Model Fictions, Structural Necessitation, and Explanatory Liberalism

Abstract

Standard accounts of scientific explanation presuppose that the explanans of a good explanation must be true. Scientific models pose a conundrum to this presumption: how can models explain their targets despite representing their properties in highly idealized, *viz.* literally false, ways? In this paper, I identify a feature of model explanations that has been overlooked: the representation of empirical regularities as necessities. This feature of structural necessitation allows us to drop the requirement of truth in order to accommodate model explanations. At the same time, it helps us avoid explanatory anarchism.

1 Introduction

Explanatory conservatism is the view that the explanans of an explanation must be 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. 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 embraced 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). Despite its 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 actually are 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 models (Bokulich 2008a, Strevens 2008, Batterman 2009, Bokulich 2011, Kennedy 2012, Batterman and Rice 2014, Rice 2015), probably for good reason: given explanatory conservatism, how can models explain their target systems if they misrepresent them so badly? Many of these writers have therefore rejected explanatory conservatism. Yet, a rejection of explanatory conservatism poses a risk: if truth is not a constraint on model explanation, accounts of model explanation must somehow avoid collapsing into *explanatory anarchism* where "any model explanation goes". For this reason, even those how have rejected strict explanatory conservatism have still tended to retain milder forms of it.

In this paper, I identify a feature of model explanation that has hitherto been overlooked: the representation of contingent regularities as necessities within the model. This feature, which I call *structural necessitation*, helps us accommodate the fact that models explain their targets despite their highly idealized and, strictly speaking, false assumptions: we can give up on conservatism and embrace *explanatory liberalism*. At the same time, the notion helps us avoid explanatory anarchism.

The paper is structured as follows. In Section 2 I briefly review the extant literature on model explanations, which has been predominantly conservative. I pose a puzzle for the conservatives to which I don't see a satisfactory solution consistent with explanatory conservatism. In Section 3 I advance my account of structural necessitation and show how it successfully addresses several issues of model explanations, including the threat of explanatory anarchism. In Section 4 I conclude this paper.

2 Mild explanatory conservatism and a puzzle

The fact that models require us to rethink the epistemic practices of science is widely recognized. For example, in the literature on understanding (as distinct from explanation), several authors have argued that scientific understanding doesn't require that the target system is accurately represented by the model (Elgin 2004, 2007, Rohwer and Rice 2013, de Regt 2015, Rohwer and Rice 2016, Reutlinger et al. 2018, Doyle et al. forthcoming).[1] Philosophers concerned with the ontology of models have suggested that we conceive of scientific models as analogous to fictional characters in literary fictions, such as Sherlock Holmes, Anna Karenina, or Tom Sawyer, and modelling practices as games of "make-believe" (Frigg 2010, Toon 2010, 2012, Levy 2015). Yet, despite these recent developments, explanatory conservatism is still widely held. For example, Trout (2002), perhaps the first to highlight understanding as an independent topic from explanation, states that "when it comes to explanation, there is no substitute for simply being (approximately) right" (230). More recently, although Rohwer and Rice (2013) hold that idealised models can give us understanding, they regard "the veridical representation of the features of any real-world system" as necessary condition for explanation (353). Similarly, Reutlinger et al. (2018) grant that what they describe as "toy models", namely "highly simple and highly idealised models", can give us how-possibly understanding, but are sceptical that toy models can give us how-actually understanding when they are not embedded in a true "framework theory".

Even those who have rejected strong forms of explanatory conservatism have not rejected explanatory conservatism outright.[2] For example, although Strevens (2008) has claimed that model explanations are "always better than their veridical counterparts" (300), on his account idealisations explain because they correctly single out and flag only those causal factors that make a difference to the explanandum phenomenon.[3] In other words, idealized models explain by virtue of getting the relevant causal facts *right*.[4] Similarly, Bokulich (2008a, 2009, 2011) has claimed that "fictional" semiclassical models in quantum mechanics are not merely "calculational tools" but give us "genuine physical insight" which is "deeper" than the explanations provided by quantum mechanics itself (Bokulich 2008a, 230-233). At the same time, Bokulich has called on explanatory conservatism to demarcate explanatory from non-explanatory fictions (Bokulich 2012, 735). More specifically she has demanded of good model explanations that 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"

[1] Sullivan and Khalifa (forthcoming) push back against the idea that idealisations (and models) confer any epistemic value.

