Model Fictions, Structural Necessitation, and Explanatory Liberalism

Abstract

Some philosophers have argued that scientific models can explain their targets despite representationally distorting them. In this paper, I propose an account of model explanation based on the notion of structural necessitation, namely the representation of empirical regularities as necessities within a model. Although my account is much more liberal than its competitors, the notion of structural necessitation also helps to avoid explanatory anarchism.

1 Introduction

What constitutes a good scientific explanation? According to the received view, it is indispensable that the explanans be true, or at least approximately true. Explanations whose explanantia refer to entities that don't exist or whose explanantia severely misrepresent the target system, on this view, cannot be genuinely explanatory. This view, which I want to refer to as *explanatory conservatism*, seems immediately plausible. Consider a casual example: I step outside the house and get soaking wet. The explanation of this event is that it rains heavily. If it hadn't been true that it rained, "it rains heavily" would not have been an explanation of my getting soaking wet. It is therefore no wonder that the main philosophical accounts of scientific explanation have all abided by explanatory conservatism. In the classical deductive-nomological model of explanation, an explanandum is explained only when "the sentences constituting the explanans", which must contain a law of nature, are true (Hempel 1965, 248). In the long tradition of causal explanation, it has always been presumed that only *actual* causes, rather than just potential ones, can be explanatory of the explanandum (Salmon 1984, Woodward 2003, Strevens 2008). Likewise, mechanistic accounts of biological explanation deem genuinely explanatory only those explanations that cite the mechanism which *actually* produces the phenomenon to be explained (Machamer et al. 2000, Craver 2007).[1] Despite its

[1] Colombo et al. (2015) have recently argued that any realist commitments are gratuitous to accounts of mechanistic explanations: mechanisms can instead be viewed as models that are to one or another degree supported by the evidence, are coherent with background knowledge, and facilitate instrumental success

intuitiveness, however, explanatory conservatism faces a challenge when it comes to scientific models.

Scientific models are ubiquitous: in their attempts to understand the world, physicists assume frictionless planes, frictionless pendula, massless springs, and falling objects experiencing no air resistance. Evolutionary biologists assume infinite populations to model evolution without genetic drift, the Lotka-Volterra model makes a number of idealizing assumptions about predator and prey populations, economists assume perfectly rational agents in economic transactions, and so on. All of these models distort reality: they are radically different from what the target systems are actually like.

Philosophical puzzles associated with the representational function and the ontology of models have long been debated (Hesse 1966, Cartwright 1983). Only fairly recently have philosophers started to pay closer attention to the explanatory function of some models (Bokulich 2008a, b, Strevens 2008, Bokulich 2011). However, the explanatory function of models seems to be especially hard to square with their patent falsity: given explanatory conservatism, how can models explain their target systems if they misrepresent them so badly?

In this paper, I argue that we should dispense with explanatory conservatism when it comes to models. Instead, I propose a form of *explanatory liberalism,* according to which scientific models are genuinely explanatory by virtue of the structural necessitation they engender by postulating fictions: they represent contingent regularities as necessary relations in those fictions. Whether or not the explanans is true, I will argue, is extraneous to the explanatory function of scientific models. Since not all scientific models accomplish structural necessitation, my account successfully averts a collapse into the 'anything goes' of explanatory anarchism.

The paper is structured as follows. In Section 2 I briefly review recent discussions concerning the nature of scientific models as imagined objects in fictions. This will help us not only to prepare the grounds for my account of structural necessitation in fictions, but also to draw a line between fictive and fictional models. In Section 3 I discuss Strevens' account of fictive model explanations and Bokulich's account of fictional model explanation, which, contrary first appearances, ultimately espouse mild forms of explanatory conservatism. In Section 4 I advance my account of structural necessitation show how it allows us to be explanatory liberalists. In Section 5 I conclude this paper.

---

with regards to the explanandum phenomenon. I take Colombo et al.'s account to be concerned mostly with making sense of explanatory practices of biologists, which involve mechanisms. Colombo et al. seem less interested in conditions for a good mechanistic explanation.

2 Models as fictions?

Scientific models come in many shapes and forms: concrete models, theoretical models, toy models, computer models, analogue models, scale models, data models, and others (Frigg and Hartmann 2012). Here we are interested in one specific form of models, namely models which represent the target systems as something they are not, by means of our faculty of imagination. Before we can ask whether scientific models can be put to real explanatory work, it will be useful to get a handle on some of the more recent debates regarding model-world relations. After all, for a model to be explanatory of some parts of the world, it must be relatable *somehow* to its target system.

Recent discussions have distinguished between two broad ways of characterizing the model-world relationship. Indirect views posit models as separate entities that in one way or another relate to the target system. Direct views, in contrast, have it that model talk is really just 'make-believe' talk that is directly about the target system, rather than about a separate entity (Toon 2012, Levy 2015). Those who assume that models constitute separate entities characterize them as either *abstract objects* or as *concrete* possibilia / hypotheticals. For example, Giere (1990), in his classical discussion of models of harmonic oscillators, characterizes models as abstract objects which are individuated or even 'defined', as Giere says, by the 'equations' to be found in 'standard texts', such as textbooks on classical mechanics. Once scientific models are constructed, they can successfully represent real and concrete target systems by virtue of being similar *in some respect* and to *some degree* to them (81).[2]

A number of reasons have been offered in the literature for not regarding models as abstract entities. Some have argued that, if models really were abstract objects, they could not be compared to their concrete targets (as presupposed by e.g. Giere), since abstract objects are normally understood to have no spatiotemporal extension, whereas concrete objects do (Thomson-Jones 2010). Others have argued that if models were abstract objects, then model descriptions are literally false of them, even though we might want to say that there is "some truth" to model descriptions with regards to models (Contessa 2010, see also below). Others still have rejected the view of models as abstract objects for maybe less compelling reasons. They find this view hard to square with (i) the intuition that models are products of our imagination, and with (ii) the fact that models particularly in biology

[2] Some of the core elements of Giere's account have recently been defended by Weisberg (2012). Others have criticized Giere's view of models as too closely tied to *theoretical* models (Morgan and Morrison 1999). It is also worth noticing that Giere is sympathetic to the 'semantic view', which views theories as families or sets of models. There are different formulations of the semantic approach. For an excellent overview see Winther (2016).

do not seem to lend themselves easily to a mathematical description (apparently presupposing that abstract objects would require this) (Levy 2015).

