Penultimate Draft

"Models and Explanation"

Published in L. Magnani and T. Bertolotti (eds.)
Handbook of Model-based Science (Springer, 2017): 103-118.

Alisa Bokulich
Philosophy Department
Boston University
abokulic@bu.edu

Summary

Detailed examinations of scientific practice have revealed that the use of idealized models in the sciences is pervasive. These models play a central role in not only the investigation and prediction of phenomena, but in their received scientific explanations as well. This has led philosophers of science to begin revising the traditional philosophical accounts of scientific explanation in order to make sense of this practice. These new model-based accounts of scientific explanation, however, raise a number of key questions: Can the fictions and falsehoods inherent in the modeling practice do real explanatory work? Do some highly abstract and mathematical models exhibit a non-causal form of scientific explanation? How can one distinguish an exploratory "how-possibly" model explanation from a genuine "how-actually" model explanation? Do modelers face tradeoffs such that a model that is optimized for yielding explanatory insight, for example, might fail to be the most predictively accurate, and vice versa? This chapter explores the various answers that have been given to these questions.

4.1 Overview

Explanation is one of the central aims of science, and the attempt to understand the nature of scientific explanation is at the heart of the philosophy of science. An explanation can be analyzed as consisting of two parts, a phenomenon or event to be explained, known as the *explanandum*, and that which does the job of explaining, the *explanans*. On the traditional approach, to explain a phenomenon is either to deduce the explanandum phenomenon from the relevant laws of nature and initial conditions, such as on the deductive-nomological (DN) account (Hempel 1965 [1]), or to trace the detailed causal chain leading up to that event, such as on the causal-mechanical account (Salmon 1984 [2]). Underlying this traditional approach are the assumptions that, in order to genuinely explain, the explanans must be entirely true, and that the more complete and detailed the explanans is, the better the scientific explanation.

As philosophers of science have turned to more careful examinations of actual scientific practice, however, there have been three key observations that have challenged this traditional approach: first, many of the phenomena scientists seek to explain are incredibly complex; second, the laws of nature supposedly needed for explanation are either few and far between or entirely absent in many of the sciences; and third, a detailed causal description of the chain of events and interactions leading up to a phenomenon are often either beyond our grasp or not in fact what is most important for a scientific understanding of the phenomenon.

More generally, there has been a growing recognition that much of science is a model-based activity.[1] Models are by definition incomplete and idealized descriptions of the systems they describe. This practice raises all sorts of epistemological questions, such as, how can it be that false models lead to true insights? And most relevant to our discussion here, how might the extensive use of models in science lead us to revise our philosophical account of scientific explanation?

4.2 The Explanatory Function of Models: How Model-Based Explanations Work

Model-based explanations (or model explanations, for short) are explanations in which the explanans appeals to certain properties or behaviors observed in an idealized model or computer simulation as part of an explanation for why the (typically real-world) explanandum phenomenon exhibits the features that it does. For example, one might explain why sparrows of a certain species vary in their feather coloration from pale to dark by appealing to a particular game theory model: although coloration is unrelated to fitness, such a polymorphism can be a badge of status that allows the sparrows to avoid unnecessary conflicts over resources; dark birds are dominant and displace the pale birds from food sources. The model demonstrates that such a strategy is stable and successful, and hence can be used as part of the explanation for why we find this polymorphism among sparrows (Maynard Smith 1982 [4]; see Potochnik (manuscript [5]) for further discussion).

[1] For an overview of many different types of models in science, and some of the philosophical issues regarding the nature and use of such models, see Frigg and Hartmann (2012 [3]).

There are of course many perils in assuming that just because we see a phenomenon or pattern exhibited in a model that it therefore explains why we see it in the real world: the same pattern or phenomenon could be produced in multiple, very different ways, and hence it might be only a phenomenological model at best, useful for prediction, but not a genuine explanation. Explanation and the concomitant notion of understanding are what we call success terms: if the purported explanation is not in fact right (right in some sense that will need to be spelled out) and the understanding is only illusory, then it is not in fact a genuine explanation. Determining what the success conditions are for a genuine explanation is the central philosophical problem in scientific explanation.

Those who have defended the explanatory power of models have typically argued that further conditions must be met in order for a model's exhibiting of a salient pattern or phenomenon to count as part of a genuine explanation of its real-world counterpart. Not all models are explanatory, and an adequate account of model explanation must provide grounds for making such discriminations. As we will see, however, different approaches have filled in these further requirements in different ways.

One of the earliest defenses of the view that models can explain is Ernan McMullin's (1978 [6]) "hypothetico-structural" account of model explanations. In a hypothetico-structural (HS) explanation, one explains a complex phenomenon by postulating an underlying structural model whose features are causally responsible for the phenomenon to be explained. McMullin notes that such models are often tentative or metaphorical, but that a good model explanation will lay out a research program for the further refinement of the model. On his account, the justification of the model as genuinely explanatory involves a process known as de-idealization, where features that were left out are added back in or a more realistic representation of those processes is given. More specifically he requires that one be able to give a theoretical justification for this de-idealization process, so that it is not merely an ad hoc fitting of the model to the data. He writes,

> [If] techniques for which no theoretical justification can be given have to be utilized to correct a formal idealization, this is taken to count against the explanatory propriety of that idealization. The model itself in such a case is suspect, no matter how good the predictive results it may produce. (McMullin 1985 [7], p. 261)

He further notes that a theoretical justification for the de-idealization process will only succeed if the original model has successfully captured the real structure of the phenomenon of interest.

As an example, McMullin (1984 [8]) describes the fertility of the continental drift model in explaining why the continents seem to fit together like pieces of a puzzle and why similar fossils are found at distant locations. The continental drift model involved all sorts of idealizations and gaps: most notably, the chief proponent of this approach, Alfred Wegener, could offer no account of the forces or mechanisms by which the massive continents could move. We now know that the continental drift model is strictly speaking false, and has been supplanted by plate tectonics. But as McMullin notes, the continental drift model nonetheless captures key features of the real structure of the phenomenon of interest, and hence succeeds in giving genuine explanatory insight.

While McMullin's account of HS model explanations fits many cases, there are other examples of model explanations in the sciences that do not seem to fit his account. First, there seem to be examples of model explanations where the idealizations are ineliminable, and hence they cannot be justified through anything like the de-idealization analysis that McMullin describes (Batterman 2005a [9]). Second, not all models are related to their target phenomena via an idealization: some models represent through a fictionalization (Bokulich 2009 [10]). Third, insofar as McMullin's HS model explanations are a subspecies of causal explanations, they do not account for non-causal model explanations. These sort of cases will be discussed more fully in subsequent sections.

Another early account of the explanatory power of models is Nancy Cartwright's (1983 [11]) "simulacrum" account of explanation, which she introduces as an alternative to the deductive-nomological (DN) account of explanation and elaborates in her book *How the Laws of Physics Lie*. Drawing on Pierre Duhem's (1914/1954 [12]) theory of explanation, she argues,

> To explain a phenomenon is to find a model that fits it into the basic framework of the theory and that thus allows us to derive analogues for the messy and complicated phenomenological laws which are true of it. (Cartwright 1983 [11], p. 152).

According to Cartwright, the laws of physics do not describe our real messy world, only the idealized world we construct in our models. She gives the example of the harmonic oscillator model, which is used in quantum mechanics to describe a wide variety of systems. One describes a real-world helium-neon laser as if it were a van der Pol oscillator; this is how the phenomenon becomes tractable and we are able to make use of the mathematical framework of our theory. The laws of quantum mechanics are true in this model, but this model is just a simulacrum of the real world phenomenon. By 'model', Cartwright means "a specially prepared, usually fictional description of the system under study" (Cartwright 1983 [11], p. 158). She notes that while some of the properties ascribed to the objects in the models are idealizations, there are other properties that are pure fictions, hence one should not think of models in terms of idealizations alone.