[2] Recent rejections of strong explanatory conservatism can be found in the works of Bokulich (2008a), Strevens (2008), Batterman (2009), Bokulich (2011), Kennedy (2012), Batterman and Rice (2014), Colombo et al. (2015), Rice (2015).

[3] "Veridical models", in contrast, list *all* causal influences, even if they don't make a difference to the explanandum, such as Mars gravitational pull on an apple falling to the ground on Earth.

[4] Kennedy (2012) argues for a similar view in the case of two idealized models in astrophysics.

(ibid.; added italics).[5] The underlying structures or causes, in turn, are of course accurately described by quantum mechanics. Thus, model fictions explain their targets only insofar their misrepresentations can be related to true bits of the relevant underlying theory.

Mild explanatory conservatism, as one might call the view that makes room for the idealisations of the explanatory targets while sticking to the core conservative principle that requires that explanantia be true, is at no risk of collapsing into explanatory anarchism of the "any explanation goes": explanatory models are delineated from non-explanatory ones by way of their correct identification of difference-makers or by their connection to true theory. Yet, even mild explanatory conservatism presents us with a puzzle: if it ultimately is the actual causes or the true theory that explain the targets, then why do scientists take a detour to models in the first place? Why do they not simply use the "truest" and presumably explanatorily best theories that they possess when trying to explain their explanatory targets? Call this the *model detour problem*.

Mild and staunch explanatory conservatives have given different answers to the model detour problem. Each of them faces difficulties. Bokulich (2008a, 2009, 2011) has argued that model explanations allow us to answer a wider range of what-if-things-had-been-different questions than true theories would (and thereby to gain more understanding of the target systems than we would with true theories alone). It is for this reason that scientists don't just use true theories when they explain their targets. However, as Schindler (2014) points out, there is a tension between this claim and her requirement that model fictions be justified by true theories. Either model explanations are deeper than the explanations of the justifying theory and parts of the model explanation remain unjustified by the underlying theory, or all model statements can be translated into statements of the underlying theory and the explanations provided by the model are not deeper than the explanation of the underlying theory. Bokulich cannot have her cake and eat it too.

According to Strevens (2008), the idealisations of models, contrary to what one may think, do not "lie" about the causally relevant factors (by distorting them), but just highlight those factors that don't make a causal difference to the explanandum phenomenon. This is an interesting suggestion. However, if Strevens is right that the purpose of idealization is a highlighting of non-difference makers, a much more straightforward way to achieve the same goals would be for scientists to simply state which factors are causally relevant and which ones are not. Or scientists could simply just

[5] The idea of a justification of model fictions by true theories via a translation key is already present in her earliest work on the topic (Bokulich 2008a, b).

use what Strevens calls a "canonical model", namely a model containing only those factors that make a causal difference to the explanandum phenomenon. It's not at all clear on Strevens account why scientists ought to prefer the detour of describing causally irrelevant factors in terms of idealization.

Sullivan and Khalifa (forthcoming) argue that falsehoods in models provide no epistemic value but merely aid "in easing calculations and making things salient". In other words, models just make life more convenient for scientists. Furthermore, they argue, that "only the parts of the idealization that approximate their de-idealized counterparts provide the aforementioned epistemic goods [of understanding]".[6] On this view of models, it would seem that if scientists were just cognitively more endowed than they actually are, they could do much better by not taking a detour to models and by using the calculationally more demanding and more complex true theories. In other words, on Sullivan and Khalifa's view models are a bit like crutches: any positive value they may have springs from our own limitations.

The "models as crutches" view is challenged by recent works which have argued that idealisations play an indispensible part in certain explanations in physics and biology (Batterman 2009, Kennedy 2012, Rice 2015). For example, treatments that falsely model gases as continuous fluids in order to account for 'shocks' in gases (i.e., areas of high molecular density within a gas created by external pressure). Without this idealization, Batterman (2009) contends, there wouldn't be any explanation of the target phenomenon. Similarly, Rice (2015) argues that false assumptions in "optimality explanations" in biology, such as the assumption that populations are infinite and that organisms mate randomly, cannot be removed from the model "without consequently eliminating the explanation being offered" (601). Moreover, Rice (forthcoming) argues more generally that models typically cannot be decomposed into accurate and inaccurate parts, as assumed by several conservatives (Strevens 2008, Sullivan and Khalifa forthcoming), but are instead to be viewed as "holistically distorted representations". If all this is correct, then models play a much stronger role in explanation than some conservatives would have it.