Instead, it has been proposed that scientific models are better viewed as concrete possibilia or concrete hypotheticals, i.e., imagined objects that *would* be concrete, if they were real (Godfrey-Smith 2009, Contessa 2010). Such a view of models lends itself particularly well to regarding scientific models as *objects of fiction*, analogous to fictional characters in literary fictions, such as Sherlock Holmes, Anna Karenina, or Tom Sawyer.[3] Just like in literary fictions, we may say that there is truth and falsehood *in the fiction* and truth or falsehood of descriptions in the fiction as compared to the actual world. One can accordingly distinguish between internal and external sentences, respectively (Contessa 2010). Thus, *internally*, it is true in the Bohr model of the atom that the atom that the electron orbits the nucleus just like it is a truth in the fiction that Sherlock Holmes lives on 221b Baker Street. Likewise, it would be false to say that in the Bohr model electrons crash into the nucleus just as it would be false to say that Sherlock Holmes is a crooked gangster. Conversely, *externally*, we know that it is simply false that the electron orbits the nucleus (instead: there are only certain probabilities that the electrons is located in a certain region in the atom), just as we know that Sherlock Holmes never existed. There are of course also fictions that are true both internally and externally. For example, it is true both in Tolstoi's *War and Peace* and in reality that the French invaded Russia. Likewise, it is true both in models of the aether and in reality that light has wave-like properties (even though aether models presumed longitudinal instead of transverse light waves).

Several authors have sought to explore the analogy of models to literary fictions with the help of Kendall Walton's antirealist account of fictional entities, as expounded in his *Mimesis as Make-Believe* (Walton 1990). Although the resulting accounts differ in a number of ways, particularly about whether model fictions should be conceived of as entities separate from the target system, they all embrace one basic Waltonian idea: scientific practices involving models can be construed as games of make-believe, just like one can construe literary fictions as resulting from games of make-believe (Frigg 2010, Toon 2010, 2012, Levy 2015).

Apart from this narrower sense in which models can be conceived of as analogous to works of fiction in literature, there is also a wider (and minimal) sense in which models are fictitious: they are literally false descriptions or, more appropriately, *imaginings* of their

[3] It should be said, though, that some philosophers have defended the view that literary fictions refer to abstract objects. See Kroon and Voltolini (2011) for more. There are also philosophers who have defended dual (Meinongian) accounts of model fictions in which models are both abstract and concrete (Contessa 2010).

target systems. But not all models are false to the same degree. Whereas some models postulate entities that idealise and simplify objects that we know exist, others postulate entities that we know do not exist.[4] Accordingly, it is possible to 'de-idealise' the former to bring them closer to the actual target system (McMullin 1985), but it is not possible with the latter. Although the distinction is probably not a sharp one, there clearly is a difference between working with models of the pendulum that ignore friction on the one hand, and models that postulate phlogiston when seeking to accommodate processes of combustion on the other hand. With Suárez (2009) one may refer to models that incorrectly describe real entities as *fictive* and reserve the label *fictional* for models that describe non-existent entities.

In this paper, I want to argue, in contrast to extant accounts, that both fictive and fictional models can be genuinely explanatory. On the account that I will present in Section 4, fictive and fictional models explain by representing empirical regularities as necessities. The entities invoked for this purpose are imagined substructures of the target system and we can speak of internal truths within models and we can ask whether the postulates by the models are (externally) true or not. In models of a fictive character, some posits of the model will be true, others not, whereas in models of a fictional character, there will not be any posits that are externally true. Although this will not be central in what follows, I believe with the majority of philosophers working on models as fictions that models are best conceived of as concrete possibilia, not abstract objects. At any rate, on my account model fictions are separate from the target system, but, as we shall see, relate to it in a surprisingly unproblematic way.

3 Mild explanatory conservatism

Although it is widely agreed that scientific models play a crucial instrumental function in scientific practice, and although there has been much discussion about how to best construe the representational function of imaginary models, another epistemological function of models, namely explanation, has not received much specific attention. Exceptions are the works of Michael Strevens and Alisa Bokulich, who have argued that fictive and fictional models can be genuinely explanatory, respectively. It is to their works which we will turn in a moment. We shall see that, despite first appearances, both Strevens and Bokulich ultimately espouse mild forms of explanatory conservatism.

Before proceeding, we should note that there exists a sizable literature about understanding, that, for better or for worse, is fairly distinct from the literature about

[4] This distinction has been drawn by a number of authors, including Levy (2012) and Bokulich (2009).

explanation. Particularly interesting for our purposes, many theorists about understanding, in contrast to their colleagues working on explanation, have been quite liberal about requirements of truth (Elgin 2004, 2007, de Regt 2015, Doyle et al. forthcoming, Reutlinger et al. forthcoming). At the same time, theorists about understanding have often rather little to say about explanation or are explicitly committed to explanatory conservatism. For example, Trout (2002), who perhaps is the first to highlight understanding as an independent topic from explanation, states very clearly his commitment to explanatory conservatism: "when it comes to [the quality of] explanation, there is no substitute for simply being (approximately) right" (230). Likewise, Reutlinger et al. distinguish (as others have before) between how-actually and how-possibly understanding and grant toy models (i.e., their topical focus) only how-possibly understanding, on the assumption that how-actually understanding requires the provision of true explanations. We therefore will bracket off the understanding literature in what follows.

3.1 Strevens' optimal models explanations

In his book-length treatment of causal explanation, Strevens (2008) dedicates one chapter to model explanations, and more specifically idealized explanations. Contrary to 'pragmatic' accounts, which view idealized explanations merely as 'stepping stones' to true explanations and as explanatorily suboptimal, Strevens believes that good model explanations are explanatorily optimal. That is, true explanations cannot improve on them. In fact, Strevens even claims that model explanations are "always better than their veridical counterparts" (2). Prima facie, this seems to put him squarely into the camp of those who believe that models can provide explanation, and in fact *better* explanations than true explanations.

The main example Strevens adduces to make his case is the explanation offered by the kinetic theory of gases for Boyle's law, according to which a gas's pressure and volume are inversely proportional ($P \propto \frac{1}{V}$ where P=pressure, V=volume)). Although the kinetic theory (KT) correctly assumes that gases consist of molecules in motion, it makes a number of idealizing assumptions about gases: it *falsely* ignores the molecule's extension, *falsely* ignores intermolecular forces, *falsely* ignores any collisions between molecules, and *falsely* invokes classical mechanics to describe processes which are quantum mechanical. In the terminology introduced at the end of the previous section, KT is thus a *fictive* model—the label 'theory' KT thus carries mostly for historical reasons. How can we make sense of the fact that KT explains Boyle's law despite the fact that it is so highly idealized?

According to Strevens' *kairetic* account (from greek: *kairos*=decisive), we can determine the explanatory relevance of causal factors in the explanation of a phenomenon by asking what factors would make a (decisive) *difference* to whether or not the phenomenon obtains. Causal difference makers for a phenomenon P, in Strevens account, are determined by a causal model containing as premises statements asserting the presence of causes for P, so that those premises deductively entail P. If we can remove a premise A of that causal model of P without P ceasing to obtain, then A is not a difference-maker for P, and accordingly, is not required in the explanation of P. For example, although the Moon does exert a gravitational pull on objects falling to the ground on Earth, it makes no difference to whether or not the apple falls to the ground, or even to the speed of its fall.