Although Cartwright's simulacrum account is highly suggestive, it leaves unanswered many key questions, such as when a model should or should not be counted as explanatory. Mehmet Elgin and Elliott Sober (2002 [13]) offer a possible emendation to Cartwright's account that they argue discriminates which sorts of idealized causal models can explain. The key, according their approach, is to determine whether or not the idealizations in the model are what they call "harmless." A harmless idealization is one that if corrected "wouldn't make much difference in the predicted value of the effect variable" (Elgin and Sober 2002 [13], p. 448). They illustrate this approach using the example of optimality models in evolutionary biology. Optimality models are models that determine what value of a trait maximizes fitness (is optimal) for an organism given certain constraints (e.g., the optimal length of a bear's fur, given the benefits of longer fur and the costs of growing it, or the optimal height at which crows should drop walnuts in order to crack open the shells, given the costs of flying higher, etc.). If organisms are indeed fitter the closer a trait is to the optimal value, and if natural selection is the only force operating, then the optimal value for that trait will evolve in the population. Thus optimality models are used to explain why organisms have trait values at or near the optimal value (e.g., why crows drop walnuts from an average of 3 meters high (Cristol and Switzer 1999 [14])).

As Elgin and Sober note, optimality models contain all sorts of idealizations: "they describe evolutionary trajectories of populations that are infinitely large in which reproduction is asexual with offspring always resembling their parents, etc." (Elgin and Sober 2002 [13], p. 447). Nonetheless they argue that these models are genuinely explanatory when it can be shown that the value described in the explanandum is close to the value predicted by the idealized model; when this happens we can conclude that the idealizations in the model are harmless (p. 448). Apart from this concession about "harmless" idealizations, Elgin and Sober's account of explanation remains close to the traditional DN account in that they further require (i) the explanans must cite the cause of the explanandum; (ii) the explanans must cite a law; (iii) all of the explanans propositions must be true (p. 446), though their condition (iii) might better be stated as all the explanans propositions are *either* true *or harmlessly false*.

As a general account of model explanations, however, one might argue that the approaches of Cartwright, Elgin and Sober are too restrictive. As noted before, this approach still depends on there being laws of nature from which the phenomenon is to be derived, and such laws just might not be available. Moreover, it is not clear that explanatory models will contain only harmless idealizations. There may very well be cases in which the idealizations make a difference (are not harmless) and yet are essential to the explanation (see, for example, Batterman 2009 [15] and Kennedy 2012 [16]) .

While the simulacrum approach of Cartwright, especially as further developed by Elgin and Sober, largely draws its inspiration from the traditional DN approach to explanation, there are other approaches to model explanation that are tied more closely to the traditional causal-mechanical approach to explanation. Carl Craver (2006 [17]), for example, has argued that models are explanatory when they describe mechanisms. He writes "...the distinction between explanatory and non-explanatory models is that the [former], and not the [latter] describe mechanisms" (p. 367). The central notion of mechanism, here, can be understood as consisting of the various components or parts of the phenomenon of interest, the activities of those components, and how they are organized in relation to each other.

Craver imposes rather strict conditions on when such mechanistic models can be counted as explanatory; he writes, "to characterize the phenomenon correctly and completely is the first restrictive step in turning a model into an acceptable mechanistic explanation" (p.369).[2] Craver analyzes the example of the Hodgkin-Huxley mathematical model of the action potential in an axon (nerve fiber). Despite the fact that this model allowed Hodgkin and Huxley to derive many electrical features of neurons, and the fact that it was based on a number of fundamental laws of physics and chemistry, Craver argues that it was not in fact an explanatory model. He describes it instead as merely a phenomenological model because it failed to accurately describe the details of the underlying mechanism.

A similar mechanistic approach to model explanation has been developed by David Michael Kaplan (2011 [19]), who introduces what he calls the mechanism-model-mapping (or 3M) constraint. He defines the 3M constraint as follows:

> A model of a target phenomenon explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the target mechanism that produces, maintains, or

[2] Some have argued that if one has a complete and accurate description of the system or phenomenon of interest, then it is not clear that one has a model at all, since models are by definition incomplete and, in some respects at least, inaccurate descriptions of the systems they describe (Bokulich 2011 [18]).

> underlies the phenomenon, and (b) the (perhaps mathematical) variables in the model correspond to causal relations among the components of the target mechanism. (Kaplan 2011 [19], p. 347)

Kaplan takes this 3M constraint to provide a demarcation line between explanatory and non-explanatory models. He further notes that "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" (p. 347). Models that do not comply with 3M are rejected as non-explanatory, being at best phenomenological models, useful for prediction, but giving no explanatory insight. In requiring that explanatory models describe the "*real* components and activities in the mechanism that are *in fact* responsible for producing the phenomenon" (Craver 2006 [17], p. 361; Kaplan 2011 [19], p. 353) Craver and Kaplan rule out the possibility that fictional, metaphorical, or strongly idealized models can be explanatory.

One of the most comprehensive defenses of the explanatory power of models is given by Alisa Bokulich (2008a [20], 2008b [21]; 2011 [18]; 2012 [22]), who argues that model explanations such as the three discussed previously (McMullin, Cartwright-Elgin-Sober, and Craver-Kaplan), can be seen as special cases of a more general account of the explanatory power of models. Bokulich's approach draws on James Woodward's counterfactual account of explanation, in which "the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways" (Woodward 2003 [23], p.11). She argues that model explanations typically share the following three features: First, the explanans makes essential reference to a scientific model, which, as is the case with all models, will be an idealized, abstracted, or fictionalized representation of the target system. Second, the model explains the explanandum by showing how the elements of the model correctly capture the patterns of counterfactual dependence in the target system, enabling one to answer a wide-range of what Woodward calls "what-if-things-had-been-different" questions. Finally, there must be what Bokulich calls a "justificatory step", specifying the domain of applicability of the model and showing where and to what extent the model can be trusted as an adequate representation of the target for the purpose(s) in question (Bokulich 2011 [18], p. 39; see also Bokulich 2012 [22], p. 730). She notes that this justificatory step can proceed bottom-up through something like a de-idealization analysis (as McMullin, Elgin and Sober describe), top-down through an overarching theory (such as in the semiclassical mechanics examples Bokulich (2008a [20], 2008b [21]) discusses), or through some combination.

Arguably one of the advantages of Bokulich's approach is that it is not tied to one particular conception of scientific explanation, such as the DN or mechanistic accounts. By relaxing Woodward's manipulationist construal of the counterfactual condition, Bokulich's approach can even be extended to highly abstract, structural, or mathematical model explanations. She argues that the various "subspecies" of model explanation can be distinguished by noting what she calls the "origin" or ground of the counterfactual dependence. She explains, it could be either

> the elements represented in the model *causally producing* the explanandum (in the case of causal model explanations), the elements of the model *being the mechanistic parts which make up* the explanandum-system whole (in the case of mechanistic model explanations), or the explanandum being a consequence of the laws cited in the model (in the case of covering law model explanations). (Bokulich 2011 [18], p. 40).

She goes on to identify a fourth type of model explanation, which she calls structural model explanation, in which the counterfactual dependence is grounded in the typically mathematical structure of the theory, which limits the sorts of objects, properties, states, or behaviors that are admissible within the framework of that theory (Bokulich 2011 [18], p. 40). Bokulich's approach can be thought of as one way to flesh out Margaret Morrison's suggestive, but unelaborated, remark that "the reason models are explanatory is that in representing these systems, they exhibit certain kinds of structural dependencies" (Morrison 1999 [24], p. 63).