In sum, the model detour problem, i.e., question why scientists use models to explain their targets, is still an open question on explanatory conservatism. In what follows, I want to show that we can accommodate the explanatory power of model explanations and thereby avoid the model detour problem without having to hold onto the truth requirement of explanatory conservatism (mild or strong).

---

[6] Sullivan and Khalifa's main focus is understanding, but they also discuss the aforementioned accounts of model explanation.

3 Model fictions and structural necessitation

Not all scientific models idealise to the same degree: some contain more severe distortions than others. Following Suárez (2009), it is useful to distinguish two kinds of 'model fictions': those that distort the target content of entities that we know exist, which he calls 'fictive', and those that postulate entities we know do not exist, which he calls 'fictional'. Although the distinction is probably not a sharp one, there clearly is a difference between models towards the 'fictive' end of the scale, such as models that represents real pendulums as frictionless, and models towards the 'fictional' end, such as models that postulate classical electron orbits in the explanation of quantum phenomena.[7] Another way of describing this contrast is by way of variables: whereas fictive models tend to set parameters of real world entities (such as the actual pendulum or an actual population) to zero or infinity (e.g. no friction or infinite population size), fictional models 'invent' variables for which there is no real world counterpart.[8] A real problem case for explanatory conservatives would be a fictional models that explain their targets without being justifiable by true theory (contrary to what Bokulich would require) without correctly identifying the difference-makers (contra Strevens), and without being de-idealisable (contra Sullivan and Khalifa). I shall argue that there is at least one such model, namely the caloric theory of heat.

Here is the plan for the remainder of the paper. In Section 3.1 will first highlight a feature of model explanations that has been overlooked in previous discussions: explanatory models represent their targets as necessities. I call this feature *structural necessitation*. I will first illustrate structural necessitation with a model that has been widely used in discussions of model explanation, namely the kinetic theory of gases and its explanation of the ideal gas law. After explaining what philosophical work structural necessitation can do for us with regard to several issues in model explanation in Section 3.2, I will then show in Section 3.3 that the ideal gas law is also necessitated by the caloric theory of heat, i.e., a model that cannot be de-idealised, because it postulates entities that clearly don't exist. Allaying conservatives' fears, I will argue in Section 3.4. that structural necessitation and other standard constraints on scientific theorizing help us avoid the clutches of explanatory anarchism.

[7] Electron orbits are ruled out by the Heisenberg uncertainty relation: there cannot be a simultaneously well-defined position and momentum (Bokulich 2008a, b, 2011).

[8] See also Strevens (2008) for thinking of idealization in terms of setting variables to extreme values.

3.1 Structural necessitation in fictive model explanations

The kinetic theory of gases (KT) is an oft-used example in the philosophy of scientific explanation and understanding (Hempel 1970, Friedman 1974, Salmon 1984, Cushing 1991, de Regt 1996, Woodward 2003, Elgin 2004, de Regt and Dieks 2005, Elgin 2007, de Regt 2015, Doyle et al. forthcoming, Rice forthcoming, Sullivan and Khalifa forthcoming). KT idealises actual 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. The numerous idealizations and literally false assumptions of KT allow us to derive 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. KT is widely regarded to explain IGL, which summarizes two laws which I want to keep separate in what follows: Boyle's law ($P \propto \frac{1}{V}$) and the Gay-Lussac law ($P \propto T$).

Consider first 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 gas density, resulting in turn in more molecule-container wall collisions (as molecules have to travel shorter distances in the same time unit before hitting the container wall). Molecule collisions, in turn, are taken cause gas pressure. The Gay-Lussac law, KT explains thus: if the temperature of the gas container is increased (whilst holding the volume fixed), the speed of the gas molecules will increase, which in turn will increase the frequency with which the molecules hit the container wall. The higher the frequency with which the molecules hit the container wall, the higher the gas pressure.

It is quite apparent that this explanation provided by KT lends itself to a counterfactual analysis, which has been championed by Bokulich (2008a, 2011) in the context of model fictions. 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 the gas pressure increased, the temperature would have risen, and vice versa, had the temperature increased, the pressure would have increased (under the condition that the volume remains fixed). It would thus seem that Bokulich's identification of counterfactuals between the model and regularities is important. But it is not the whole story.