In a *textbook explanation* of Boyle's law, three fundamental assumptions are made: the gas pressure is proportional to the frequency of molecule collisions with the wall, the frequency of the molecule-wall collisions is proportional to the gas density, and the gas density is inversely proportional to its volume. From these assumptions it follows that pressure is inversely proportional to volume. The assumptions are justified only under the condition of the idealisations we mentioned already above, the details of which we will skip here (see Strevens 2008).

Strevens discusses the explanatory relevance of all causal factors in a *veridical* model of the 'Boylean behavior' of gases one-by-one and concludes about the idealized textbook explanation that "though non-veridical, it contains *all* the elements of the *best* explanation of the law" (Strevens ms, 23, added emphasis).[5] For example, although intermolecular forces do make a difference to the van-der-Waals equation (which is more accurate at low temperatures and high pressures), they make no difference to the Boylean behavior of gases. That is, with regards to the causes it deems explanatorily relevant, the idealized explanation surprisingly turns out to be *identical* to what Strevens calls the *canonical model*, namely the model that contains *only* difference-makers. This is then also why "the idealized model does not suffer for its falsehoods" (ibid.). Still, there is one important difference between idealized and canonical models: in contrast to canonical models, idealized models state causal factors that are explanatorily *irrelevant*—just so as to then disregard them by 'idealizing' them, i.e., by "assigning them infinite, zero, or some other value" (27). So why do scientists not simply state canonical models and instead use their idealized counterparts? According to Strevens, this has to do with communication: In

[5] Here I refer to an unpublished manuscript by Strevens entitled "Why Explanations lie: Idealisation in Explanation" which, in an altered version, forms chapter 8 of his book *Depth* (Strevens 2008). I use Strevens' paper rather than his book because his focus is more firmly on false assumptions in scientific explanations.

contrast to canonical models, the explanatory irrelevance of certain causal factors is "underlined in a dramatic way—more dramatic, certainly, than a canonical explanation's knowing silence" (30). Thus, according to Strevens, the function of idealisations is therefore not, as it might first appear, to state that "some non-actual factor is relevant to the explanandum", but on the contrary, that "some actual factor is irrelevant" (25). This is also why the idealized model is "vastly superior" to the veridical model, since the latter, in contrast to the former, implies that certain causal factors are difference-makers, when they are in fact not (31). Strevens concludes:

> Do idealizing explanations lie? If taken literally, yes. If understood correctly, no: what appear to be lies are true claims about the identity of the nondifference-makers. Explanations lie, then, in order to communicate a certain kind of truth, a truth about what does not matter to the causal production of the explanandum. (35)

In sum, idealizing explanations are good explanations (and in fact *better* than their veridical counterparts), *not* because of their distortions, but because they correctly identify all those causes that make a difference to the occurrence of the explanandum phenomenon. So what carries the actual load of explanatory work, for Strevens, are truths, not falsehoods, about the target system. For this reason, Strevens' account is to be classified as a form of mild explanatory conservatism. Where Strevens differs from other proponents of explanatory conservatism is that he regards it as detrimental to the explanation of the target system when a model takes into consideration *all* the prima facie relevant causal factors (such as intermolecular forces).

It is worth pointing out that, although model explanation has not been a focus of the fictionalists, Levy (2015), in his Waltonian account of model fictions, has recently embraced the kind of explanatory conservatism defended by Strevens. Commenting on Strevens' example, Levy writes

> we evaluate the truth of parts of what the ideal gas model says, when interpreted as about real gases. When we do so, we find that it is partly true, i.e. with respect to the role of the motion of particles; and partly untrue, i.e. with respect to the role of collisions among particles. We know that *the untrue part does not affect the explanandum* (Boyle's law), so *we regard the partial truth contained in the ideal gas model as an explanation of Boyle's law* (Levy 2015, 793, added emphasis).

Thus, for Levy, a model fiction successfully represents and explains a target system in virtue of (parts of) its descriptions of the target system being (externally) true. The untrue parts of the model do not contribute to the explanatory power of the model. Levy helps himself to the notion of partial truth developed by Yablo (2014), the details of which need not concern us here. However, one may raise a number of questions already about the very idea, for example: how many falsehoods can a model contain before it ceases to be explanatory? If it is only the true statements in a model (and not the false ones) that can do any explanatory work, why do we use models in the first place?

These issues may not be insurmountable, and in fact, Strevens has answers to some of them. In particular, as already mentioned, Strevens believes that the model's falsehood are simply there to highlight those factors that do not make a causal difference to the explanandum phenomenon. At any rate, it should be clear at this point that Strevens' account of fictive explanations is inapplicable to fictional explanations, where the entities figuring in the explanandum are known not to exist. This is where we shift our attention to Bokulich's account of structural model explanations.

3.2 Bokulich's structural model explanations

In contrast to Strevens, Bokulich is concerned with fictional, rather than just fictive, model explanations. Bokulich has articulated three conditions that summarize her view on how fictional models explain:[6]

1. *Reference to a model*. The explanans must make 'essential' reference to a scientific model, whereby that model "involves a certain degree of idealization and/or fictionalization" (Bokulich 2011, 39).
2. *Counterfactual dependence.* The model explains the explanandum "by showing how the elements of the model correctly capture the pattern of counterfactual dependence of the target system". Bokulich also speaks of an isomorphic relation between the counterfactual structure of the model and the counterfactual structure of the phenomenon. By virtue of this isomorphism, "the model should also be able to give information about how the target system would behave, if the elements represented in the model were changed in various ways" (ibid.).
3. *The justificatory step.* The 'justificatory step' specifies the "domain of applicability of the model" and shows "where and to what extent the model can be trusted as an adequate representation of the target for the purpose(s) in question" (Bokulich 2012, 730).[7]

Bokulich has illustrated her account primarily with examples from 'semi-classical' physics where classical periodic orbits are deployed to explain phenomena in the quantum realm (such as the morphology of 'scarred wavefunctions' and resonances in the absorption spectrum of Rydberg atoms). The simplest of her examples is the explanation of the

---

[6] The status of these conditions is not quite clear. Sometimes Bokulich speaks of them as if they were at least necessary (and perhaps jointly sufficient) conditions for structural model explanations. At other times, she presents them merely as *descriptions* of structural models.

[7] Bokulich's formulation of the third condition has changed. In her (2011 39), Bokulich characterizes this condition as "specifying what the domain of applicability of the model is and showing that the phenomenon in the real world to be explained falls within that domain".

Balmer series of the spectral lines of hydrogen by the Bohr model of the atom, which postulated classical circular electron orbits, with discrete quantum transitions between the orbits. The Bohr model was superseded by quantum mechanics in the mid-1920s; classical electron orbits are now believed not to exist. Nevertheless, the false Bohr model, according to Bokulich, explains the Balmer series, because it allows us answer a number of 'what-if-things-had-been-different-questions', such as "how the spectrum would change if the orbits were elliptical rather than circular" (Bokulich 2008a). That is, we are able to exploit the counterfactual dependencies that holds between the model and the explanandum phenomenon in order to gain understanding of the target system. Bokulich is adamant that semiclassical models in the quantum realm with their postulation of the 'fiction' of electron orbits are not merely convenient 'calculational tools' or 'phenomenological models' that just "save the phenomena". Instead, she insists, these models give us "genuine physical insight" and indeed "deeper" explanations than quantum mechanics (ibid., 230, 233). Here, Bokulich's claims clearly mirror those of Strevens about fictive model explanations.

