Salmon 1984, p. 132). Others, meanwhile, have proposed that to understand a phenomenon requires being able to control and reproduce it. The physicist Richard Feynman, for example, famously stated "that which I cannot create, I do not understand" (Eliasmith and Trujillo 2014, p. 1). In this regard, explanatory goal (1) is thought to have explanatory power because it is a sign that we have satisfied explanatory goals such as (2), (4), or (5).[5] It is not hard to see why this is the case. After all, to claim that a model provides us with an understanding of a given phenomenon, but which fails to provide any insights into how the phenomenon is produced, when or where the phenomenon occurs, or how to reproduce or intervene in it, calls into question what it even means to claim that we have an understanding of it at all.

Indeed, *any* explanatory goal that is satisfied in isolation of the others runs into a similar problem. A model that allows us to predict the phenomenon, but which provides us with no understanding of it, does not identify any principles or constraints that influence its behaviour, identifies no underlying causes for it, and provides no insight into how to manipulate, control or reproduce it, is typically not treated as explanatory (Keil 2006; Legare et al. 2009; Braverman et al. 2012).

Or consider models which identify mechanisms. A model which identifies a mechanism, but fails to provide any understanding of how it works, cannot accurately predict its resulting behaviour, does not identify any general principles or constraints on its behaviour, and does not allow us to intervene in its workings, is typically not considered explanatory *even by mechanists*. This is because one of the primary guides for determining if we have successfully identified the proper causal mechanism, and thus provided an explanation, is whether our mechanistic account lets us accurately *predict* the phenomenon (Machamer et al. 2000), *understand* it

[5] One might object that this simply reflects an ambiguity in the term "understanding" as opposed to any deeper claim regarding the interdependence between the goal of understanding and the other explanatory goals (special thanks to a blind referee for pointing out this worry). While constraints on space limit my ability to address this problem at length here, it should be sufficient for my purpose to highlight the fact that almost every definition of understanding involves some sort of cognitive component in which the target phenomenon is made intelligible to the inquirer (for psychological studies that support this, see: Keil 2006; Braverman et al. 2012; Waskan et al. 2014a, b, c. See also: Potochnik 2015). This very minimal shared criterion of "understanding" is sufficient to show the interdependence between it and the other goals, as each of the other goals has been defended as essential for explanation on the grounds that the psychological intelligibility of the phenomenon is contingent on their attainment. That being said, this point is still contentious and may deserve greater exploration.

(Glennan 2002), or *manipulate* it (Craver 2006; Bechtel 2008). A mechanistic model which fails to do any of these things is taken to have failed as an explanation.

What all this shows is that forcing us to choose *between* explanatory goals is explanatorily detrimental, as each goal in isolation fails to explain without the others. The different explanatory goals support one another, since their explanatory value is partially defined in terms of their relationship to one another. Moreover, the more explanatory goals we can satisfy, the easier it becomes to satisfy the others given their interconnected nature. Thus instead of an explanatory pluralism or an explanatory monism, what we have instead is an explanatory *holism*; one where the goals of explanation reinforce and depend on one another, and cannot be cut off from one another without undermining their explanatory power in the process. As such, the nature of explanation may simply not allow us to adjudicate between the different explanatory goals, since it is their very interdependence that makes them explanatory.

Thus we are now left with a problem. If the different explanatory goals depend on one another for their explanatory power, then any model which satisfies one goal at the cost of others will inevitably cut that goal off from the very things that *make it* explanatory. The irony of this is that the essential trade-offs that individual models must make in order to gain a degree of explanatory power by satisfying certain goals at the cost of others undermines the very explanatory power of those goals in the process. For example, if what makes the goal of identifying causal mechanisms explanatory is that it allows us to better predict the phenomenon of interest, or that it provides information needed to intervene in the workings of the system, then a model which can only identify mechanisms by *sacrificing* its ability to predict the phenomenon and intervene in its workings *undermines* the very reason for considered that goal explanatory in the first place.[6]

This means that we cannot grant priority to any one explanatory goal over another, since their interconnected nature is essential to their explanatory status. This is not to say that we never have good reasons for prioritizing one sort of scientific goal over another for other sorts of scientific reasons of course. Scientists rightfully care about things other than just explanation. A medical professional who cares about treating a patient may wish to use a model that allows them to control and manipulate the phenomenon over one that identifies abstract principles that the phenomenon obeys, since that is best suited for the task of treatment. However, the fact that we have good reason to prioritize one goal over another for a given scientific purpose does not mean that such a goal is therefore *more integral to the very nature of scientific explanation* than the others. We must be careful not to confuse prioritizations of goals for different pragmatic purposes with an *explanatory* prioritization of one goal over another. The fact that we sometimes care more about one goal over others does not *ipso facto* show that satisfying the goal is thereby