A fictive model such as KT also allows us to *represent empirical (and contingent) regularities as necessities*. 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.

Broadly speaking, model explanation by structural necessitation is a form of the *modal conception* of explanation. The modal conception of explanation, according to Salmon (1998), "says that a good explanation shows that what did happen had to happen" (321). Salmon conceived of the modal conception mostly in terms of a relation between a *physically necessary* law of nature and an event-to-be-explained (Salmon 1984).[9] Although this makes the modal conception sound similar to the classical DN model of explanation, Salmon points out that in contrast to the classical DN model of explanation, in the modal conception the explanandum is not explained by deriving it from a law, but rather by showing that the explanandum event is "physically necessary relative to the explanatory facts" (Salmon 1984, 111). The modal conception hasn't had many followers. Recently, however, Lange (2013, 2017) has revived the conception in the context of mathematical explanations in science and what he calls "explanation by constraint", which put necessary constraints on the laws of nature (as for example conservation principles which constrain force laws of various kinds to conserve energy). For Lange, mathematical explanations are modally "stronger" than explanations by constraint, which in turn are modally stronger than what he considers "causal necessities".[10]

Structural necessitation, contrary to Salmon's idea of what the modal conception amounts to, is not about an explanatory relation holding between a law and an

[9] Salmon traces the modal conception back to D.H. Mellor (1976), von Wright (1971), and even Aristotle. Mellor has a modal conception of causal explanation according to which "things could not have happened otherwise than the [causal] explanandum says" (Mellor 1976, 235) and von Wright believes that identifies the necessity of the explanandum's occurrence as the main source of explanation in the DN model.

[10] Lange also speaks of a "modal pyramid" whose rungs can be determined by counterfactuals. E.g., conservation of energy would have held even if the force laws had been different ones than the ones we know.

explanandum *event*. Rather, it holds between the laws of the model and the empirical regularities in nature. Structural necessitation thus explains the particular contingent *form* of empirical *relations*. What it shares with Salmon's modal conception is that structural necessitation, too, "shows that what did happen had to happen". More specifically what gives us understanding in model explanations of the target system, on the notion of structural necessitation, 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. 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 (Grüne-Yanoff 2013, Rohwer and Rice 2013, Reutlinger et al. 2018)). Because empirical relations are represented in the model in terms of the entities it postulates, we can also speak of an *isomorphism* between the empirical regularities and the relations in the model that represent those regularities as necessities. For example, an increase in gas pressure in KT is represented as an increase of molecule-wall collisions (isomorphism 1), which in turn results from a decrease of gas volume, which in KT is represented as an increase in gas density (isomorphism 2).[11]

Of what kind are the necessities in structural necessitation? Clearly the necessities within KT, minimally, are intended to be *physical* necessities (on the same modal level as Lange's "causal necessities") that arise in virtue of the postulated substructures. 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 *physical necessities in a fiction*. Although there will have to be said much more about the nature of these necessities in the future, 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, the classical laws of physics apply to the highly idealised molecules postulated by KT, even though the molecules of real gases have rather different properties.

---

[11] Bokulich (2011) also speaks of isomorphisms between the counterfactual structure of the model and the counterfactual structure of the target. However it's not clear from her writing what these isomorphisms amount to beyond there being a counterfactual dependence relation between the model and the target. The concept of isomorphism does not occur in her earlier or later works on the topic (Bokulich 2008a, b, 2009, 2012). For more on isomorphism, and in particular partial isomorphism, in the context of models see da Costa and French (2003).

An issue that has been raised against the modal conception of explanation is that it fails to accommodate statistical and indeterministic explanations, in particular, when it comes to such explanations of *events* (Salmon 1984, 1998). 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, which is in many ways idealised, 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, the Mendelian 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 explanation is supposed to explain regularities (probabilistic or non-probabilistic), 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. However, in this case, a good account of explanation must accommodate only what the Mendelian model actually aims to explain, namely the character distributions under normal or ideal experimental circumstances.