Bokulich's account faces a number of challenges (cf. Belot and Jansson 2010, Schindler 2014, King 2016). One of these challenges concerns the *problem of explanatory asymmetry*: what are the conditions in the account of explanation that ensure that the explanans explains the explanandum and not vice versa? E.g., a good account of explanation should not sanction the length of the shadow of a flagpole as legitimate explanans for the height of the flagpole. But since the counterfactual dependence relation between the model and the explanandum is an isomorphic one according to Bokulich, it would seem that this relation lacks the directionality that a solution to the problem of explanatory asymmetry requires.

One might be tempted to think that Bokulich has an easy way out of the problem of explanatory asymmetry, at least when it comes to model explanations of the phenomena, since the first condition of her account states that the explanans must make reference to a scientific model. However, if this is a solution to the problem, then the problem of asymmetry cannot really be a problem for anyone, as this solution would amount to no more than stipulating that the intended explanans has got to be the actual explanans. The solution seems thus too cheap. Ideally, an account of explanation should equip us with all the resources needed to recover the explanatory asymmetry without such ad hoc stipulations.[8] Bokulich has attempted a different solution, which we shall discuss in a moment.

Another challenge for Bokulich, which I want to mention here, is the demarcation of genuine and non-genuine explanations involving fictions. Let us refer to this problem as the *problem of demarcation*. An example for a non-genuine explanation involving fictions, which was brought up in the discussion by Belot and Jansson (2010), is Ptolemaic astronomy. In Ptolemaic astronomy, the phenomenon of planetary retrogressions, i.e., the apparent periodic backwards motion of the planets as seen in the night sky, is accounted for by the notorious device of the epicycle (see Fig. 1). Such examples potentially pose a problem for Bokulich's account, because it appears to license such examples as genuinely explanatory even though we intuitively might want to deny epicycles any explanatory power. Bokulich accepts this challenge and specifies that, although the Ptolemaic system can be said to "make very precise predictions about the locations of the planets in the night sky", one should refrain from granting that it provides an actual explanation of retrogressions. Instead, retrogressions receive their genuine explanation in the Copernican system. The Ptolemaic system is merely a "phenomenological model" (Bokulich 2012, 730 and 735).

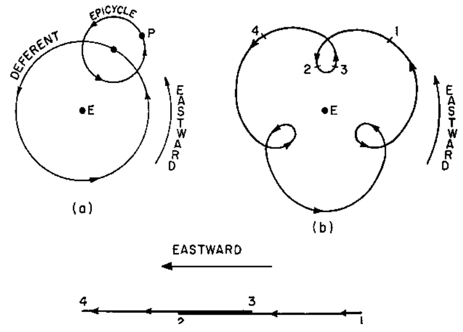

Fig. 1: retrogressive motion. From Kuhn (1957)

Bokulich response to the problem of asymmetry and demarcation is as follows. Bokulich does not believe that there is anything intrinsically different about non-explanatory and explanatory fictions. In order to nevertheless be able to distinguish between the two, she appeals to "a contextual relevance relation set by the current state of scientific knowledge" that determines the items of an 'explanatory store' that scientists can use to explain the phenomena around them (ibid., 734). She specifies that "what is to count as an adequate fictional representation is something that has to be negotiated by the relevant scientific community and will depend on the details of the particular science, the nature of the target system, and the purposes for which the scientists are deploying the model" (ibid.). This notion of a relevance relation, she believes, is enough to rule the Ptolemaic model of astronomy non-explanatory: "given the relevance relation set by contemporary science, epicycles are irrelevant to the explanation of retrograde motion"

[8] An account that manages to do so is e.g., Woodward's, where explanatory asymmetry reflects causal asymmetry, which can be identified by means of interventions.

(735). She also believes that appealing to the aforementioned relevance relation solves the problem of asymmetry, as in the famous barometer-storm example, since "behaviors of barometers *is not relevant* for explaining the behavior of storms" (734; added emphasis). But particularly with this example, it is not quite clear how Bokulich's contextual explanatory relevance relation is supposed to solve the problem of explanatory asymmetry, unless implicit appeal is made to a causal relationship. This, Bokulich must (and wants) to avoid in her account of explanatory fictions. Also, it is worth pointing out that van Fraassen (1980), from whom Bokulich seems to borrow the idea of explanatory relevance, has argued that in the famous flagpole-shadow example, the explanatory relation can in fact run in *both* directions. But clearly, the spectrum of hydrogen does not explain the jumping of electrons from one to another orbit in the Bohr model. If anything, it's the other way around.

Bokulich emphasizes that the explanatory store may also contain explanatory fictions. What fictions are allowed to enter? Only ones that give us genuine physical insight into the problem at hand, Bokulich stipulates. If our planetary system had been a geocentric one, she explains, then the Ptolemaic system would have been explanatory. But since we live in a sun-centred system, it is not. In contrast, "the classical periodic orbits of the electrons in the quantum dot *are* an explanatory fiction", because, "given the relevance relation set by contemporary physics, the classical periodic orbits are able to capture, in their fictional representation, *real features of the quantum dynamics* in the dot" (ibid., added emphasis). Periodic orbits are thus explanatory because "they are a fictional representation that is able to *generate genuine knowledge of the true underlying quantum dynamics*" (ibid.; added emphasis). Bokulich concludes that "realism comes in to distinguish explanatory from non-explanatory fictions" (735), and adds that in explanatory fictions such as semiclassical mechanics, there is a "well-defined translation key" which allows us to translate statements about the fictions "to statements about the *underlying structures or causes* of the explanandum phenomenon" (ibid.; added italics).[9] In other words, in order to address the problem of demarcation, Bokulich espouses a mild form of explanatory conservatism.[10]

---

[9] The idea of a 'translation key' between fiction and real structures is already present in Bokulich's earlier work (although in much less explicit form). For example, in illustration of the justificatory step she writes about Gutzwiller's periodic orbit theory that it "specifies precisely how the classical trajectories can properly be used to model certain features of the quantum dynamics" (Bokulich 2008a).

[10] In her most recent publication on the topic, Bokulich speaks of *credential fictions,* which are individuated "in relation to a veridical representation" (Bokulich 2016).