[6] For a straightforward example of this sort of model, consider the use of large scale graph-based models to characterize certain organizational features of complex biological mechanisms. Such models are often necessary for representing organizational features like complex feedback loop, but can only do so by idealizing away from many of the structural and behavioural features of the system needed for both manipulation and prediction (for details and discussion, see Bechtel 2015).

more explanatory than satisfying others.[7] Given the interdependent nature of the explanatory goals, the assumption of explanatory prioritization is self-undermining.

With this in mind, let us return to the question of whether the Fitzhugh–Nagumo model and the WDS model count as good scientific explanations. Philosophers on both sides of these debates assume that there is some set of criteria we can appeal to in order to determine if the model definitely *does* or *does not* provide an explanation. Yet, this assumes that an individual model has the ability to satisfy not only a particular explanatory goal, but also all the other goals needed for sustaining that goal's explanatory power. Yet, given the necessary explanatory trade-offs that each model must make, this is simply not possible. Neither the Fitzhugh–Nagumo model nor the WDS model can satisfy all the goals that are needed to grant either one of them explanatory power. Yet this will inevitably be the case with *any* individual scientific model we employ.

The mistake being made is the assumption that scientific explanations must be provided by *individual* models. To ask whether a given model provides an explanation is to assume that the explanation can be contained in, or conveyed by, a single model or representation. If we reject this claim, then we can interpret our explanatory practices in a way which dissolves these philosophical disputes. Scientists rarely, if ever, point to a single model as being *the* explanation of a given phenomenon. Instead, they appeal to many different models when engaging in the act of explaining a complex phenomenon (for numerous examples, see: Trumpler 1997; Mitchell 2003; Weisberg 2013, p. 103; Hochstein 2016a; Miłkowski 2016). Once we give up the idea that explanations must be provided by individual representations, then the holistic nature of explanation and the trade-offs made by individual models becomes reconcilable.

As noted above, the explanatory interdependence between the different goals means that satisfying one explanatory goal allows us to more easily satisfy others. This means that even though individual models must sacrifice their ability to satisfy some goals in order to successfully satisfy others, the information we gain from a model can be used to help us construct *other* models which satisfy other goals that the initial model had to trade off. This in turn generates information which can be used to build even more models which satisfy even more explanatory goals, and to refine previous models in light of the new goal being satisfied. The more models we develop which satisfy different explanatory goals, the more our entire collection of models becomes explanatory by allowing us to satisfy the entire interconnected set. As a result, scientific explanations are distributed across sets of different models which satisfy different goals, and in doing so help to improve the explanatory power of one another. This explanatory interdependence can be seen clearly when we observe how models like the Fitzhugh–Nagumo model and the WDS model influence, and are influenced by, the construction of other scientific models which satisfy distinct explanatory goals.

Consider the WDS model. In order for the WDS model to accurately account for the bacteriophage reproductive cycle, particular mechanisms in the evolutionary processes must be present and active, while other potentially disruptive factors must

[7] Thanks to Natalia Washington for encouraging me to emphasize this distinction.

be absent. Knowing these mechanistic details tells us when and how to apply the optimality model, even if the inclusion of this information into the model *itself* may inhibit its ability to work properly. As Potochnik notes:

> Because optimality models use highly simplified assumptions as placeholders for complex dynamics, their successful use depends upon evolutionary dynamics that the models themselves do not explicitly represent. In other words, optimality models are epistemically dependent on unrepresented dynamics. Information about these unrepresented dynamics helps establish whether an optimality model's simplifying assumptions are problematic, and thus how successful the model is.
>
> Genetic constraints are a prime example of how unrepresented dynamics can have unanticipated effects on evolution. Genetic transmission can involve a host of complications, such as epistasis (different genes with interacting effects) and pleiotropy (one gene with different unrelated effects). Such complicating factors may cause an evolutionary outcome to deviate widely from an optimality model's predictions. For this reason, optimality models are epistemically dependent on unrepresented features of genetic transmission. (2010, p. 226)