Finally, let me add a remark on the regularities explained model explanations. I have characterized those laws as contingent. One may object, however, that on certain accounts of lawhood, laws as conceived of not as contingent but as necessities themselves (Armstrong 1978, Dretske 1978, Armstrong 1983). If one were to adopt such a view of lawhood, the objection continues, my notion of structural necessitation would be vacuous. Let us note, however, that even on such accounts the necessities associated with lawhood are continent. So even if one were to subscribe to such a view of lawhood, contingent necessities could still be represented as non-contingent. There are in fact reasons not to hold such a view in the first place. Lewis (1983) has criticized that the notion has not much explanatory force in the context of elucidating lawhood. Also, many laws of nature are ceteris paribus laws that hold only under certain conditions (Reutlinger et al. 2017). This fact seems to be hard to reconcile with a view that treats laws of nature as inviolable connection between universals.

3.2 Advantages of the account

I now want to advertise some benefits of integrating the notion of structural necessitation into accounts of model explanation.

3.2.1 Models: no detours

In Section 2 I pointed out that explanatory conservatives face the model detour problem: given their commitments, the use of models in the quest of providing explanations looks suboptimal and is only justified by our own cognitive limitations. On my view, in contrast, the use of models in science is justified by their real explanatory purchase. The model detour problem thus dissolves.

3.2.2 Indispensable idealisations

As mentioned in Section 2, there are philosophers who have claimed that the idealisations of some models cannot be de-idealised without losing the explanation that those models provide—without explanatory substitute of true theories (Bokulich 2008a, Batterman 2009, Batterman and Rice 2014, Rice 2015, forthcoming). On the account offered here, models can explain their targets even if their assumptions cannot de-idealised, so long as they represent their targets as necessities.

3.2.3 Asymmetry

A traditional problem for accounts of explanations has been 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 of the height of the flagpole (but only vice versa, the height of the flagpole as the explanans of the length of the shadow).[12]

On counterfactual accounts of model explanation, such as Bokulich's, explanatory relations supervene on counterfactual dependencies. But counterfactual dependence is a symmetric relation. In Bokulich's own example it is e.g. not only true that had the shape of the electron orbit been different, the spectral lines would have been different, but also vice versa, it is true that had the spectral lines been different, the shape of the electron orbits would have been different. On what basis can we then say that the model explains the explanandum phenomenon *but not vice versa*? (see Schindler 2014). Structural necessitation offers a natural solution to the problem of asymmetry: it is the model that represents contingent regularities as necessities, and not vice versa.[13]

[12] Khalifa et al. (forthcoming) call the problem of asymmetry a "surefire way to embarrass a theory of explanation".

[13] Bokulich (2012) has tried to address this problem by way of what she calls a "contextual relevance relation". See Jansson (2015) and Khalifa et al. (forthcoming) for other proposals of how to solve the symmetry problem without appeal to causation.

3.2.4 Explanatory demarcation: beyond the dichotomy

A pressing (and legitimate) concern of explanatory conservatives is explanatory anarchism: our philosophical theory of model explanation must avoid the slide to the "any explanation goes". But if it's the case that fictional models explain their targets, as I shall argue in a moment, then what constrains the class of good model explanations, if not truth? I want to suggest that structural necessitation does.

Explanatory conservatism assumes that there are only two classes of model explanations: approximately true and explanatory models on the one hand, and false and non-explanatory models on the other hand. On the view defended here, this dichotomy is too simple. Instead, I propose, there are in fact at least three classes of model explanations: On the one hand, there are fictive models that explain, for example, by correctly identifying difference-makers. Some of those fictive models will be models whose explanation in part consists of representing their targets as necessities. On the other hand, there are models that are entirely fictional (and e.g. fail to identify difference-makers) but which nevertheless explain by structural necessitation. Figure 1 visualizes these possibilities.

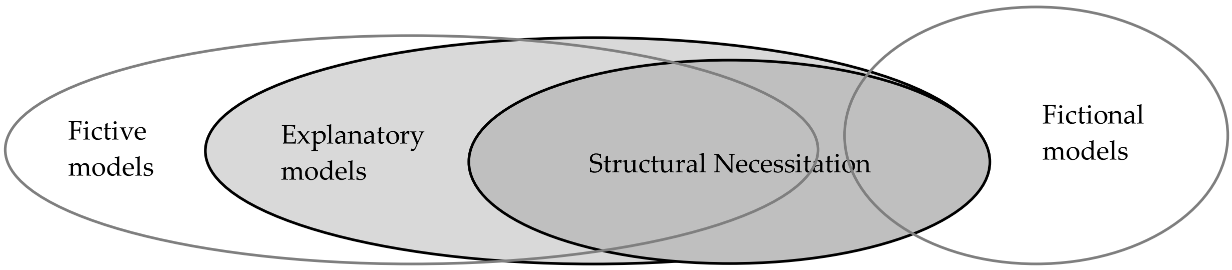

Figure 1: Fictive, fictional, and explanatory models and structural necessitation.

