Explanation by structural necessitation

Abstract

In this paper I explore a type of explanation according to which contingent regularities are explained by representing them as a necessities in a model. I consider this type of explanation, which I call *explanation by structural necessitation*, a species of the modal conception of explanation. I discuss a classic objection to the modal conception and compare explanations by structural necessitation to Lange's explanations by constraints.

1 Introduction

Salmon (1989, 1998) famously distinguished between the epistemic, ontic, and modal conception of scientific explanation. In the epistemic conception, the explanandum is explained by making it expected by way of an argument, containing a law of nature. In the ontic conception, which Salmon himself preferred, the explanandum is fitted into the causal-mechanistic structure of the world. And finally, in the modal conception the explanandum has got to occur by necessity.

The epistemic conception, in the form of the classical DN model of explanation, dominated the first couple of decades of philosophical ruminations about scientific explanation since the mid-twentieth century (Hempel 1965). The fortunes started to shift in the mid-1980s towards the ontic conception with Salmon's work, and subsequently became the preferred conception, in particular with Woodward's work in the early 2000s (Woodward 2003).

The modal conception for a long time didn't have strong advocates. This changed with the recent work by Lange (2013, 2017) and his account of explanation by constraint. In this paper, I want to discuss another form of the modal conception of explanation. I call it *explanation by structural necessitation*.

I will begin the paper by giving a short exposition of what I consider to be an exemplar of the kinds of explanations in which I'm interested in this paper, namely the explanation of the ideal gas law by the kinetic theory of gases (Section 2). In Section 3 I will introduce my account of explanation by structural necessitation. In Section 4 I will address Salmon's objections to the modal conception and in Section 5 I will compare my account to Lange's. In Section 6 I will discuss the problem of explanatory demarcation, before I will then conclude the paper in Section 7.

2 The kinetic theory of heat: derivation, counterfactuals, and Woodward

The kinetic theory of gases (KT), developed during the mid-and late 19th century, postulates that gases consist of molecules in motion and makes a number of idealizing assumptions about their properties, such as that the extension of the molecules is negligible and that there are no intermolecular collisions.[1] KT's explanandum is the ideal-gas law (IGL), namely $PV = nRT$, 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. For what follows it is useful to know that IGL is in fact a summary of two empirical laws: Boyle's law ($P \propto \frac{1}{V}$) and the Gay-Lussac law ($P \propto T$). The former states that the pressure of a gas is inversely proportional to the gas volume and the latter tells us that pressure is proportional to temperature. How does KT explain IGL?

Textbook presentations of KT discuss derivations of IGL from KT. What follows is a truncated version of q derivation of part of IGL, namely Boyle's law.[2] First of all, let *l* be the distance between two opposite container walls and v_x = the molecule velocity along the x-axis. Then for a molecule on a 'round trip' between opposite sides of the container wall, the number of collisions of a molecule with the container wall per second is ${v_x}/{2l}$. The total momentum change per second for all molecules (*N*) in the container with average speed $\overline{v_x}$ is $N\frac{m\overline{v_x^2}}{l}$. By Newton's second law, the change of momentum is equivalent to the average net force exerted by the molecules on a container wall per second. Since pressure is defined as the force exerted perpendicularly on an area (here: l^2), $P = N\frac{m\overline{v_x^2}}{l^3}$. But l^3 is just the volume of a cubical container, so we can write $P = \frac{Nm\overline{v_x^2}}{V}$. Assuming that there is no preferred direction of a molecule's path in the three spatial dimensions, $\overline{v_x^2} = \frac{1}{3}\overline{v^2}$, and $PV = \frac{1}{3}Nm\overline{v^2}$. Since the *total* kinetic energy of translation KE_{trans} of all molecules is $N \times \frac{1}{2}m\overline{v^2}$, this gives us $PV = \frac{2}{3}KE_{trans}$. This is the so-called *gas pressure equation*: it relates the macrovariables of gas pressure and volume to the microvariable of translational energy of molecules. Finally, under the assumption that molecular speed (v^2) depends only on temperature (to be proven separately), it follows from the gas pressure equation and the ideal gas law $PV = nRT$ that the gas pressure is proportional to the volume of the gas when T is constant. This is Boyle's law.