More recently Collin Rice (forthcoming [25]) has drawn on Bokulich's account to develop a similar approach to the explanatory power of models that likewise uses Woodward's counterfactual approach without the manipulation condition. He writes,

> The requirement that these counterfactuals must enable one to, in principle, *intervene* in the system restricts Woodward's account to specifically causal explanations. However, I think it is a mistake to require that all scientific explanations must be causal. Indeed, if one looks at many of the explanations offered by scientific modelers, causes are not mentioned. (Rice forthcoming [25], p. 20)[3]

Rice rightly notes that the question of causation is conceptually distinct from the question of what explains. He further requires on this approach that model explanations provide two kinds of counterfactual information, namely both what the phenomenon depends on and what sorts of changes are irrelevant to that phenomenon. Following Robert Batterman (2002 [26], 2005a [9], 2009 [15]), he notes that for explanations of phenomena that exhibit a kind of universality, an important part of the explanation is understanding

[3] Compare this to Bokulich's statement "I think it is a mistake to construe all scientific explanation as a species of causal explanation, and more to the point here, it is certainly not the case that all model explanations should be understood as causal explanations. Thus while I shall adopt Woodward's account of explanation as the exhibiting of a pattern of counterfactual dependence, I will not construe this dependence narrowly in terms of the possible causal manipulations of the system" (Bokulich 2011 [18], p. 39).

that the particular causal details or processes are irrelevant--the same phenomenon would have been reproduced even if the causal details had been different in certain ways.

As an illustration, Rice discusses the case of optimality modeling in biology. He notes that optimality models are not only highly idealized, but also can be understood as a type of equilibrium explanation, where "most of the explanatory work in these models is done by *synchronic mathematical representations of structural features of the system*" (Rice forthcoming [25], p. 8). He connects this to the counterfactual account of model explanation as follows:

> Optimality models primarily focus on noncausal counterfactual relations between structural features and the system's equilibrium point. Moreover, these features can sometimes explain the target phenomenon without requiring any additional causal claims about the relationships represented in the model." (Rice forthcoming [25], p. 17)

These causal details are irrelevant because the structural features cited in the model are multiply realizable, indeed this is what allows optimality models to be used in explaining a wide variety of features across a diversity of biological systems.

In the approaches to model explanations discussed here, two controversial issues have arisen that merit closer scrutiny: First, whether the fictions or falsehoods in models can themselves do real explanatory work (that is, even when they are neither "harmless", "de-idealizable", nor eliminable), and second, whether many model explanations illustrate an important, but often overlooked, non-causal form of explanation. These issues will be taken up in turn in the next two sections.

4.3 Explanatory Fictions: Can Falsehoods Explain?

Models contain all sorts of falsehoods, from omissions, abstractions, and idealizations to outright fictions. One of the most controversial issues in model explanations is whether these falsehoods, which are inherent in the modeling practice, are compatible with the explanatory aims of science. Julian Reiss in the context of explanatory models in economics has called this tension the "explanation paradox": he writes,

> [T]hree mutually inconsistent hypotheses concerning models and explanation are widely held: (1) economic models are false; (2) economic models are nevertheless explanatory; and (3) only true accounts explain. Commentators have typically resolved the paradox by rejecting either one of these hypotheses. I will argue that none of the proposed resolutions work and conclude that therefore the paradox is genuine and likely to stay. (Reiss 2012 [27], p. 43)[4]

The field has largely split into two camps on this issue: those who think it is only the true parts of models that do explanatory work, and those who think the falsehoods play an essential role in the model explanation. Those in the former camp rely on things like “de-idealization” and “harmless” analyses to show that the falsehoods do not get in the way of the true parts of the model that do the real explanatory work. Those in the latter camp have the challenging task of showing that some idealizations are essential and some fictions yield true insights.

The "received view" is that the false parts of models only concern those things that are explanatorily irrelevant. Defenders of the received view include Michael

[4] This paradox, and some criticisms to Reiss's approach (such as Mäki 2013 [28]) are explored in a special issue of the journal *Journal of Economic Methodology* (volume 20, issue 3).

Strevens, who in his book detailing his kairetic[5] account of scientific explanation, writes, “No causal account of explanation--certainly not the kairetic account--allows nonveridical models to explain" (Strevens 2008 [29], p. 297). He spells out more carefully how such a view is to be reconciled with the widespread use of idealized models to explain phenomena in nature, by drawing the following distinction:

> The content of an idealized model, then, can be divided into two parts. The first part contains the difference-makers for the explanatory target. . . . The second part is all idealization; its overt claims are false but its role is to point to parts of the actual world that do not make a difference to the explanatory target. (Strevens 2008 [29], p. 318)

In other words, it is only the true parts of the model that do any explanatory work. The false parts are harmless, and hence should be able to be de-idealized away without affecting the explanation.

On the other side, a number of scholars have argued for the counterintuitive conclusion that sometimes it is in part *because* of their falsehoods--not despite them--that models explain. Robert Batterman (2002 [26], 2005a [9], 2009 [15]), for example, has argued that some idealizations are explanatorily ineliminable, that is, the idealizations or falsehoods themselves do real explanatory work. Batterman considers continuum model explanations of phenomena such shocks (e.g., compressions traveling through a gas in a tube) and breaking drops (e.g., the shape of water as it drips from a faucet). In order to explain such phenomena, scientists make the idealization that the gas or fluid is a continuum (rather than describing it veridically as a collection of discrete gas or water molecules). These false continuum assumptions are essential for obtaining the desired

[5] Strevens takes the term kairetic from the ancient Greek word *kairos*, meaning crucial moment (Strevens 2008 [28], p. 477).

explanation. In the breaking drops case, it turns out that different fluids of different viscosities dripping from faucets of different widths will all exhibit the same shape upon breakup. The explanation depends on a singularity that exists only in the (false) continuum model; such an explanation does not exist on the de-idealized molecular dynamics approach (Batterman 2009 [15], pp. 442-443). Hence, he concludes, "continuum idealizations are explanatorily ineliminable and . . . a full understanding of certain physical phenomena cannot be obtained through completely detailed, non-idealized representations" (Batterman 2009 [15], 427). If such analyses are right, then they show that not all idealizations can be de-idealized, and moreover, those falsehoods can play an essential role in the explanation.

Alisa Bokulich (2008a [20], 2008b [21]; 2009 [10]; 2012 [22]) has similarly defended the view that it is not just the true parts of models that can do explanatory work, arguing that in some cases even fictions can be explanatory. She writes, "some fictions can give us genuine insight into the way the world is, and hence be genuinely explanatory and yield real understanding" (Bokulich 2009 [10], p. 94). She argues that some fictions are able to do this by capturing in their fictional representation real patterns of structural dependencies in the world. As an example, she discusses semiclassical models whereby fictional electron orbits are used to explain peculiar features of quantum spectra. Although, according to quantum mechanics, electrons do not follow definite trajectories or orbits (i.e., such orbits are fictions), physicists recognized that puzzling peaks in the recurrence spectrum of atoms in strong magnetic fields have a one-to-one correspondence with particular closed classical orbits:

> The resonances . . . form a series of strikingly simple and regular organization, not previously anticipated or predicted. . . . The regular type resonances can be

> physically rationalized and explained by classical periodic orbits of the electron on closed trajectories starting at and returning to the proton as origin. (Main et al. 1986 [30], pp. 2789-2790, quoted in Bokulich 2009 [10], p. 99)

As she explains, at no point are these physicists challenging the status of quantum mechanics as the true, fundamental ontological theory; rather, they are deploying the fiction with the express recognition that it is indeed a literally false representation.[6] Nonetheless it is a representation that is able to yield true physical insight and understanding by carefully capturing in its fictional representation the appropriate patterns of counterfactual dependence of the target phenomenon.

Bokulich (2008a [20], 2008b [21]; 2009 [10]; 2012 [22]) offers several such examples of explanatory fictional models from semiclassical mechanics, where the received explanation of quantum phenomena appeals to classical structures, such as the Lyapunov (stability) exponents of classical trajectories, that have no clear quantum counterpart. Moreover, she notes that these semiclassical models with their fictional assumption of classical trajectories are valued not primarily as calculation tools (often they require calculations that are just as complicated), but rather are valued as models that provide an unparalleled level of physical insight into the structure of the quantum phenomena. Bokulich is careful to note that not just any fiction can do this kind of explanatory work, indeed most fictions cannot. She shows more specifically how these semiclassical examples meet the three criteria of her account of model-based explanation, discussed above (see, for example, Bokulich 2009 [10], p. 106).