In other words, a scientific model which identifies the presence of relevant evolutionary mechanisms and potentially disruptive factors provides invaluable information as to whether an optimality model will work, and under what conditions. Likewise, if we have a successful optimality model, then we can often use this information to determine that certain evolutionary mechanisms must be present, while others are absent. If satisfying one explanatory goal helps us to satisfy others, then building a scientific model which satisfies one goal will similarly help us to create and apply a different model in order to satisfy another. This understanding of scientific explanation does not force us to choose *between* models, but instead acknowledges their explanatory interdependence.[8]

[8] It is worth noting that Potochnik draws a very different conclusion from this interdependence between models than I do. While she grants that there is an epistemic interdependence between the different models, she insists that optimality model remain explanatorily independent from the other models. She argues that the model which identifies the high-level causal pattern is the best explanation for why a particular trait occurs. Other models, like those that identify essential evolutionary mechanisms, may be needed to effectively construct and apply an optimality model, but it is the optimality model that provides the explanation independently of those models.

Yet I propose that this interpretation is incorrect. The mechanistic details are essential to our explanation of the phenomenon, since the presence or absence of certain evolutionary mechanisms (such as epistasis and pleiotropy) is essential for the phenomenon to display the patterns represented in the optimality model. In other words, the explanation as to why the trait appears is not merely because it is locally optimal, it is because it is locally optimal *in virtue of the presence or absence of certain key mechanistic facts*. These facts are part of the explanation as to why the trait occurs as it does, and are only identified by the mechanistic model, not the optimality model. Thus the mechanistic model not only provides context for the optimality model, it provides relevant explanatory information as to *why* the optimal trait occurs. And so to suggest that the optimality model's explanatory power is independent of the mechanistic model is extremely misleading.

What appears *prima facie* to be a case of explanatory independence is instead a case in which our pragmatic interests shift our *attention* from one model to another. This shift in attention should not be

Take the Fitzhugh–Nagumo model. While the Fitzhugh–Nagumo model itself does not directly describe the underlying mechanisms of any particular action potential, it provides essential constraints needed in the discovery and understanding of those mechanisms. Not any kind of mechanism will be successfully characterized by the time and voltage dependencies that the Fitzhugh–Nagumo model describes. As such, identifying these dependencies allowed scientists in the 1960s and 70s to use them as a guide for discovering the possible mechanisms responsible for them. This was not a one-way street either. Learning more about the underlying mechanisms of the action potential similarly allowed scientists to further refine and improve upon the regularities previously described, and to develop more accurate representations of them (for details and discussion, see: Trumpler 1997; Hochstein 2016a). In this respect, a model which satisfied one explanatory goal provided essential information that helped to guide the creation of other models used to satisfy others.

Conclusion

So what must a scientific model do in order to provide a good scientific explanation? In this paper, I have argued that this very question presupposes that individual models have greater representational powers than they in fact do. Individual models can provide information needed to satisfy some of the goals thought to be constitutive of explanation, but only by undermining their explanatory power in the process. And so when forced to choose *between* individual scientific models, there can be no explanatory victors. Instead, we must appeal to *collections* of models in our explanatory practices. Collections of models provide an explanation when they satisfy the web of interconnected goals that justify the explanatory power of one another. In this regard, many current disputes in the philosophy of science regarding which sorts of models provide explanations, and which do not, are misguided. They implicitly assume that explanations must be provided by individual models. By abandoning this idea, we can better understand and reconcile the interconnected nature of our different explanatory goals, as well as the necessary trade-offs that our different models must make.

Footnote 8 continued
confused with a shift in *explanatory content* however. Once the mechanistic model is used to identifying the relevant evolutionary mechanisms, we shift our focus to the optimality model in order to satisfy explanatory goals that our mechanistic model could not provide. It only appears like the optimality model is explanatorily independent from the mechanistic model because it seems like the explanatory content is only available to us once we have the optimality model in hand, and not when we have the mechanistic model. But this perception is deceptive, since in order to generate the optimality model we must *already have available to us* the information from the mechanistic model. So by the time we apply the optimality model, the explanatory information available to us is being conveyed by both the mechanistic *and* optimality models together. It only seems like the optimality model is providing an independent explanation because the mechanistic information has been pushed into the background as we focus our attention on the optimality model, and so appears invisible. But it is only when the information from our optimality model is used to *supplement* the information from our mechanistic model that we begin to generate an explanation. The explanatory contents of the models are not independent, but deeply dependent on one another.