On the classical Deductively-Nomological model of explanation, KT explains IGL because IGL can be deductively derived from the laws of KT. However the DN model is beset with problems (see Woodward 2014) and it is generally agreed that it doesn't successfully capture important features of scientific explanations. One such important feature concerns counterfactual dependence relations, as highlighted by Woodward (2003) and by others since then (e.g. Bokulich 2011, Saatsi and Pexton 2013, Reutlinger 2016, Reutlinger and Saatsi 2018).

It is easy to see that IGL describes counterfactual dependency relations, e.g. had T had another value than it actually has, P would have had a different value (under the condition that V stays fixed).[3] Generalizations such as IGL allow us to answer what-if-things-had-been-different-

---

[1] KT is called a theory for historical reasons. Given its idealizations it is perhaps more appropriately described as a model.

[2] E.g. Holton and Brush (2001).

[3] Woodward discusses the ideal gas law briefly on pages 234 and 250-1 of his book (Woodward 2003).

questions" (w-questions, for short), which Woodward considers essential for identifying causal relations and for understanding the relevant physical systems.

Although Woodward does not explicitly mention KT in his book, he does discuss a "microscopic strategy" for explaining why P changed to P′ after a change of V to V′ (Woodward 2003, 231-232). In this microscopic strategy, one would consider the initial state of the system (consisting of the initial positions and momenta of 6 $x10^{23}$ molecules in the container) and the evolving energy and momentum "of each molecule in terms of its initial state, the successive collisions it undergoes with other molecules and the laws governing those collisions". The microscopic strategy would try to explain the new value of P by "aggregating the energy and momentum transferred by each molecule to the walls of the container". Yet Woodward deems such an explanatory strategy unsatisfactory because it fails to answer the relevant w-questions: it doesn't tell us determinately what value P would have had, had the initial microstate been different. That is so because any macrostate (such as a particular value of P) is compatible with basically any initial state of molecules and its evolution (with different molecule trajectories). Woodward concludes that the microscopic strategy "fail[s] to provide the explanation of the macroscopic behavior of the gas we are looking for", because it "omits information that is crucial to an explanation of pressure" (namely counterfactual information). For Woodward, IGL simply does a "better job" at explaining the change from P to P′. Woodward believes that this lesson is a general one and that it is often the case that "lower-level" explanations such as KT do worse than "upper-level" explanations such as IGL (233).

I believe Woodward's dismissal of the "microscopic strategy" does not do full justice to KT;[4] KT is *standardly* presented as an explanation of IGL in physics textbooks. Of course physicists could be deceived about KT (and in addition many of the philosophers writing about explanation), but I think we can do better than this.

Contrary to Woodward I think that we *can* identify counterfactual dependencies between KT and IGL, despite the underdetermination of macrostates by microstates of gases. For example, an increase in the frequency of molecule-wall collisions *will* result in a higher value of P, no matter what the initial conditions of the molecules might have been. That is so because in KT molecule-wall collisions just underlie the macrophenomenon of pressure. Thus, contra Woodward, we *can* say how the value of P would have changed, had the frequency of the molecule-wall collisions changed.

In fact, as we shall see in the remainder of the paper, theories or models like KT do even more than provide counterfactual information; they tell us why the counterfactual dependencies described by empirical generalizations like IGL *have got to be* the way they are.

[4] Note that Woodward mentions "statistical mechanics" on p. 223 of his book (referring the reader to the paragraph I discuss here), which is a term that is often used to subsequent developments of KT.

3 Explanation by structural necessitation

IGL, as an empirically discovered relationship, is contingent: in a different world, it might have been the case that the pressure of a gas actually *decreases* when the temperature increases (in contradistinction to Gay-Lussac's law), and that the pressure of a gas actually *increases* when the gas container volume decreases (in contradistinction to Boyle's law). In another possible world, that is, the counterfactual relations between the macrovariables in IGL might have been reversed. So why does IGL have the form that it actually does have in the actual world?