[6] Interestingly this was one of the Hans Vaihinger's criteria for a *scientific* fiction, namely that there must be "an express awareness that the fiction is just a fiction" (Vaihinger 1911/1952 [31], p. 98).

A more pedestrian example of an explanatory fiction, and one that brings out some of the objections to such claims, is the case of light rays postulated by the ray (or geometrical) theory of optics. Light rays are strictly speaking, a fiction. The currently accepted fundamental theory of wave optics denies that they exist. Yet light rays seem to play a central role in the scientific explanation of lots of phenomena, such as shadows and rainbows. The physicists Dan Kleppner and John Delos, for example, note, "When one sees the sharp shadows of buildings in a city, it seems difficult to insist that light-rays are merely calculational tools that provide approximations to the full solution of the wave equation" (Kleppner and Delos 2001 [32], p. 610). Similarly, Batterman argues, "one cannot explain various features of the rainbow (in particular, the universal patterns of intensities and fringe spacings) without ultimately having to appeal to the structural stability of ray theoretic structures called caustics—focal properties of families of rays" (Batterman 2005b [33], pp. 154-155). Batterman is quite explicit that he does not think that an explanatory appeal to these ray-theoretic structures requires reifying the rays; they are indeed fictions.

Some, such as Gordon Belot, want to dismiss ray optics models as nothing but a mathematical device devoid of any physical content outside of the fundamental (wave) theory. He writes,

> The mathematics of the less fundamental theory is definable in terms of that of the more fundamental theory; so the requisite mathematical results can be proved by someone whose repertoire of interpreted physical theories included only the latter. (Belot 2005 [34], p. 151)

The point is roughly this: it looks like in Batterman's examples that one is making an explanatory appeal to fictional entities from a "less fundamental" theory that has been superseded (e.g., ray optics or classical mechanics). However, all one needs from that superseded theory is the mathematics--one doesn't need to give those bits of mathematics a physical interpretation in terms of the fictional entities or structures. Moreover, that mathematics appears to be definable in terms of the mathematics of the true "fundamental" theory. Hence, those fictional entities are not in fact playing an explanatory role.

Batterman has responded to these objections, arguing that in order to have an explanation, one does in fact need the fictional physical interpretation of that mathematics, and hence the explanatory resources of the non-fundamental theory. He explains,

> Without the physical interpretation to begin with, we would not know *what* boundary conditions to join to the differential equation. Neither, would we know *how* to join those boundary conditions to the equation. Put another way, we must examine the physical details of the *boundaries* (the shape, reflective and refractive details of the drops, etc.) in order to set up the *boundary conditions* required for the mathematical solution to the equation. (Batterman 2005b [33], p. 159)

In other words, without appealing to the fictional rays we would not have the relevant information we need to appropriately set up and solve the mathematical model that is needed for the explanation.

In a paper with Lina Jansson, Belot has raised similar objections against Bokulich's arguments that classical structures can play a role in explaining quantum phenomena. They write,

> Bokulich and others see explanations that draw on semiclassical considerations as involving elements of classical physics as well as of quantum physics. . . . But there is an alternative way of thinking of semiclassical mechanics: . . . starting with the formalism of quantum mechanics one proves theorems about approximate solutions--theorems that happen to involve some of the mathematical apparatus of classical mechanics. But this need not tempt us to think that there is [classical] physics in our explanations. (Belot and Jansson 2010 [35], p. 82)

Once again we see the objection that it is just the bare mathematics, not the mathematics with its physical interpretation that is involved in the explanation. On Bokulich's view, however, it is precisely by connecting that "mathematical apparatus" to its physical interpretation in terms of classical mechanics, that one gains a deeper physical insight into the system one is studying. On her view, explanation is importantly about advancing understanding, and for this the physical interpretation is important.[7] Even though classical mechanics is not the true fundamental theory, there are important respects in which it gets things right, and hence reasoning with fictional classical structures within the well-established confines of semiclassical mechanics, can yield explanatory insight and deepen our understanding.

As we have seen, these claims that fictions can explain (in special cases such as ray optics and classical structures) remain controversial and involve subtle issues. These debates are not entirely new, however, and they have some interesting historical antecedents, for example, in the works of Niels Bohr and James Clerk Maxwell. More specifically, when Bohr is articulating his widely misunderstood "correspondence principle",[8] he argues that one can explain why only certain quantum transitions between stationary states in atoms are allowed by appealing to which harmonic components appear in the Fourier decomposition of the electron's classical orbit (see Bokulich 2008a

[7] Potochnik (forthcoming [5], Chapter 5) has also argued for a tight connection between explanation and understanding, responding to some of the traditional objections against this association. More broadly she emphasizes the communicative function of explanation over the ontological approach to explanation, which makes more room for non-veridical model explanations than the traditional approach.

[8] For an accessible discussion of the various interpretations (and misinterpretations) of the correspondence principle see Bokulich's (2010 [36]) entry on the correspondence principle for the online *Stanford Encyclopedia of Philosophy*.

[20], Section 4.2 and references therein). He does this even long after he has conceded to the new quantum theory that classical electron trajectories in the atom are impossible (i.e., they are a fiction). Although Heisenberg used this formulation of the correspondence principle to construct his matrix mechanics, he argued that "[it] must be emphasized that this [correspondence] is a purely formal result" (Heisenberg 1930 [37], p. 83), and should not be thought of as involving any physical content from the other theory. Bohr, by contrast, was dissatisfied with this interpretation of the correspondence principle as "pure mathematics", arguing instead that it revealed a deep *physical* connection between classical and quantum mechanics. Even earlier, we can see some of these issues arising in the work of Maxwell, who, in exploiting the utility of fictional models and physical analogies between disparate fields, argued, "My aim has been to present the mathematical ideas to the mind in an embodied form . . . not as mere symbols, which convey neither the same ideas, nor readily adapt themselves to the phenomena to be explained" (Maxwell 1855/1890 [38], p. 187; for a discussion see Bokulich 2015 [39]).

Three other challenges have been raised against the explanatory power of fictional models. First, there is a kind of slippery-slope worry, that once we admit some fictional models as explanatory, we will not have any grounds on which to dismiss other fictional models as nonexplanatory. Bokulich (2012 [22]) in her paper "Distinguishing Explanatory from Nonexplanatory Fictions" introduces a framework for addressing this problem. Second, Samuel Schindler (2014 [40]) has raised what he sees as a tension in Bokulich's account. He claims that on one hand she says semiclassical explanations of quantum phenomena are autonomous in the sense that they provide more insight than the quantum mechanical ones. Yet on the other hand, she notes that semiclassical models are justified through semiclassical theory, which connects these representations as a kind of approximation to the full quantum mechanics. Hence, they cannot be autonomous. This objection seems to trade on an equivocation of the term 'autonomous': in the first case 'autonomous' is used to mean "a representation of the phenomenon that yields more physical insight" and in the second case 'autonomous' is used to mean "cannot be mathematical justified through various approximation methods." These seem to be two entirely different concepts, and hence not really in tension with each other. Moreover, Bokulich never uses the term 'autonomous' to describe either, so this seems to be a misleading reading of her view.

Schindler also rehearses the objection, raised by Belot and Jansson (2010 [35]), that by eliminating the interventionist condition in Woodward's counterfactual approach to explanation she loses what he calls "the asymmetry-individuating function," by which he means her account seems susceptible to the traditional problem of asymmetry that plagued the DN account of explanation (for example that falling barometers could be used to explain impending storms or shadows could used to explain the height of flag poles, to recall Sylvain Bromberger's well-known examples). This problem was taken to be solved by the causal approach to explanation, whereby one secures the explanatory asymmetry simply by appealing to the asymmetry of causation. It is important to note, however that this is not an objection specifically to Bokulich's account of structural model explanation, but rather is a challenge for any noncausal account of explanation.[9]

[9] Bokulich outlines a solution to the problem of asymmetry for her account in Bokulich (2012 [22]).