Acknowledgements There are many I owe a great deal of thanks for assistance with earlier drafts of this paper. This includes Callie Philips, Anya Plutynski, Tim Kenyon, Doreen Fraser, Nathan Haydon, Ian McDonald, Mark Povich, Carl Craver, and Peter Blouw. Special thanks in particular go to Lauren Olin, Joseph McCaffrey and Natalia Washington for in-depth discussions, feedback, and encouragement. I would also like to offer thanks to the blind referees of this paper. Their feedback was not only constructive and insightful, but essential in helping to shape the paper.

References

Achinstein P (1983) The nature of explanation. Oxford University Press, New York

Batterman R (2001) The devil in the details: asymptotic reasoning in explanation, reduction, and emergence. Oxford University Press, Oxford

Batterman R (2002) Asymptotics and the role of minimal models. Br J Philos Sci 53:21–38

Bechtel W (2008) Mental mechanisms: philosophical perspectives on cognitive neuroscience. Lawrence Erlbaum Associates, New York

Bechtel W (2015) Can mechanistic explanation be reconciled with scale-free constitution and dynamics? Stud Hist Philos Sci Part C: Stud Hist Philos Biol Biomed Sci. doi:10.1016/j.shpsc.2015.03.006

Bechtel W, Abrahamsen A (2005) Explanation: a mechanistic alternative. Stud Hist Philos Biomed Sci 36:421–441

Bogen J (2005) Regularities and causality; generalizations and causal explanations. Stud Hist Philos Sci Part C 36:397–420

Braverman M, Clevenger J, Harmon I, Higgins A, Horne Z, Spino J, Waskan J (2012). Intelligibility is necessary for explanation but accuracy may not be. In: Proceedings of the thirty-fourth annual conference of the cognitive science society

Bull JJ (2006) Optimality models of phage life history and parallels in disease evolution. J Theor Biol 241:928–938

Bull JJ, Pfennig DW, Wang I-N (2004) Genetic details, optimization and phage life histories. Trends Ecol Evol 19(2):76–82

Chemero A, Silberstein M (2008) After the philosophy of mind: replacing scholasticism with science. Philos Sci 75:1–27

Chirimuuta M (2014) Minimal models and canonical neural computations: the distinctness of computational explanation in neuroscience. Synthese 191(2):127–153

Craver C (2006) When mechanistic models explain. Synthese 153(3):355–376

Craver C, Kaplan D (under review) Are more details better? On the norms of completeness for mechanistic explanations

Dretske F (1994) If you can't make one, you don't know how it works. Midwest Stud Philos 19(1):468–482

Eliasmith C (2010) How we ought to describe computation in the brain. Stud Hist Philos Sci Part A 41:313–320

Eliasmith C, Trujillo O (2014) The use and abuse of large-scale brain models. Curr Opin Neurobiol 25:1–6

Fitzhugh R (1960) Thresholds and plateaus in the Hodgkin-Huxley nerve equations. J Gen Physiol 43(5):867–896

Glennan S (2002) Rethinking mechanistic explanation. Philos Sci 69(S3):S342–S353

Gopnik A (2000) Explanation as orgasm and the drive for causal knowledge: the function, evolution, and phenomenology of the theory formation system. In: Keil F, Wilson R (eds) Cognition and explanation. MIT Press, Cambridge, pp 299–323

Gould S, Lewontin R (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B 205:581–598

Hempel C (1965) Aspects of scientific explanation. Free Press, New York

Hempel C, Oppenheim P (1948) Studies in the logic of explanation. Philos Sci 15:135–175

Hochstein E (2016a) One mechanism, many models: a distributed theory of mechanistic explanation. Synthese 193(5):1387–1407

Hochstein E (2016b) Giving up on convergence and autonomy: why the theories of psychology and neuroscience are codependent as well as irreconcilable. Stud Hist Philos Sci 56:135–144