This is where KT comes in: it tells us that IGL *must* have the form that it does have and that, for example, a gas's pressure *must* increase when the volume is decreased (in accordance with Boyle's law). In other words, KT necessitates IGL. Because KT tells us why the *relations* between the variables of IGL have to be the way they are, I shall call this feature of KT *structural necessitation* (see Table 1).

IGL regularity	**Explanation in terms of structural necessitation in KT**
$P \propto {}^{1}/_{V}$ (Boyle's law)	FC ∝ gas density *Not possible*: FC ∝ 1 / gas density
$P \propto T$ (Gay-Lussac's law)	FC ∝ speed of molecules *Not possible*: FC ∝ 1 / speed of molecules

Table 1: KT and Boyle's and Gay-Lussac's laws (as parts of IGL). In KT contingent regularities are represented as necessities. FC= frequency of molecule-wall collisions. KT presupposes the following relations between macro- and microvariables: P ∝ FC, V ∝ 1/gas density, and T ∝ molecular speed.

Let us try to characterize KT's explanation of IGL more broadly. First of all, what KT does in order to explain IGL is to postulate a *model* underlying the phenomena that consists of *entities* and their *activities*. In other words, KT postulates a *mechanism* for IGL (Machamer et al. 2000). Although mechanisms have standardly been interpreted in realist terms, this need not be so (Colombo et al. 2015). And as mentioned, there are several aspects about KT that are unrealistic (e.g. the neglect of intermolecular forces, neglect of spatial extension of molecules, etc.).

Second, in the model there is an *internal* counterfactual structure, based on the entities and activities in the postulated mechanism, which is *isomorphic* to the *external* counterfactual structure of the empirical regularity. For example, IGL supports the counterfactual "had the temperature increased, the pressure would have increased" (left hand side of Table 1). This structure is isomorphic to the structure in KT of "had the speed of the molecules increased, the frequency of the molecule-wall collisions would have increased" (right hand side of Table 1).[5]

Third, and crucially, the model's counterfactual structure *represents* the counterfactual structure of the empirical regularity as a *necessity*. This is done by the model *ruling out* a possible counterfactual structure that could have obtained empirically. E.g., KT, by postulating a

---

[5] An isomorphism between counterfactual structures in the model and the explanandum is also a condition in Bokulich's (2011) account. See also fn 7. For more on isomorphism, and in particular partial isomorphism, in the context of models see da Costa and French (2003).

mechanism and by employing the laws of classical mechanics, rules out that the frequency of the molecule wall collisions would decrease (rather than increase) if the speed of the molecules had increased (see right hand side of Table 1). In other words, the representation of an empirical regularity as necessity within a model allows us to see *why* empirical laws *have to take* the form that they actually do take. The understanding provided by explanations by structural necessitation is therefore not 'how-possibly' understanding, which has recently been identified as particularly relevant to explanations provided by scientific models (Grüne-Yanoff 2013, Rohwer and Rice 2013, Reutlinger et al. 2018), but rather 'how-necessarily' understanding.

That models represent regularities as necessities is central to their explanations: given the isomorphism between the counterfactual structures of the model and the empirical regularity, it would seem that there wouldn't be any extra counterfactual information that we could gain from the model that is not already contained in the empirical regularity itself. The model would seem simply redundant. For example, if one knows that the pressure of a gas will increase when the temperature increases (so long as the volume stays fixed), then one doesn't gain any additional counterfactual insight from knowing that the frequency of the container-wall collisions (corresponding to pressure) increases when the molecular speeds (corresponding to pressure) increase. So again, the third feature of explanations by structural necessity (representation as necessity) is quite central to the kinds of explanations under focus here. Let us consider another example.