Let us consider an example of a fictional model that is widely agreed not to explain its target, namely the Ptolemaic model of our planetary system (Belot and Jansson 2010, Bokulich 2012).[14] Notoriously, the system employed the entirely fictional device of the epicycle to save various celestial phenomena. Conservatives deem the model non-explanatory because it does not latch onto the real structure of the world. I now want to show that in addition to that, the Ptolemaic model fails to represent its target regularities as necessities. Structural necessitation thus affords another, independent, way of constraining the set of good model explanations. What's more, it offers a way of demarcation that is epistemically less demanding than the conservative's solution: we don't need to be able to presuppose some privileged access to the truth or rely on an

[14] Belot and Jansson (2010) use this example in their criticism that Bokulich's account cannot rule out non-explanatory models. See Bokulich (2012) for a reply and Schindler (2014) for an assessment.

inference to the best explanation (as realists would) and we do not need to have the privilege of historical hindsight in order to determine whether a model is explanatory or not.

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 Figure 2).

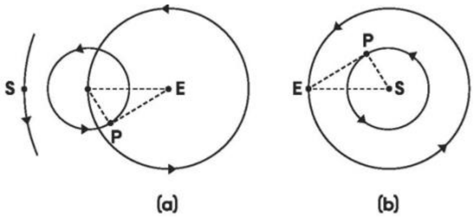

Figure 2: maximum elongation as accounted for in the Ptolemaic and the Copernican system (a and b, respectively). From Kuhn (1957).

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. We can thus disqualify some fictions, or more accurately, some *uses* of fictions as non-explanatory without having to appeal to (the epistemically more demanding) notion of truth.[15] All we require is the notion of structural necessitation.

Although the notion of structural necessitation helps us to rule out some fictional models as non-explanatory, the explanatory liberalism I defend is liberal indeed. In the following section I want to show that the same contingent regularity that KT represents as a necessity, namely the ideal gas law, is in fact also be represented as such on the utterly false caloric theory of heat.

[15] Whether or not a fiction is explanatory, I think, depends on the context in which it is used. If it is used in a context in which it allows the structural necessitation of regularities, then it plays an important part in the explanation of the those regularities.

3.3 Structural necessitation and explanatory liberalism

In contrast to KT, where heat is correctly represented as molecular motion (with a number of idealizing assumptions), in the caloric theory of heat (CT) heat is conceived of as a *substance*. CT furthermore supposes that caloric particles attract matter particles and that caloric particles repel each other (Figure 3).

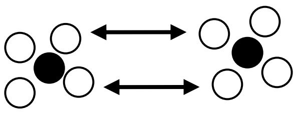

Figure 3: Caloric particles (white) repel each other but are attracted by matter particles (black).

CT was a strong competitor to KT for decades. CT was held by the likes of Joseph Black, who used CT to explain changes of state, 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 law, a reduction of gas volume, on CT, would *have to* lead to an increase in the pressure on the container wall, as the mutual repulsion of caloric particles would be stronger the closer the particles are to each other. With regards to Gay-Lussac's law, an increase in the amount of caloric would *have to* result in an increase in pressure, because the supply of further caloric particles would increase the net amount of repulsion between them (all whilst the volume remains fixed, of course).[16]

Table 1 shows that CT represents IGL as a necessity just like KT represents IGL as a necessity. In both KT and CT structural necessitation is an important part of their explanation of IGL. Even though we may have good reasons to say that KT is a *better* explanation than CT because it e.g. identifies the correct difference makers (more on that in a moment), this does not legitimize the claim that CT provides *no* explanation of IGL. On the contrary CT, just like KT, shows us why IGL has to have the form that it does have (on the assumptions of CT) and thereby explains it.

[16] 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).