There is a tension between Bokulich's proposal for how the explanation provided by model fictions are justified and her claim that these explanations are "deeper" than the explanations afforded by the theories justifying them (Schindler 2014). If it really is the case, as Bokulich claims, that fictional explanations of quantum phenomena that make reference to the fiction of periodic electron orbits give us "deeper" insight into those phenomena than quantum mechanics (Bokulich 2008b, 230), and that quantum mechanics itself "fails to provide adequate understanding" of such phenomena (230), is explanatorily "deficient" (230) and "incomplete" as compared to the model fiction (233), then it is hard to see how quantum mechanics could justify the explanations provided by the model fiction. In particular, if there exists a counterfactual dependence between the model fiction and the explanandum phenomenon, and statements about the model can be translated into statements about the underlying structures or causes of the phenomenon, as Bokulich claims, then any counterfactual dependence that holds between the model and the phenomena will also hold between the theory justifying the model and the phenomena. So either the model explanations are deeper than the justifying theory at the cost of some statements in the model remaining unjustified, or all model statements can be translated into statements of the theory at the cost of losing the explanatory autonomy which the above quotes by Bokulich seem to suggest. On her account, Bokulich cannot have both deeper model explanations *and* a full justification of those models. One could of course give up on either a full justification or the claim that models are explanatorily deeper (and in that sense explanatorily 'autonomous'). But giving up the former would rob Bokulich of her solution to the problem of demarcation. Giving up on the latter would require further arguments why explanatory fictions are not just convenient, simplifying devices, rather than genuinely explanatory (as Bokulich has claimed).

4 Structural necessitation and explanatory liberalism

It is time now to present my own account of model fiction explanations. I will argue that a crucial, and yet hitherto overlooked, component of the explanation provided by the kinetic theory of gases is the representation of contingent regularities as physical necessities in a fictional model. I call this feature of model explanations *structural necessitation*. I will demonstrate that this feature is not only to be found in the fictive KT, but also in the fictional caloric theory of heat, which postulated heat particles which are known not to exist. On this basis, I shall argue that we should give up explanatory conservatism for explanatory liberalism. At the same time, I will show that structural necessitation successfully avoids the treaded explanatory anarchism of 'anything goes'.

4.1 Structural necessitation in fictive model explanations

As we've mentioned above, the kinetic theory of gases (KT) idealises gas molecules as point particles with zero extension but with mass and perfect elasticity (amongst other things) and is thus – despite its historical label 'theory' – adequately described as a fictive model. Despite these numerous idealizations and literally false assumptions, KT allows us to derive not only Boyle's law ($P \propto \frac{1}{V}$) but also the Gay-Lussac law ($P \propto T$), both of which are summarized in the ideal gas law $PV = nRT$, where P=pressure, V=the volume of the gas container, T=temperature, R=the ideal gas constant, and n=the amount of substance of gas in moles.

Consider once more Boyle's law. When holding fixed the temperature of a gas, Boyle's law says that if we decrease the gas volume in a container, the gas pressure will rise, and if we increase the volume, the gas pressure will fall. In KT, this is explained by a decrease in the container volume corresponding to an increase in intermolecular space inside the container resulting in fewer molecule-container wall collisions (as molecules have to travel longer before hitting the container wall). Molecule collisions, in turn, are taken to correspond to gas pressure. The Gay-Lussac law KT explains thus: if the temperature of the gas container is increased (whilst holding the volume fixed), this will be taken to correspond to an increase of the speed of the gas molecules. Since an increase of the speed of gas molecules will increase the frequency with which the molecules hit the container wall, the pressure of the gas will increase.

It is quite apparent that the explanations provided by KT lend themselves to a counterfactual analysis. For example, had we increased the number of collisions of gas molecules (by decreasing the container volume), the pressure would have risen. Also, the empirical laws of Boyle and Gay-Lussac, likewise, can be rephrased in counterfactual terms—with no reference to gas molecules. E.g., had we increased the gas pressure, the temperature would have risen, and vice versa, had we increased the temperature, the pressure would have increased (under the condition that the volume remains fixed). It would thus seem that Bokulich's identification of counterfactuals and isomorphic relations between model and regularities is important.[11] But it is not the whole story. A fictive model such as KT also allows us to *represent empirical (and contingent) regularities as necessities*.

---

[11] Interestingly, Woodward (2003), from whom Bokulich heavily borrows, seems to think that explanations provided by KT of the ideal gas law is inferior to explanations derived from interventions of the variables in the self-same law (232).

Both Boyle's and Gay-Lussac's law are contingent: the world could have been such that a gas's pressure *falls*—rather than rises—when its volume is reduced, and such that a gas's pressure *falls*—rather than rises—when its temperature increases, respectively. KT represents these regularities as necessities: with regards to Boyle's law, a reduction in a gas's volume *has got to* result in an increase (rather than a decrease) of molecule collisions per time (and therefore pressure), as the distance traveled by the molecules before colliding with the container walls is shortened. Since a reduction in volume, in KT, *cannot* result in a decrease in collisions—as that would require an increase of space—the contingent law by Boyle is represented as a necessity within KT. This structural necessitation engendered by KT, I surmise, is an important reason for KT's capability of explaining Boyle's law. Similar reasoning applies to KT's explanation of Gay-Lussac law.

On counterfactual accounts of explanation, explanatory relations supervene on counterfactual dependencies. As we noted above, Bokulich's counterfactual account of explanatory fictions, faces the problem of explanatory asymmetry: if there is a counterfactual dependence between the model and the explanandum phenomenon, then on what basis can we say that the model explains the explanandum phenomenon *but not vice versa*?[12] The notion of structural necessitation helps us to see why this is the case: only the model represents contingent regularities as necessities, and not vice versa.

On the notion of structural necessitation, what gives us understanding in model explanations of the target system is that it allows us to see why empirical laws, on the supposition of the model's assumptions, *have to take* the form that they take. Accordingly, models thus allow us not only to answer what-if-things-had-different-questions (as in counterfactual accounts), but also questions about why certain changes in the explanandum *had to happen*. Models such as KT thus give us 'how-necessarily' understanding, as compared to mere 'how-actually' understanding (championed by proponents of causal and mechanistic explanations), and 'how-possibly' understanding (famously defended by Woodward (2003) and recently in the realm of toy models by Reutlinger et al. (forthcoming)). Moreover, it should be stressed that in contrast to extant counterfactual accounts of explanation, which have sought to exploit the counterfactual dependencies between events or variables, structural necessitation is a representational relation that explains the particular contingent *form* of empirical *relations*.

Contrary to what has been claimed by Lange (2013) in his discussion of mathematical necessities, I would not want to claim the necessities, models such as KT

[12] Other counterfactual accounts of explanation have invoked the notion of interventions to avoid such problems (Woodward 2003). Bokulich explicitly rejects this notion. See Schindler (2014) for more details.

represent regularities as, to be modally stronger than the ‘causal necessities’, or that models would explain “by showing the explanandum to be more necessary” than causal explanations (485). Instead, I consider the necessities engendered by KT as modally compatible with causation. So what kind of necessity are we talking about here then?