The Bohr model of the atom (proposed in 1913) postulated that electrons moved in stable classical orbits around the nucleus. How did the model explain the characteristically discrete spectral lines of hydrogen captured by the Rydberg formula $\frac{1}{\lambda} = R(\frac{1}{{n_1}^2} - \frac{1}{{n_2}^2})$?[6] First, the model proposes a (non-classical) *mechanism* consisting of entities and their activities: electrons, their orbiting the nucleus, and their 'jumping' between orbits. Second, there is a counterfactual structure in the Bohr model that is part of its explanation (e.g., had the electron jumped to higher energy orbit, the atom would have absorbed a photon of an energy-equivalent frequency), and there is an isomorphism between the counterfactual structure of the Bohr model and the counterfactual structure encoded in the empirical law (e.g., had n_2 been changed, the wavelength λ would have changed).[7] Third, and crucially, the structure of the Bohr model *represents* the regularities and counterfactual dependencies captured by the Rydberg formula as *necessities*: the electron can occupy only discrete stationary orbits within the model. Whenever the electron 'jumps' from an orbit with higher energy to an orbit with lower energy, it emits energy in the form of photons (the reverse is true when the atom absorbs energy). Now, given that in the Bohr model it *is not possible* for electrons to occupy positions outside of the fixed stationary orbits, the model explains why the measured line spectrum of hydrogen *has got to be* discrete rather than continuous.

[6] In the Rydberg formula, n_1 determines the kind of spectral line series, e.g. the Balmer series for n_1=2.

[7] That there is an isomorphism between counterfactual structures in the Bohr model and its explanandum has been pointed out by Bokulich (2011), although it is not entirely clear from her writing what precisely the counterfactual structures are supposed to be between which an isomorphism is supposed to hold.

In the two examples we've discussed so far, the necessity concerned in the explanations is a *physical* necessity: it is brought about by a combination of basic assumptions and physical laws that can be derived from more fundamental theories (classical mechanics, in particular). But there are also some explanations by structural necessitation which derive their necessity from a combination of basic assumptions and *mathematics*. Let us consider two examples.

Dalton's atomism (proposed in the early 19th century) explains the laws of definite proportions and the law of multiple proportions. The former law says that chemical elements always combine in the same ratios. E.g. oxygen always making up 8/9 and hydrogen 1/9 of the mass of water. The latter law says that whenever two elements form more than one compound, the ratio of those compounds will be multiples of each other, as e.g. in CO and CO_2, where a certain amount of carbon combines with exactly twice as much oxygen in the second compound as it does in the first. Dalton's atomism explains not only the fact *that* we observe these ratios in the two laws, but also why we do *not* observe any intermediate ratios. E.g. we wouldn't observe that one unit of carbon combines with 1 ½ as much oxygen as it does in CO. Such a combination is ruled out by the assumption of indivisible atoms: it is for this reason that units of elements combine with each other only in integer multiples.

In sum, Dalton's account provides a *model postulating entities* (one kind of atom per chemical element) that combine in certain ways (namely as integral, fixed entities, not e.g. by splitting) in order to *represent* empirical regularities (the laws of definite proportions and multiple proportions) as necessities. Representing these regularities in this way *explains* why they *have to* take the form that they actually do take ('integer' proportions, rather than 'fractional' proportions). The necessities concerned in the explanation are brought about by a combination of the assumption of the indivisibility of atoms and (the thereby justified use of) natural numbers (rather than e.g. rational numbers) in the determination of multiple compounds. Let us consider a final example.

Mendel discovered in his hybridization experiments with pea plants in the mid-19th century that when he self-fertilized the offspring of plants from the first filial generation that resulted from crossing purebred (recessive) white and (dominant) purple flower plants, the second filial generation (F2) would consist of a 3:1 ratio of purple vs. white coloured plants. Mendel explained this by invoking of unitary genetic 'factors' (as he called them) that come in pairs for each organism (alleles). Mendelian model – just like the Daltonian model – represents regularities as necessities by means of basic (combinatorial) mathematics: for each hybridized pair of plants there are exactly four possible combinations of alleles (two per plant). Together with the principle of dominance, according to which traits of the dominant genetic 'factors' are always expressed phenotypically, the self-fertilization of plants from F1 *has got to* result in a 3:1 ratio in F2. A 2:2 ratio in F2, for example, would be impossible on the model.[8]