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

Table 1: KT, CT, and Boyle's and Gay-Lussac's laws. In both KT and CT contingent regularities are represented as necessities. FC: frequency of collisions with container walls. SR: strength of caloric repulsion.

One might be tempted to think that CT explains by virtue of latching onto some correct structure that is identified by KT in a way that structural realists have argued for the structural continuity in the history of science despite radical theoretical change (Votsis and Schurz 2012). On this picture, CT could be granted explanatory power by the critic but only in virtue of the explanatory power of 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 in his false theory of the aether. But the structural continuity between CT and KT is just IGL, i.e., an *empirical* regularity. All that CT and KT do is represent this regularity as necessities. And they do so in radically different ways. That is not enough for structural realist arguments. Thus, insofar as there is structural continuity between KT and CT, it isn't of the right kind and insofar as the structure is of the right kind, it is radically different.

It must be emphasized that explanatory liberalism is consistent with the view that KT provides a *better* explanation of IGL than CT does, because KT arguably identifies at least some causal difference-makers (in particular, the fact that gases consist of molecules and that heat is not a substance). For instance, 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 of gases. Similarly, has made many successful predictions (such as the specific heat ratios of gases (de Regt 1996)) and has overall provided a very *fertile* research programme (Musgrave 1976). In contrast, CT soon ran into problems that it could not solve, such 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. More generally speaking, *theoretical virtues* such as scope and fertility can help us to narrow down which of those explanations that manage to represent their target regularities as necessities are good truth candidates. Those would obviously be *better* explanations than models that have very limited scope and fertility. Nevertheless the account presented here, in contrast to explanatory conservatism, allows us to accommodate the fact that CT was widely-held in the scientific community also for its explanatory accomplishments. In other words, whereas on explanatory conservatism there are only good and approximately true models on the one hand, and bad and false models on the other hand, explanatory liberalists can allow for fictive models to be *better* explanations of their target systems than fictional models *without* having to conclude that fictional models must therefore be non-explanatory.

3.4 Overshooting the target?

Even though the notion of structural necessitation helps us to reject certain model fictions as non-explanatory (such as the Ptolemaic model of our planetary system), one may object that it leads to an account of model explanation that is still too liberal. Consider for example a model that postulates 'hammering pixies' to explain the behavior of gases.[17] 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 result in more of them getting 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, why should we assume that the pixies always hammer with the same frequency? Maybe the reduction in volume actually makes them want to hammer less frequently? Or why should we assume that they keep hammering against the wall at all when the volume is compressed? Perhaps they get so scared by the moving walls and all fly to the centre of the container in order to comfort each other/ In other words, this fictional mechanism seems too rich and

[17] I owe this example to <blinded for peer-review>.

unconstrained in its resources to allow for an explanation that could compete with KT or CT.

Let us look past these specific 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. Some have even argued that idealisations of models are justified by the applicability of mathematical apparatus (Rice forthcoming). 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. Third, and perhaps most importantly, models that are not confirmed by the evidence should not be candidates for good model explanations. It is hard to see that the hammering-pixies mechanism would survive severe scientific testing for long. More generally speaking, there are several standard constraints on scientific theorizing that will supplement structural necessitation in singling out explanatory models.

4 Concluding remarks

In this paper I argued that we can accommodate the explanatory power of highly idealized and literally speaking false models by the notion of structural necessitation. This notion allows us to be liberalists about the truth requirement endorsed by explanatory conservatives without having to give in to explanatory anarchism.

Although explanatory liberalism is counterintuitive particularly in the scientific realm, this alone shouldn't be a reason for rejecting explanatory liberalism outright. We have learned to live with the fact that the micro-physics described by our best theory is very far removed from the phenomena as we encounter them in the everyday realm. There is no reason why philosophy should answer higher demands of consistency when it comes to reconciling science and common sense.

Bibliography

Armstrong, D.M. 1978. *A theory of universals*. Cambridge: Cambridge University Press.

———. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

Batterman, R.W. 2009. Idealization and modeling. *Synthese*, **169** (3): 427-446.

Batterman, R.W. and C.C. Rice. 2014. Minimal model explanations. *Philosophy of Science*, **81** (3): 349-376.