Since many examples of explanatory models purport to be non-causal explanations, we will examine this topic more fully in the next section.

Another context in which this issue about the explanatory power of fictional models arises is in connection with cognitive models in psychology and cognitive neuroscience. Daniel Weiskopf, for example, discusses how psychological capacities are often understood in terms of cognitive models that functionally abstract from the underlying real system. More specifically, he notes, "In attempting to understand the high level dynamics of complex systems like brains, modelers have recourse to many techniques for constructing such indirect accounts . . . *reification, functional abstraction,* and *fictionalization*" (Weiskopf 2011 [41], p. 328). By reification he means "positing something with the characteristics of a more or less stable and enduring object, where in fact no such thing exists" (p. 328). He gives as an example the positing of symbolic representations in classical computational systems, even though he notes that nothing in the brain seems to 'stand still' or be manipulable in the way symbols do. Functional abstraction, he argues occurs when we "decompose a modeled system into subsystems and other components on the basis of what they do, rather than their correspondence with organizations and groupings in the target system" (p. 329). He notes that this occurs when there are cross-cutting functional groupings that don't map onto the structural or anatomical divisions of the brain. He notes that this strategy emphasizes "networks, not locations" in relating cognition to neural structures. Finally, there is also fictionalization, which, as he describes, "involves putting components into a model that are known not to correspond to any element of the modeled system, but which serve an essential role in getting the models to operate correctly" (Weiskopf 2011 [41], p. 331). He gives as an example of a fiction in cognitive modeling what are called 'Fast Enabling Links' (FELs), which are independent of the channels by which cells actually communicate and are assumed to have functionally infinite propagation speeds, allowing two cells to fire in synchrony (p. 331). Despite being false in these ways, some modelers take these fictions to be essential to the operation of the model and not likely to be eliminated in future versions.

Weiskopf concludes that models involving reifications, functional abstractions, and fictions, can nonetheless in some cases succeed in "meeting the general normative constraints on explanatory models perfectly well" (p. 332), and hence such models can be counted as genuinely explanatory. Although Weiskopf recognizes the many great successes of mechanistic explanations in biological and neural systems, he wants to resist an "imperialism" that attempts to reduce all cases of model explanations in these fields to mechanistic model explanations.

More recently Cameron Buckner (forthcoming [42]) has criticized Weiskopf's arguments that functionalist models involving fictions, abstractions, and reification can be explanatory and defended the mechanist's maxim (e.g., as articulated by Craver and Kaplan) that only mechanistic models can genuinely explain. Buckner employs two strategies in arguing against Weiskopf: first, in cases where the models do explain, he argues that they are really just mechanism sketches, and where they cannot be reconstructed mechanistically, he dismisses them as impoverished explanations. He writes,

> Concerning fictionalization and reification, I concede that models featuring such components cannot be interpreted as mechanism sketches, but argue that interpreting their nonlocalizable components as natural kinds comes with clear costs in terms of those models' counterfactual power. . . . Functional abstraction,

> on the other hand, can be considered a legitimate source of kinds, but only on the condition that the functionally abstract models be interpreted as sketches that could be elaborated into a more complete mechanistic model. (Buckner forthcoming [42], p. 3)

An essential feature of mechanistic models seems to be that their components are localizable. Weiskopf argues, however, that his functional kinds are multiply realizable, that is, they apply to many different kinds of underlying mechanisms, and that in some cases they are distributed in the sense that they ascribe to a given model component capacities that are distributed amongst distinct parts of the physical system. Hence, without localization, such models cannot be reconstructed as mechanistic models.

What of Buckner's claim that fictional models will be impoverished with regard to their counterfactual power? Consider again Weiskopf's example of the fictional FELs, which are posited in the model to allow the cells to achieve synchrony. Buckner argues explanations involving models with FELs are impoverished in that if one had a true account of synchrony, that model explanation would support *more* counterfactual knowledge. It is not clear, however, that this objection undermines the explanatory power of models involving FELs per se; rather it seems only to suggest that if we knew more and had the true account of synchrony we might have a *deeper* explanation[10] (at least on the assumption that this true account of synchrony would allow us to answer a wider range of what-if-things-had-been-different questions). However, the explanation involving the fiction might still be perfectly adequate for the purpose for which it is being deployed, and hence it need not even be counted as impoverished. For example, there might be some explananda (ones other than the explanadum of "how do cells achieve

[10] For an account of explanatory depth, see Hitchcock and Woodward (2003 [43]).

synchrony") for which it simply doesn't matter *how* cells achieve synchrony; the fact that they *do* achieve synchrony might be all that is required for some purposes.

Weiskopf is not alone in trying to make room for non-mechanistic model explanations; Elizabeth Irvine (forthcoming [44]) and Lauren Ross (2015 [45]) have also recently defended non-mechanistic model explanations in cognitive science and biology. Their approaches argue for non-causal forms of model explanation, which we will turn to next.

4.4 Explanatory Models and Non-Causal Explanations

Recently there has been a growing interest in non-causal forms of explanation. Similar to Bokulich's (2008a [20], 2008b [21]) approach, many of these seek to understand non-causal explanations within the context of Woodward's (2003 [23]) counterfactual approach to explanation without the interventionist criterion that restricts his account specifically to causal explanation (e.g., Saatsi and Pexton 2013 [46] and Rice forthcoming [25]). Non-causal explanations are usually defined negatively as explaining by some means *other than* citing causes, though this is presumably a heterogeneous group. We have already seen one type of non-causal model-based explanation: Bokulich’s (2008a [20], 2008b [21]) structural model explanations in physics. More recently, examples have been given in fields ranging from biology to cognitive science. Highly mathematical model explanations are another type of non-causal explanation, though not all mathematical models are non-causal. A few recent examples are considered here.

In the context of biology and cognitive science, Elizabeth Irvine (forthcoming [44]) has argued for the need to go beyond the causal-mechanical account of model explanation and defends what she calls a non-causal structural form of model explanation. She focuses specifically on reinforcement learning (RL) models in cognitive science and optimality models in biology. She notes that although RL and optimality models can be construed as providing causal explanations in some contexts, there are other contexts in which causal explanations miss the mark. She writes,

> In the account developed here, it is not the presence of idealisation or abstraction in models that is important, nor the lack of description of causal dynamics or use of robustness analyses to test the models. Instead, it is the bare fact that some models and target systems have equilibrium points [that] are highly O-robust with respect to initial conditions and perturbations. . . .This alone can drive a claim about non-causal structural explanations. (Irvine forthcoming [44], p. [11])

By O-robustness, Irvine means a robust convergence to an optimal state across a range of interventions, whether it be an optimization of fitness or an optimization of decision making strategies. Her argument is that since interventions (in the sense of Woodward) don't make a difference to the convergence on the optimal state, that convergence cannot be explained causally, and is instead due to structural features of the model and target system it explains.

Another recent approach to non-causal model explanation is Batterman and Rice's (2014 [47]) minimal model explanations. Minimal models are models that explain patterns of macroscopic behavior for systems that are heterogeneous at smaller scales. Batterman and Rice discuss two examples of minimal models in depth: the Lattice Gas Automaton model, which is used to explain large-scale patterns in fluid flow, and Fisher's Sex Ratio model, which is used to explain why one typically finds a 1:1 ratio of males to females, across diverse populations of species. In both cases, they argue,

> these minimal models are explanatory because there is a detailed story about why the myriad details that distinguish a class of systems are irrelevant to their large-scale behavior. This story demonstrates, rather than assumes, a kind of stability or robustness of the large-scale behavior we want to explain under drastic changes in the various details of the system. (Batterman and Rice 2014 [47], p. 373)

They make two further claims about these minimal model explanations. First, they argue that these explanations are "distinct from various causal, mechanical, difference making, and so on, strategies prominent in the philosophical literature" (Batterman and Rice 2014 [47], p. 349). Second, they argue that the explanatory power of minimal models cannot be accounted for by any kind of mirroring or mapping between the model and target system (what they call the "common features" account). Instead, these non-causal explanations work by showing that the minimal model and diverse real-world systems fall into the same universality class. This latter claim has been challenged by Marc Lange (2015 [48]) who, though sympathetic to their claim that minimal models are a non-causal form of model explanation, argues that their explanatory power does in fact derive from the model sharing features in common with the diverse systems it describes (i.e., the "common features" account Batterman and Rice reject).