Clearly the necessities within KT, minimally, are intended to be *physical* necessities that arise in virtue of the postulated substructures. But again, within KT those substructures do not exist in the form postulated by KT: KT entirely ignores intermolecular forces and employs classical mechanics in a realm in which quantum mechanics is supposed to reign, amongst other things. Thus, we may say that KT represents empirical regularities as *necessities in a fiction.* Although there will have to be said much more about the nature of these necessities in the future, but they should in principle be no more puzzling than actualities in fictions. In many fictional stories, the laws of physics are the same as in the actual world. So when Sherlock Holmes drops his pen – everything else being equal – it will *have to* fall to the floor by virtue of the physical necessity of the law of gravity. Sherlock Holmes’ pen will never actually fall to the floor, as Sherlock Holmes does not exist. Likewise, molecules as envisioned by KT (as non-colliding and extensionless) do not exist. Nevertheless, in KT, the classical laws of physics apply to it.

Thinking about model explanations in this way allows one to overcome some of the issues that have been debated in the recent models as fiction literature. For example, as mentioned in Section 2, it is sometimes seen as problematic to view the fictive or fictional entities described by models as distinct from the target system (Toon 2010, Levy 2015). On my account, there is no need to compare the entities in the model with the target entities in order to say how well the model explains the target system. All we need to do is check whether the regularities in question are represented as necessities in the model. That is, in the first instance, we need to determine that there exists a structural isomorphism between the relations in the model and the relations *in the world*, before we can then determine whether the model represents these relations as necessities. The connection between the model and the real world is thus established unproblematically, via the first of these two steps.

Broadly speaking, model explanation by structural necessitation is a form of the *modal conception* of explanation (also embraced by Lange (2013)). An issue that has been raised against the modal conception is that it fails to accommodate statistical and indeterministic explanations, in particular, when it comes to such explanations of *events* (Salmon 1984). For example, quantum mechanics does not predict (or explain) any particular measurement of a quantum system; it only predicts certain probabilities of measurement outcomes. Likewise, Mendelian genetics does not predict (or explain) any particular color of peas, but only certain probability distributions. But, as already mentioned, my account of model explanation is meant to be an account of *regularities*, not of single events. So, in the Mendelian example, so far as it can be conceived of as a model (which I think it can), the model would explain the probability distributions of dominant and recessive traits in white and purple flower pea plants, but not why any particular pea flower has the color does it does. Also, my account of model explanations is supposed to explain regularities (probabilistic or non-probabilistic ones), not deviations from these regularities. Thus, it may of course well be that the actual character distributions in a *particular* sample of pea hybridization deviate from what the Mendelian model would have us expect. But my account of model explanation is not supposed to account for such deviations and should accordingly not be held accountable for it.

4.2 Structural necessitation and the caloric theory of heat

Interestingly, KT is not the only model that manages to represent IGL as necessities. The caloric theory of heat (CT) is another such theory. In contrast to KT, where, as we just saw, heat is conceived of as molecular motion, in CT heat is conceived of as a *substance*, namely so-called caloric. CT furthermore supposes that caloric particles attract matter particles and that caloric particles repel each other (Fig. 2). CT for long was a strong competitor to KT. It was held by the likes of Joseph Black, who used CT to explain changes of states, Antoine Lavoisier, who even included caloric as one of the known chemical elements of the time in his field-defining *Traite´ elementaire de chimie*, and Sadi Carnot, who lay the groundwork for the later development in thermodynamics on the basis of CT (Chang 2003). It was only in the late 19th century that CT was entirely abandoned. Interestingly, CT was also capable of explaining Boyle's law and Gay Lussac's law. With regards to the former, a reduction of gas container space would increase the pressure on the container wall, as the mutual repulsion of caloric particles would be stronger the closer the particles were to each other. With regards to Gay-Lussac's law, an increase in the amount of caloric, corresponding to a temperature rise, would result in an increase in pressure would result, because the supply of further caloric particles would increase the net amount of repulsion between them (all

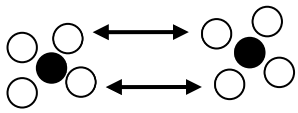

Fig. 2 Caloric particles (white) repel each other but are attracted by matter particles

whilst the volume remains fixed, of course).[13] Tab. 1 summarizes the relations we just identified between our two laws and KT and CT, respectively.

Regularity	**Theory**	**Explanation in terms of structural necessitation**
$P \propto \frac{1}{V}$ (Boyle's law, at constant temperature)	KT	FC: frequency of collisions with container walls ↑(↓) ∝ gas density ↓(↑) *Not possible*: FC ↓(↑) ∝ gas density ↓(↑)
	CT	SR: strength of caloric repulsion ↑(↓) ∝ gas density↓(↑) *Not possible*: SR ↓(↑) ∝ gas density ↓(↑)
$P \propto T$ (Gay-Lussac's law, at constant volume)	KT	frequency of collisions with container walls ↑(↓) ∝ speed of molecules ↑(↓) *Not possible*: FC ↑(↓) ∝ speed of molecules ↓(↑)
	CT	strength of caloric repulsion ↑(↓) ∝ amount of caloric ↑(↓) *Not possible*: SR ↑(↓) ∝ amount of caloric ↓(↑)

Tab. 1: KT, CT, and Boyle's and Gay-Lussac's laws. In both KT and CT contingent regularities are represented as necessities. Downward arrows indicate decreases in the values of the variable in question and upward arrows increases.

From this comparison it should be pretty clear that, insofar one accepts that structural necessitation might be an important part of the explanation of IGL (which one should), one cannot deny that CT, in virtue of accomplishing the relevant structural necessitation, is explanatory of IGL too. But then it is not only the case that fictive models, such as KT, can explain but also that fictional models, such as CT, can. Obviously, in CT, there are no statements that are externally true, to use the terminology introduced in Section 2. There are only internal truths, such as that caloric particles repel each other, which is true in CT, but externally very wrong, since there is no caloric. This of course threatens the 'anything goes' of explanatory anarchism. How can it be avoided?

One might be tempted to think that CT explains by virtue of latching onto some correct structure, a la structural realism.[14] So, CT, in our example, would be explanatory simply because it correctly identified the structure identified in the partially true KT. Explanatory conservatism would be saved. Unfortunately, however, this suggestion won't work. The structure at issue is simply not of the right kind.

In the realism debate, realists have sought to identify *theoretical* structures that were retained through theory-change, e.g. Fresnel's correct identification of the wave equations

[13] In order for these explanations to work, it must be supposed that caloric particles, by virtue of their attraction to ordinary matter particles, attach to the container wall. This follows from the basic postulates mentioned in the main text. For further comparisons between KT and CT see (Votsis and Schurz 2012).

[14] This is what Votsis and Schurz (2012) suggest.

in his false theory of the aether. But the continuity between CT and KT is just IGL, i.e., an *empirical* regularity. But that is not enough for realism. The theoretical representations of CT and KT are radically different, whereby KT's are partially and CT's obviously false. Both KT and CT manage to represent IGL as a necessity, but they do so in radically different ways.