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.
———. 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.
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.
Craver, C.F. 2007. *Explaining the brain : mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
Cushing, J.T. 1991. Quantum theory and explanatory discourse: endgame for understanding? *Philosophy of Science*, **58** (3): 337-358.
da Costa, N.C. and S. French. 2003. *Science and partial truth: A unitary approach to models and scientific reasoning*: Oxford University Press.
de Regt, H.W. 1996. Philosophy and the Kinetic Theory of Gases. *The British Journal for the Philosophy of Science*, **47** (1): 31-62.
———. 2015. Scientific understanding: truth or dare? *Synthese*, **192** (12): 3781-3797.
de Regt, H.W. and D. Dieks. 2005. A contextual approach to scientific understanding. *Synthese*, **144** (1): 137-170.
Doyle, Y., S. Egan, N. Graham, and K. Khalifa. forthcoming. Non-factive understanding: A statement and a defense. *Journal of General Philosophy of Science*.
Dretske, F. 1978. Laws of nature. *Philosophy of science*, **44** (2): 248-268.
Elgin, C.Z. 2004. True enough. *Philosophical issues*, **14** (1): 113-131.
———. 2007. Understanding and the facts. *Philosophical Studies*, **132** (1): 33-42.
Friedman, M. 1974. Explanation and scientific understanding. *The Journal of Philosophy*, **71** (1): 5-19.
Frigg, R. 2010. Models and fiction. *Synthese*, **172** (2): 251-268.
Grüne-Yanoff, T. 2013. Appraising models nonrepresentationally. *Philosophy of Science*, **80** (5): 850-861.
Hempel, C.G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

———. 1970. On the 'standard conception' of scientific theories. In *Minnesota studies in the philosophy of science*, M. Radner and S. Winokur (eds.), Minnesota: Minnesota University Press, 142-163.

Hesse, M.B. 1966. *Models and analogies in science*. Notre Dame, Ind.: University of Notre Dame Press.

Jansson, L. 2015. Explanatory asymmetries: Laws of nature rehabilitated. *The Journal of Philosophy*, **112** (11): 577-599.

Kennedy, A.G. 2012. A non representationalist view of model explanation. *Studies in History and Philosophy of Science Part A*, **43** (2): 326-332.

Khalifa, K., J. Millson, and M. Risjord. forthcoming. Inference, explanation, and asymmetry. *Synthese*, **OnlineFirst**.

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.

———. 2017. *Because without cause: Non-causal explanations in science and mathematics*: Oxford University Press.

Levy, A. 2015. Modeling without models. *Philosophical studies*, **172** (3): 781-798.

Lewis, D. 1983. New work for a theory of universals. *Australasian Journal of Philosophy*, **61** (4): 343-377.

Machamer, P., L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science*, **67** (1): 1-25.

Mellor, D.H. 1976. Probable explanation. *Australasian Journal of Philosophy*, **54** (3): 231-241.

Musgrave, A. 1976. Why did oxygen supplant phlogiston? Research programs in the chemical revolution. *Method Appraisal Phys. Sci*: 181-209.

Reutlinger, A., D. Hangleiter, and S. Hartmann. 2018. Understanding (With) Toy Models. *British Journal for the Philosophy of Science*, **69** (4): 1069–1099.

Reutlinger, A., G. Schurz, and A. Hüttemann. 2017. Ceteris paribus laws. *Stanford encyclopedia of philosophy*, edited by Edward N. Zalta, <https://plato.stanford.edu/archives/spr2017/entries/ceteris-paribus/>.

Rice, C. 2015. Moving beyond causes: Optimality models and scientific explanation. *Noûs*, **49** (3): 589-615.

———. forthcoming. Models Don't Decompose That Way: A Holistic View of Idealized Models. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axx045.

Rohwer, Y. and C. Rice. 2013. Hypothetical pattern idealization and explanatory models. *Philosophy of Science*, **80** (3): 334-355.

———. 2016. How are models and explanations related? *Erkenntnis*, **81** (5): 1127-1148.

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

———. 1998. *Causality and explanation*. New York: Oxford 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.
Suárez, M. 2009. *Fictions in science: philosophical essays on modeling and idealisation*. New York: Routledge.
Sullivan, E. and K. Khalifa. forthcoming. Idealizations and Understanding: Much Ado About Nothing? *Australasian Journal of Philosophy*.
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.
von Wright, G.H. 1971. *Explanation and understanding*: Cornell 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.
Woodward, J. 2003. *Making things happen: a theory of causal explanation*. Oxford: Oxford University Press.