Lauren Ross (2015 [45]) has applied the minimal models account to dynamical model explanations in the neurosciences. More specifically she considers as an explanandum phenomenon the fact that a diverse set of neural systems (e.g., rat hippocampal neurons, crustacean motor neurons, and human cortical neurons)[11], which are quite different at the molecular level, nonetheless all exhibit the same "type I" excitability behavior. She shows that the explanation for this involves applying mathematical abstraction techniques to the various detailed models of each particular

[11] These are examples given by Ross (2015 [45], p. 48).

type of neural system and then showing that all these diverse systems converge on one and the same canonical model (known as the Ermentrout-Kopell model). After defending the explanatory power of these canonical models, Ross then contrasts this kind of non-causal model explanation with the causal-mechanical model approach:

> The canonical model approach contrasts with Kaplan and Craver's claims because it is used to explain the shared behavior of neural systems without revealing their underlying causal mechanical structure. As the neural systems that share this behavior consist of differing causal mechanisms . . . a mechanistic model that represented the causal structure of any single neural system would no longer represent the entire class of systems. (Ross 2015 [45], p. 46)

Her point is that a non-causal explanation is called for in this case because the particular causal details are irrelevant to the explanation of the universal behavior of class I neurons. The minimal models approach, as we saw above, is designed precisely to capture these sort explanations involving universality.

More generally, many highly abstract or highly mathematical model explanations also seem to fall into this general category of non-causal model explanations. Christopher Pincock, for example, identifies a type of explanation that he calls "abstract explanation", which could be extended to model-based explanations. He writes "the best recent work on causal explanation is not able to naturally accommodate these abstract explanations" (Pincock forthcoming [49] , p. 11). Although some of the explanations Pincock cites, such as the topological (graph theory) explanation for why one cannot cross the seven bridges of Königsberg exactly once in a non-backtracking circuit, seem to be genuinely non-causal explanations, it is not clear that all "abstract" explanations are necessarily non-causal. Alexander Reutlinger and Holly Andersen (manuscript [50]) have recently raised this objection against Pincock's account, arguing that an explanation's being abstract is not a sufficient condition for it being non-causal. They argue that many causal explanations can be abstract too and so more work needs to be done identifying what makes an explanation truly non-causal. This is a particularly pressing issue in model-based explanations, since many scientific models are abstract in this sense of leaving out microphysical or concrete causal details about the explanandum phenomenon.

Marc Lange (2013 [51]) has also identified a kind of non-causal explanation that he calls a "distinctively mathematical" explanation. Lange considers a number of candidate mathematical explanations, such as why one cannot divide twenty-three strawberries evenly among three children, why cicadas have life-cycle periods that are prime, and why honeybees build their combs on a hexagonal grid. Lange notes that whether these are to count as distinctively mathematical explanations depends on precisely how one construes the explanandum phenomenon. If we ask why honeybees divide the honeycomb into hexagons, rather than other polygons, and we cite that it is selectively advantageous for them to minimize the wax used, together with the mathematical fact that a hexagonal grid has the least total perimeter, then it is an ordinary causal explanation (it works by citing selection pressures). If, however, "we narrow the explanandum to the fact that in any scheme to divide their combs into regions of equal area, honeybees would use at least the amount of wax they would use in dividing their combs into hexagons. . . [t]his fact has a distinctively mathematical explanation" (Lange 2013 [50], p. 500). As Lange explains more generally,

> These explanations are non-causal, but this does not mean that they fail to cite the explanandum's causes, that they abstract away from detailed causal histories, or that they cite no natural laws. Rather, in these explanations, the facts doing the explaining are modally stronger than ordinary causal laws. (Lange 2013 [51], p. 485)

The key issue is not whether the explanans cites the explanandum's causes, but whether the explanation works *by virtue of* citing those causes. Distinctively mathematical (non-causal) explanations show the explanandum to be necessary to a stronger degree than would result from the causal powers alone.

As this literature makes clear, distinguishing causal from non-causal explanations is a subtle and open problem, but one crucial for understanding the wide-spread use of abstract mathematical models in many scientific explanations.

4.5 How-Possibly vs. How-Actually Model Explanations

Models and computer simulations can often generate patterns or behaviors that are strikingly similar to the phenomenon to be explained. As we have seen, however, that is typically not enough to conclude that the model thereby explains the phenomenon. An important distinction here is that between a 'how-possibly' model explanation and a 'how-actually' model explanation.

The notion of a how-possibly explanation was first introduced in the 1950s by William Dray in the context of explanations in history. Dray conceived of how possibly explanations as a rival to the DN approach, which he labeled 'why-necessarily' explanations (Dray 1957 [52], 161). Dray interpreted how-possibly explanations as ones that merely aim to show why a particular phenomenon or event "need not have caused surprise" (p. 157), hence they are answers to a different kind of question and can be considered complete explanations in themselves. Although Dray's approach was influential, subsequent authors have interpreted this distinction in different ways. Robert Brandon, in the context of explanations in evolutionary biology, for example writes,

> A how-possibly explanation is one where one or more of the explanatory conditions are speculatively postulated. But if we gather more and more evidence for the postulated conditions, we can move the how-possibly explanation along the continuum until finally we count it as a how-actually explanation. (Brandon 1990 [53], p. 184).

On this view the distinction is a matter of the degree of confirmation, not a difference of kind: as we get more evidence that the processes cited in the model are the processes operating in nature, we move from a how-possibly to how-actually explanation.

Patrick Forber (2010 [54]), however, rejects this interpretation of the distinction as marking a degree of empirical support, and instead defends Dray's original contention that they mark different kinds of explanations. More specifically Forber distinguishes two kinds of how-possibly explanations that he labels "global how-possibly" and "local how possibly" explanations:

> The global how-possibly explanations have theory, mathematics, simulations, and analytical techniques as the resources for fashioning such explanations. . . . The local how-possibly explanations draw upon the models of evolutionary processes and go one step further. They speculate about the biological possibilities relative to an information set enriched by the specific biology of a target system. . . . How-actually explanations, carefully confirmed by empirical tests, aim to identify the correct evolutionary processes that did, in fact, produce the target outcome. (Forber 2010 [54], p. 35)

Although Forber's distinction is conceptually helpful, it is not clear whether global versus local how-possibly explanations should in fact be seen as two distinct categories, rather than simply two poles of a spectrum.

Carl Craver draws a distinction between how-possibly models and how-actually models that is supposed to track the corresponding two kinds of explanations. He notes that how-possibly models purport to explain (unlike phenomenological models, which do not purport to explain), but they are only loosely constrained conjectures about the mechanism. How-actually models, by contrast, describe the detailed components and activities that in fact produce the phenomenon. He writes, "How-possibly models are . . . not adequate explanations. In saying this I am saying not merely that the description must be true (or true enough) but further, that the model must correctly characterize the details of the mechanism" (Craver 2006 [17], p. 361). Craver seems to see the distinction resting not just on the degree of confirmation (truth) but also on the degree of detail.