Hodgkin AL (1992) Chance and design: reminiscences of science in peace and war. Cambridge University Press, Cambridge
Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544
Hoppensteadt FC, Izhikevich EM (1997) Weakly connected neural networks. Springer, New York
Huneman P (2010) Topological exlanations and robustness in biological sciences. Synthese 177(2):213–245
Izhikevich E (2007) Dynamical systems in neuroscience: the geometry of excitability and bursting. MIT Press, Cambridge
Jackson F, Pettit P (1992) In defense of explanatory ecumenicalism. Econ Philos 8(1):1–21
Kaplan D, Bechtel W (2011a) Dynamical models: an alternative or complement to mechanistic explanations? Top Cogn Sci 3:438–444
Kaplan D, Craver C (2011b) The explanatory force of dynamical and mathematical models in neuroscience: a mechanistic perspective. Philos Sci 78(4):601–627
Keil F (2006) Explanation and understanding. Annu Rev Psychol 57:227–254
Lange M (2013) What makes a scientific explanation distinctively mathematical? Br J Philos Sci 64(3):485–511
Legare CH, Wellman HM, Gelman SA (2009) Evidence for an explanation advantage in naıve biological reasoning. Cogn Psychol 58:177–194
Levins R (1966) The strategy of model building in population biology. Am Sci 54:5
Lewontin R (1979) Fitness, survival, and optimality. In: Horn D, Stairs G, Mitchell R (eds) Analysis of ecological systems, third annual biosciences colloquium. Ohio State University Press, Columbus, pp 3–21
Lewontin R (1989) A natural selection. Nature 339:107
Lombrozo T, Carey S (2006) Functional explanation and the function of explanation. Cognition 99(2):167–204
Machamer P, Darden L, Craver CF (2000) Thinking about mechanisms. Philos Sci 67(1):1–25
Matthewson M, Weisberg M (2009) The structure of tradeoffs in model building. Synthese 170(1):169–190
Miłkowski M (2016) Unification strategies in cognitive science. Stud Log Gramm Rhetor 48(61):13–33
Mitchell S (2003) Biological complexity and integrative pluralism. Cambridge University Press, Cambridge
Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line simulating Nerve Axon. Proc Inst Radio Eng 50(10): 2061–2070
Piccinini G (2015) Physical computation: a mechanist account. Oxford University Press, Oxford
Piccinini G, Craver C (2011) Integrating psychology and neuroscience: functional analyses as mechanism sketches. Synthese 183(3):283–311
Potochnik A (2007) Optimality modeling and explanatory generality. Philos Sci 74:680–691
Potochnik A (2010) Explanatory independence and epistemic interdependence: a case study of the optimality approach. Br J Philos Sci 61(1):213–233
Potochnik A (2015) The diverse aims of science. Stud Hist Philos Sci 53:71–80
Povich M (2016) Minimal models and the generalized ontic conception of scientific explanation. Br J Philos Sci. doi:10.1093/bjps/axw019
Rice C (2015) Moving beyond causes: optimality models and scientific explanation. Noûs 49(3):589–615
Ross L (2015) Dynamical models and explanation in neuroscience. Philos Sci 81(1):32–54
Salmon W (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton
Salmon W (1989) Four decades of scientific explanation. University of Minnesota Press, Minneapolis
Schwartz J (2002) Population genetics and sociobiology. Perspect Biol Med 45(2):224–240
Strevens M (2008) Depth: an account of scientific explanation. Harvard University Press, Cambridge
Trumpler M (1997) Techniques of intervention and forms of representation of sodium-channel proteins in nerve cell membranes. J Hist Biol 30(1):55–89
Wang IN, Dykhuizen DE, Slobodkin LB (1996) The evolution of phage lysis timing. Evol Ecol 10:545–558
Waskan J, Harmon I, Horne Z, Spino J, Clevenger J (2014a) Explanatory anti-psychologism overturned by lay and scientific case classifications. Synthese 191:1013–1035

Waskan J, Harmon I, Higgins A, Spino J (2014a) Three senses of 'Explanation'. In: Bello P, Guarini M, McShane M, Scassellati B (eds) Proceedings of the 36th annual conference of the cognitive science society. Cognitive Science Society: Austin, TX, pp 3090–3095

Waskan J, Harmon I, Higgins A, Spino J (2014b) Investigating lay and scientific norms for using 'Explanation.' In: Lissack M, Graber A (eds) Modes of explanation: affordances for action and prediction. Palgrave Macmillan, pp 198–205

Weber M (2008) Causes without mechanisms: experimental regularities, physical laws, and neuroscientific explanation. Philos Sci 75:995–1007

Weisberg M (2013) Simulation and similarity: using models to understand the world. Oxford University Press, New York

Woods J, Rosales A (2010) Virtuous distortion in model-based science. In: Magnani L, Carnielli W, Pizzi C (eds) Model-based reasoning in science and technology: abduction, logic and computational discovery. Springer, Berlin, pp 3–30

Woodward J (2000) Explanation and invariance in the special sciences. British Journal for the Philosophy of Science 51:197–254

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, Oxford

Woodward J (2017) Scientific explanation. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy (Spring 2017 Edition), https://plato.stanford.edu/archives/spr2017/entries/scientific-explanation/

Zednik C (2011) The nature of dynamical explanation. Philos Sci 78(2):238–263