Maybe one would want to insist that KT explains and CT doesn't, because KT, despite the falsity of many of its assumptions, still latches onto the real causes more or less. In particular, it gets right that gases consist of molecules in motion. How could one deny CT any explanatory power by virtue of structural necessitation without begging the question? I think there are two ways. First, one could deny that structural necessitation is part of KT's explanation of IGL. This seems obviously wrong, as we have clearly demonstrated that KT *does* achieve that feat. We should also repeat that there is no contradiction in saying that an important aspect of KT's explanation of IGL is structural necessitation and the claim that KT identifies the right causes of the behavior of gases. In fact, one may say that KT achieves structural necessitation by identifying the right causes, even though identifying the right causes is not a necessary condition for structural necessitation. Second, one could grant that structural necessitation is part of both KT and CT, but deny that structural necessitation gives us any understanding of the target system. That, however, would deny that how-possibly or even how-actually explanations give us any understanding (because those modalities are entailed by necessities), which would be quite absurd, as many current accounts of explanation take those to provide genuine understanding (Woodward 2003, Bokulich 2014, Reutlinger et al. forthcoming).

4.3 Explanatory demarcation

Given the these considerations, let us return to the question raised above: how can we avoid explanatory anarchism, once we accept that fictional models can be explanatory? In my discussion of Bokulich's account (Section 3.2) I referred to this as the problem of demarcation.

Consider once more the Ptolemaic system of astronomy, which Bokulich sought to rule out as non-explanatory. The epicycle, which this system employed, clearly is a fictional theoretical device and, in contrast to the periodic orbits in the Bohr model of the atom, presumably not explanatory. How can we deem epicycles non-explanatory, if we—contrary Bokulich—do not wish to fall back on explanatory conservatism? Consider the phenomenon of maximum elongation of the inner planets Venus and Mercury, i.e., the fact that these planets can never be observed beyond a certain angle from the ecliptic (47° and 28°, respectively). In the Ptolemaic system, this fact is accounted for arbitrarily by stipulating that the centre of the epicycle on which they inner planets supposedly move would be fixed on a line connecting the sun and Earth. In contrast, in the Copernican system, the inner planets cannot possibly move away from the sun beyond a certain angle, simply because the inner planets' orbits are encompassed by the Earth's orbit around the sun (see Fig. 3).

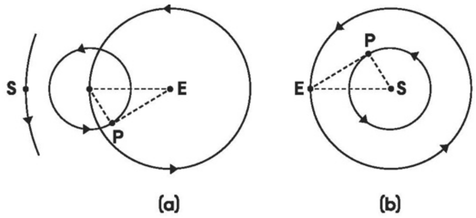

Fig. 3: maximum elongation as accounted for in the Ptolemaic and the Copernican system (a and b, respectively). From (Kuhn 1957).

That is, the Copernican system restricts the possible motions of the inner planets by design, as it were. The empirical and contingent phenomenon of maximum elongation is represented as a necessity in the Copernican system, but not in the Ptolemaic system. Although this example of course does not suffice to deny the device of the epicycle *any* explanatory import, it does go to show that the stripped-down Copernican system without any epicycles is explanatorily superior with regards to explaining maximum elongation.[15] We can thus disqualify some fictions, or more accurately, some *uses* of fictions as non-explanatory without having to resort to truth. All we require is the notion of structural necessitation.

Although the notion of structural necessitation helps us to successfully address the problem of demarcation (additionally to the problem of asymmetry, as mentioned above), it is worth pointing out that the explanatory liberalism I defend is liberal indeed. Take for example the Cartesian explanation of why the planets in our solar system all revolve around the sun *in the same direction* (Laudan 1984). Descartes explained this regularity by invoking the idea of ether vortices. What he thereby did was to represent a regularity as a necessity: the planets *could not* revolve around the sun in different directions, because they were imagined to move on vortices which have only one direction of motion. Of course, there are a plethora of reasons to reject this theory. Yet, explain it did.

One may object that the notion of structural necessitation, despite drawing a line between explanatory and non-explanatory fictions, draws a line that is still too liberal. Consider for example a model that postulates 'hammering pixies' to explain the behavior

---

[15] Even Copernicus used the epicycles to save the phenomena. See e.g. (Kuhn 1957).

of gases. Whenever the pressure of the gas rises after a decrease in the container volume (Boyle's law), the pixies' hammering against the container wall (corresponding to the gas's pressure) will increase, as a decrease in volume will make more of them get pushed to the container walls. Conversely, an increase in the gas volume should result in an decrease in the gas's pressure, as there is no more space for the pixies to roam around freely, away from the container wall.

Even though it may seem easy to come up with outrageous examples such as these that prima facie do seem analogous to the mechanisms postulated by KT or CT, it may be more difficult than it seems to replicate *workable* mechanisms. In this example, there is no reason why the pixies should be in one location and not another. Why do they not all coalesce to the container wall? Because, maybe, they are afraid of the void, or maybe because they just enjoy hammering against the wall? Moreover, why should we assume that they always hammer with the same frequency? Maybe the frequency of their hammering actually decreases when the gas volume is reduced, nullifying the above effect. In other words, this fictional mechanism seems too rich and unconstrained in its resources to offer allow for an explanation that could compete with KT or CT.

Let us look past these problems. What if someone were to come up with a workable outrageous mechanism invoking clearly false fictions such as these? First, we shouldn't forget that we're here dealing with *scientific* models. Science does not invoke supernatural powers, forces, or beings to explain the phenomena. Second, science, particularly physics, is in many ways applied math. Complex beings such as pixies do not lend themselves easily to mathematical treatment. So the pragmatics of science rules out many fancies from the get go.

Let us note finally, that explanatory liberalism does not imply that there are no grounds for choosing one model explanation that engenders structural necessitation over another one that does. Of course, there are much better reasons to believe KT to be true than there are for CT. KT has much wider explanatory scope, as it explains not only IGL, but also the properties of substances in different states, heat transfer and conduction and the specific heat of gases. CT, on the other hand, has been shown to lead to absurd consequences, such as the apparently indefinite production of heat in the boring of cannons (as famously pointed out by Count Rumford), in contradiction with CT's central tenet that heat is a substance obeying the principles of conservation. So explanatory liberalism does not imply that we shouldn't be realists or that we don't have grounds for being realists. The contrary is the case.

5 Concluding remarks

In this paper, I argued that we should give up explanatory conservatism in favour of explanatory liberalism according to which truth is extraneous to good model explanations in science. The means for liberation, I suggested, is the idea that scientific models based on fictive or fictional entities explain their target systems by representing contingencies as necessities (I called this structural necessitation).

One interesting issue to be explored further concerns the way in which the representation of empirical regularities as necessities is brought about in the model. In both KT and CT, the necessity seems to arise from the fictional mechanism consisting of the postulated entities and their properties; it is itself not stipulated. It looks as though logical deduction forms at least part of the relation between these postulates and the necessities[16], but it might not be the whole story. More work will need to be done to clarify these relations. It will also be interesting to explore how widely my account of structural necessitation might apply beyond the examples considered here.