Bokulich (2014 [55]) defends another construal of the how-possibly/how-actually distinction and applies it to model-based explanations more specifically. She considers, as an example, model-based explanations of a puzzling ecological phenomenon known as tiger bush. Tiger bush is a striking periodic banding of vegetation in semi-arid regions, such as southwest Niger. A surprising feature of tiger bush is that it can occur for a wide variety of different kinds of plants and soils, and it is not induced by any local heterogeneities or variations in topography. By tracing how scientists use various idealized models (e.g., Turing models or differential flow models) to explain phenomena such as this, Bokulich argues new insight into the how-possibly/how-actually distinction can be gained.

The first lesson she draws is that there are different levels of abstraction at which the explanandum phenomenon can be framed, which correspond to different explanatory contexts (p. 33). These different explanatory contexts can be clarified by considering the relevant contrast class of explanations.[12] Second, she argues *pace* Craver that the how-possibly/how-actually distinction does not track how detailed the explanation is. She explains,

[12] For a discussion of contrast classes and their importance in scientific explanation, see van Fraassen (1980 [56], Chapter 5).

> It is not the amount of detail that is relevant, but rather whether the mechanism represented in the model is the mechanism operating in nature. Indeed as we saw in the tiger bush case, the more abstractly the explanatory mechanism is specified, the easier it is to establish it as a how-actually explanation; whereas the more finely the explanatory mechanism is specified, the less confident scientists typically are that their particular detailed characterization of the mechanism is the actual one. (Bokulich 2014 [55], p. 334)

Hence, somewhat counterintuitively, model explanations at a more fine-grained level are more likely to be how-possibly model explanations, even when they are nested within a higher-level how-actually model explanation of a more abstract characterization of the phenomenon. She concludes that when assessing model explanations it is important to pay attention to what might be called the scale of resolution at which the explanandum phenomenon is being framed in a particular explanatory context.

4.6 Tradeoffs in Modeling: Explanation vs. Other Functions for Models

Different scientists will often create different models of a given phenomenon, depending on their particular interests and aims. Following Ron Giere we might note that "[t]here is no *best* scientific model of anything; there are only models more or less good for different purposes" (Giere 2001 [57], p. 1060). If this is right, then it raises the following questions: What are the features that make a model particularly good for the purpose of explanation? Are there tradeoffs between different modeling aims, such that if one optimizes a model for explanation, for example, then that model will fail to be optimized for some other purpose, such as prediction?

One of the earliest papers to explore this theme of tradeoffs in modeling is Richard Levins' paper "The Strategy of Model Building in Population Biology." Levins writes,

> It is of course desirable to work with manageable models which maximize generality, realism, and precision toward the overlapping but not identical goals of understanding, predicting, and modifying nature. But this cannot be done. (Levins 1966 [58], p. 422)

Levins then goes on to describe various modeling strategies that have evolved among modelers, such as sacrificing realism to generality and precision, or sacrificing precision to realism and generality. Levins in his own work on models in ecology favored this latter strategy, where he notes his concern was primarily qualitative not quantitative results, and he emphasizes the importance of robustness analyses in assessing these models.

Although Levins's arguments have not gone unchallenged, John Matthewson and Michael Weisberg have recently defended the view that some tradeoffs in modeling are genuine. They focus on precision and generality, given the relevance of this tradeoff to the aim of explanatory power. After a technical demonstration of different kinds of tradeoffs between two different notions of generality and precision, they conclude,

> These accounts all suggest that increases in generality are, ceteris paribus, associated with an increase in explanatory power. The existence of tradeoffs between precision and generality indicates that one way to increase an explanatorily valuable desideratum is by sacrificing precision. Conversely, increasing precision may lead to a decrease in explanatory power via its effect on generality. (Matthewson and Weisberg 2009 [59], p. 189)

Mapping out the various tensions and tradeoffs modelers may face in developing models for various aims, such as scientific explanation, remains a methodologically important, though underexplored topic.

More recently, Alisa Bokulich (2013 [60]) has explored such tradeoffs in the context of modeling in geomorphology, which is the study of how landscapes and coastlines change over time. Even when it comes to a single phenomenon, such as braided rivers (i.e., rivers in which there is a number of interwoven channels and bars that dynamically shift over time), one finds that scientists use different kinds of models depending on whether their primary aim is explanation or prediction. When they are interested explaining why rivers braid geomorphologists tend to use what are known as “reduced complexity models”, which are typically very simple cellular automata models with a highly idealized representation of the fluvial dynamics (Murray 2003 [61]). The goal is to try to abstract away and isolate the key mechanisms responsible for the production of the braided pattern. This approach is contrasted with an alternative approach to modeling in geomorphology known as ‘reductionist’ modeling. Here one tries to simulate the braided river in as much accurate detail and with as many different processes included as is computationally feasible, and then tries to solve the relevant Navier-Stokes equations in three dimensions. These reductionist models are the best available tools for predicting the features of braided rivers (Murray 2003 [61], p. 159), but they are so complex that they yield very little insight into *why* the patterns emerge as they do.

Bokulich uses cases such as these to argue for what she calls a division of cognitive labor among models:

> [I]f one’s goal is explanation, then reduced complexity models will be more likely to yield explanatory insight than simulation models; whereas if one’s goal is quantitative predictions for concrete systems, then simulation models are more likely to be successful. I shall refer to this as the *division of cognitive labor among models* (Bokulich 2013 [60], 121).

As Bokulich notes, however, one consequence of this division of cognitive labor is that a model that was designed to optimize explanatory insight might fail to make quantitatively accurate predictions (a different cognitive goal). She continues,

> This failure in predictive accuracy need not mean that the basic mechanism hypothesized in the explanatory model is incorrect. Nonetheless, explanatory models need to be tested to determine whether the explanatory mechanism represented in the model is in fact the real mechanism operating in nature. (Bokulich 2013 [60], p. 121)

She argues for the importance of robustness analyses in assessing these explanatory models, noting that while robustness analyses cannot themselves function as a non-empirical mode of confirmation, they can be used to identify those *qualitative* predictions or trends in the model that can appropriately be compared with observations.

4.7 Conclusion

There is a growing realization that the use of idealized models to explain phenomena is pervasive across the sciences. The appreciation of this fact has led philosophers of science to begin to introduce model-based accounts of explanation in order to bring the philosophical literature on scientific explanation into closer agreement with actual scientific practice.

A key question here has been whether the idealizations and falsehoods inherent in modeling are "harmless" in the sense of doing no real explanatory work, or whether they have an essential--maybe even ineliminable--role to play in some scientific explanations. Are such fictions compatible with the explanatory aims of science, and if so, under what circumstances? While some inroads have been made on this question, it remains an ongoing area of research. As we saw, yet another controversial issue concerns the fact that many highly abstract and mathematical models seem to exemplify a non-causal form of explanation, contrary to the current orthodoxy in scientific explanation. Determining what is or is not to count as a causal explanation turns out to be a subtle issue.

Finally, just because a model or computer simulation can reproduce a pattern or behavior that is strikingly like the phenomenon to be explained, does not mean that it thereby explains that phenomenon. An important distinction here is that between a how-possibly model explanation and a how-actually model explanation. Despite the wide agreement that such a distinction is important, there has been less agreement concerning how precisely these lines should be drawn.

Although significant progress has been made in recent years in understanding the role of models in scientific explanation, there remains much work to be done in further clarifying many of these issues. However, as the articles reviewed here reveal, exploring just how and when models can explain is a rich and fruitful area of philosophical investigation and one essential for understanding the nature of scientific practice.

References

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

[2] Salmon, W., 1984, *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

[3] Frigg, R. and S. Hartmann (2012), "Models in Science" *The Stanford Encyclopedia of Philosophy*, E. Zalta (ed.), URL=<http://plato.stanford.edu/archives/fall2012/entries/models-science/>

[4] Maynard Smith, John (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

[5] Potochnik, A. (2017) *Idealization and the Aims of Science.* University of Chicago Press.

[6] McMullin, E. (1978), “Structural Explanation”, *American Philosophical Quarterly* 15(2): 139-147.

[7] McMullin, E. (1985), “Galilean Idealization”, *Studies in History and Philosophy of Science* 16(3): 247-273.

[8] McMullin, E. (1984), “A Case for Scientific Realism” in J. Leplin (ed.) *Scientific Realism*. Berkeley: University of California Press.

[9] Batterman, R. (2005a), “Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics,” *Studies in History and Philosophy of Modern Physics* 36: 25-244.

[10] Bokulich, A. (2009), “Explanatory Fictions”, in *Fictions in Science: Philosophical Essays on Modeling and Idealization*, in M. Suárez (ed.), Routledge, pp. 91-109.

[11] Cartwright, N. (1983), *How the Laws of Physics Lie*. Oxford: Clarendon Press.

[12] Duhem, P. (1914/1954), *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

[13] Elgin, M., and E. Sober (2002), “Cartwright on Explanation and Idealization”, *Erkenntnis,* 57: 441–450.

[14] Cristol, D. and P. Switzer (1999), "Avian Prey-Dropping Behavior. II. American Crows and Walnuts". *Behavioral Ecology* 10: 220-226.

[15] Batterman, R. (2009), "Idealization and Modeling" *Synthese* 169: 427-446.

[16] Kennedy, A. (2012), “A Non Representationalist View of Model Explanation”, *Studies in History and Philosophy of Science* 43(2): 326-332.

[17] Craver, C. (2006), "When Mechanistic Models Explain", *Synthese* 153: 355-376.

[18] Bokulich, A. (2011), “How Scientific Models Can Explain”, *Synthese* 180: 33-45.

[19] Kaplan, D. M. (2011), "Explanation and Description in Computational Neuroscience". *Synthese* 183: 339-373.

[20] Bokulich, A. (2008a), *Reexamining the Quantum-Classical Relation: Beyond Reductionism and Pluralism*, Cambridge University Press.

[21] Bokulich, A. (2008b), “Can Classical Structures Explain Quantum Phenomena?” *British Journal for the Philosophy of Science* 59(2): 217–235.

[22] Bokulich, A. (2012), “Distinguishing Explanatory from Non-Explanatory Fictions”, *Philosophy of Science* 79 (5): 725-737.

[23] Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

[24] Morrison, M. (1999), “Models as Autonomous Agents”. In M. Morgan and M. Morrison (eds.) *Models and Mediators: Perspectives on Natural and Social Science* (pp. 38-65). Cambridge: Cambridge University Press.

[25] Rice, C. (forthcoming), "Moving Beyond Causes: Optimality Models and Scientific Explanation", *Noûs.*

[26] Batterman, R. (2002), *Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

[27] Reiss, J. (2012), “The Explanation Paradox”, *Journal of Economic Methodology* 19(1): 43-62.

[28] Mäki, U. (2013), "On a Paradox of Truth, or How Not to Obscure the Issue of Whether Explanatory Models Can Be True" *Journal of Economic Methodology* 20(3): 268-279.

[29] Strevens, M. (2008), *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

[30] Main, J., G. Weibusch, A. Holle, and K. H. Welge (1986), "New Quasi-Landau Structure of Highly Excited Atoms: the Hydrogen Atom", *Physical Review Letters*, 57: 2789-2792.

[31] Vaihinger, H. ([1911] 1952), *The Philosophy of 'As If': A System of the Theoretical, Practical, and Religious Fictions of Mankind.* Trans. C.K. Ogden. London: Lund Humphries.

[32] Kleppner, D. and J. B. Delos (2001), "Beyond quantum mechanics: insights from the work of Martin Gutzwiller", *Foundations of Physics* 31: 593-612.

[33] Batterman, R. (2005b), "Response to Belot's "Whose Devil? Which Details?", *Philosophy of Science* 72: 154-163.

[34] Belot, G. (2005), "Whose Devil? Which Details?", *Philosophy of Science* 52: 128-153.

[35] Belot, G and L. Jansson (2010), "Review of *Reexamining the Quantum-Classical Relation*" *Studies in History and Philosophy of Modern Physics* 41:81–83.

[36] Bokulich (2010), "Bohr's Correspondence Principle", *Stanford Encyclopedia of Philosophy*. (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/bohr-correspondence/>.

[37] Heisenberg, W. (1930), *The Physical Principles of the Quantum Theory*. C. Eckart and F. Hoyt (trans.). Chicago: University of Chicago Press.

[38] Maxwell, J.C. ([1855/56] 1890), "On Faraday's Lines of Force". Reprinted in W. Niven (ed.) *The Scientific Papers of James Clerk Maxwell* New York: Dover Press, pp. 155-229.

[39] Bokulich, A. (2015), "Maxwell, Helmholtz, and the Unreasonable Effectiveness of the Method of Physical Analogy" *Studies in History and Philosophy of Science* 50: 28-37.

[40] Schindler, S. (2014), "Explanatory Fictions--For Real? "*Synthese* 191: 1741-1755.

[41]Weiskopf, D. (2011), "Models and Mechanism in Psychological Explanation" *Synthese* 183: 313-338.

[42] Buckner, C. (2015) "Functional Kinds: A Skeptical Look" *Synthese* 192: 3915-3942.

[43] Hitchcock, C. and Woodward, J. (2003), "Explanatory Generalizations: Part II. Plumbing Explanatory Depth", *Noûs* 37(2): 181–99.

[44] Irvine, E. (forthcoming), "Models, Robustness, and Non-Causal Explanation: A Foray into Cognitive Science and Biology", *Synthese*

[45] Ross, L. (2015), "Dynamical Models and Explanation in Neuroscience", *Philosophy of Science* 82 (1): 32-54.

[46] Saatsi, J. and M. Pexton (2013), "Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations", *Philosophy of Science* 80(5): 613-624.

[47] Batterman, R. and C. Rice (2014), "Minimal Model Explanations", *Philosophy of Science* 81(3): 349-376.

[48] Lange, M. (2015), "On 'Minimal Model Explanations': A Reply to Batterman and Rice", *Philosophy of Science* 82(2): 292 -305.

[49] Pincock, C. (forthcoming), "Abstract Explanations in Science", *British Journal for the Philosophy of Science*

[50] Reutlinger, A. and H. Andersen (manuscript) "Are Explanations Non-Causal by Virtue of Being Abstract?"

[51] Lange, M. (2013), "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for the Philosophy of Science* 64: 485-511.

[52] Dray, W. (1957), *Law and Explanation in History.* Oxford: Oxford University Press.

[53] Brandon, R. (1990), *Adaptation and Environment.* Princeton: Princeton University Press.

[54] Forber, P. (2010), "Confirmation and Explaining How Possible", *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 41: 32-40.

[55] Bokulich, A. (2014), "How the Tiger Bush Got Its Stripes: 'How Possibly' vs. 'How Actually' Model Explanations" *The Monist* 97(3): 321-338.

[56] van Fraassen, B. (1980), *The Scientific Image*. Oxford: Oxford University Press.

[57] Giere, R. (2001), "The Nature and Function of Models", *Behavioral and Brain Sciences* 24(6): 1060.

[58] Levins, R. (1966). "The Strategy of Model Building in Population Biology", *American Scientist,* 54(4): 421–431.

[59] Matthewson, J. and M. Weisberg (2008), "The Structure of Tradeoffs in Model Building", *Synthese* 170(1): 169-190.

[60] Bokulich, A. (2013), "Explanatory Models Versus Predictive Models: Reduced Complexity Modeling in Geomorphology" in V. Karakostas and D. Dieks (eds.) *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, The European Philosophy of Science Association Proceedings 2. Cham, Switzerland: Springer.

[61] Murray, A. B. (2003). "Contrasting the Goals, Strategies, and Predictions Associated with Simplified Numerical Models and Detailed Simulations", in P. Wilcock & R. Iverson (eds.), *Prediction in Geomorphology*. Washington, DC: American Geophysical Union, pp. 151–165.