Whatever plausibility explanatory liberalism might have for model explanations in science, one might say, the idea that false explanantia can be explanatory must be outright rejected for our everyday explanatory practices (recall my causal 'getting wet' example). My leanings, too, are with explanatory conservatism in the everyday realm. However, I'm fazed by the possibility that explanatory conservatism may the appropriate view to hold in the everyday realm, even though I firmly believe that it is not in the scientific realm. I admit that this would be counterintuitive. On the other hand, we have come to live with the fact that we possess two incompatible theories at the fundamental physical level. I don't see why philosophy should think itself to be subject to higher demands.

References

Belot, G. and L. Jansson. 2010. Alisa Bokulich, Reexamining the Quantum-Classical Relation: Beyond Reductionism and Pluralism , Cambridge University Press, Cambridge (2008). *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **41** (1): 81-83.

Bokulich, A. 2008a. Can Classical Structures Explain Quantum Phenomena? *The British Journal for the Philosophy of Science*, **59** (2): 217-235.

———. 2008b. *Reexamining the quantum-classical relation : beyond reductionism and pluralism*. Cambridge: Cambridge University Press.

[16] E.g., from the property of inversely proportional mutual repulsion of caloric particles, we can deduce that a container full of caloric will exert a higher degree of force on the container walls than a container with a small amount of caloric.

———. 2009. Explanatory fictions. In *Fictions in science: philosophical essays on modeling and idealisation*, Mauicio Suarez (ed.), New York: Routledge, 91-109.
———. 2011. How scientific models can explain. *Synthese*, **180** (1): 33-45.
———. 2012. Distinguishing Explanatory from Nonexplanatory Fictions. *Philosophy of Science*, **79** (5): 725-737.
———. 2014. How the tiger bush got its stripes:'How possibly'vs.'how actually'model explanations. *The Monist*, **97** (3): 321-338.
———. 2016. Fiction As a Vehicle for Truth: Moving Beyond the Ontic Conception. *The Monist*, **99** (3): 260-279.
Cartwright, N. 1983. *How the laws of physics lie*. Oxford: Oxford University Press.
Chang, H. 2003. Preservative realism and its discontents: Revisiting caloric. *Philosophy of science*, **70** (5): 902-912.
Colombo, M., S. Hartmann, and R. Van Iersel. 2015. Models, mechanisms, and coherence. *The British Journal for the Philosophy of Science*, **66** (1): 181-212.
Contessa, G. 2010. Scientific models and fictional objects. *Synthese*, **172** (2): 215-229.
Craver, C.F. 2007. *Explaining the brain : mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
de Regt, H.W. 2015. Scientific understanding: truth or dare? *Synthese*, **192** (12): 3781-3797.
Doyle, Y., S. Egan, N. Graham, and K. Khalifa. forthcoming. Non-factive understanding: A statement and a defense. *Journal of General Philosophy of Science*.
Elgin, C.Z. 2004. True enough. *Philosophical issues*, **14** (1): 113-131.
———. 2007. Understanding and the facts. *Philosophical Studies*, **132** (1): 33-42.
Frigg, R. 2010. Models and fiction. *Synthese*, **172** (2): 251-268.
Frigg, R. and S. Hartmann. 2012. Models in science. *The Stanford Encyclopedia of Philosophy (Fall 2012 Edition)*, edited by Edward N. Zalta, <http://plato.stanford.edu/archives/fall2012/entries/models-science/>.
Giere, R.N. 1990. *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.
Godfrey-Smith, P. 2009. Models and fictions in science. *Philosophical Studies*, **143** (1): 101-116.
Hempel, C.G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
Hesse, M.B. 1966. *Models and analogies in science*. Notre Dame, Ind.: University of Notre Dame Press.
King, M. 2016. On structural accounts of model-explanations. *Synthese*, **193** (9): 2761-2778.
Kroon, F. and A. Voltolini. 2011. Fiction. *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, edited by Edward N. Zalta.
Kuhn, T.S. 1957. *The Copernican revolution: planetary astronomy in the development of western thought*. Harvard: Harvard University Press.
Lange, M. 2013. What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science*, **64** (3): 485-511.

Laudan, L. 1984. Realism without the Real. *Philosophy of Science*, **51** (1): 156-162.
Levy, A. 2012. Models, fictions, and realism: Two packages. *Philosophy of Science*, **79** (5): 738-748.
———. 2015. Modeling without models. *Philosophical studies*, **172** (3): 781-798.
Machamer, P., L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science*, **67** (1): 1-25.
McMullin, E. 1985. Galilean idealization. *Studies in History and Philosophy of Science Part A*, **16** (3): 247-273.
Morgan, M.S. and M. Morrison. 1999. *Models as mediators: Perspectives on natural and social science*. Cambridge: Cambridge University Press.
Reutlinger, A., D. Hangleiter, and S. Hartmann. forthcoming. Understanding (With) Toy Models. *British Journal for the Philosophy of Science*.
Salmon, W. 1984. *Scientific Explanation and Causal Structure of the World*. Princeton: Princeton University Press.
Schindler, S. 2014. Explanatory fictions—for real? *Synthese*, **191** (8): 1741-1755.
Strevens, M. 2008. *Depth: an account of scientific explanation*. Cambridge, Mass.: Harvard University Press.
———. ms. Why Explanations Lie: Idealization in Explanation.
Suárez, M. 2009. *Fictions in science: philosophical essays on modeling and idealisation*. New York: Routledge.
Thomson-Jones, M. 2010. Missing systems and the face value practice. *Synthese*, **172** (2): 283-299.
Toon, A. 2010. The ontology of theoretical modelling: models as make-believe. *Synthese*, **172** (2): 301-315.
———. 2012. *Models as make-believe: Imagination, fiction and scientific representation*. Basingstoke: Palgrave Macmillan.
Trout, J.D. 2002. Scientific Explanation And The Sense Of Understanding. *Philosophy of Science*, **69** (2): 212-233.
van Fraassen, B. 1980. *The Scientific Image*. Oxford: Oxford University Press.
Votsis, I. and G. Schurz. 2012. A frame-theoretic analysis of two rival conceptions of heat. *Studies in History and Philosophy of Science Part A*, **43** (1): 105-114.
Walton, K.L. 1990. *Mimesis as make-believe: On the foundations of the representational arts*. Harvard: Harvard University Press.
Weisberg, M. 2012. *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.
Winther, R.G. 2016. The Structure of Scientific Theories. *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)*, edited by Edward N. Zalta, http://plato.stanford.edu/archives/spr2016/entries/structure-scientific-theories/.
Woodward, J. 2003. *Making things happen: a theory of causal explanation*. Oxford: Oxford University Press.
Yablo, S. 2014. *Aboutness*: Princeton University Press.