In sum, Mendel's account provides a *model postulating entities* (genetic 'factors') that combine in certain ways (one allele from each organism, either dominant or recessive) so as to represent

---

[8] The ratios can also be expressed as probability distributions (e.g. 75% vs. 25%). The point remains that Mendel's model necessitates such distributions. See also Section 4.

empirical regularities (Mendel's laws) as necessities. Representing these regularities in this way explains why they *have to* take the form that they actually do take (ruling out any other proportion than 3:1 in the F2 generation).

It is easy to see that both the Daltonian and the Mendialian models provide answers to what-if-things-had-been different questions. In the Daltonian model, we can e.g. say what amount y of hydrogen (Y) would have combined with oxygen (X), had the amount of oxygen been x (namely multiples of the relevant atomic weights, in a fixed ratio). In the Mendelian model, we can say what ratio would have resulted if alleles in the paired gametes had been either all dominant or all recessive (namely a 4:0 ratio of either purple or white). But if all the models did was to provide this counterfactual information, there would be a concern that the models are actually redundant. E.g. had we combined y hydrogen molecules with oxygen, x oxygen molecules would have been required seems redundant when compared to the counterfactual mentioned above.[9] But such a view fails to see that what does the crucial explanatory work in the model is that, conditional on the truth of the postulates of the model, the law of constant and multiple proportions *have got to be* the way they are.

3.1 Explanatory asymmetry

Explanations are generally viewed as asymmetric relations: explanantia explain explananda, but not vice versa. For example, the length of a flagpole explains the length of the shadow it casts, but the length of the shadow does not explain the height of the flagpole.[10]

Explanatory asymmetry is easily accommodated by causal accounts of explanation: causes explain their effects, but effects don't explain their causes. In Woodward's counterfactual account of causal explanation, a variable X explains a variable Y if we can intervene on X so as to bring about a change in Y, but not vice versa. For example, we intervene on the height of the flagpole to change the length of the shadow, but we cannot, vice versa, intervene on the length of the shadow to bring about a change in the height of the flagpole.

In the recent trend in interest toward counterfactual dependence relations that are explanatory but not causal (see Reutlinger and Saatsi 2018), the problem of explanatory asymmetry arises anew. For example, in Bokulich (2011)'s account of model explanation, according to which models are explanatory of their targets by virtue of an isomorphism between the counterfactual structures in the model and the target (see fn 7), there is nothing that would ensure that the model explains the target, but not vice versa. In Bokulich's own example of the Bohr model, it is e.g. true that "had the spectral lines been different, the structures in the model would have been different".

---

[9] Needham (2004) has argued a very similar point, although he does not consider counterfactual information.
[10] Van Fraassen (1980) suggests that the directionality of explanation is pragmatic and that even for the famous flagpole there are situations in which the length of the shadow explains the height of the flagpole. For an important criticism see Kitcher and Salmon (1987). For more recent contributions on the problem of asymmetry see Jansson (2015), Khalifa et al. (forthcoming), and Kostić (2020).

But the spectral lines don't explain the structures in the model; it's the other way around (Schindler 2014).[11]

Structural necessitation offers a natural solution to the problem of asymmetry: it is the model that represents – and thereby explains – contingent regularities as necessities, and not vice versa. For example, the Bohr model of the atom explains the spectral lines of hydrogen, and not vice versa, because – as mentioned above – the structure of the Bohr model shows that the spectral lines *have got to be* discrete.

3.2 Necessities and contingencies

Of what kind are the necessities in explanations by structural necessitation? As we mentioned already briefly before, they are intended to be either *physical* or *mathematical* necessities. For example, when KT represents IGL as a necessity, it does so with the goal of showing that it is physically necessary for the gas pressure to rise when the volume of the gas container decreases. Importantly, the necessities in both cases are *conditional* on the truth of the postulated entities. In KT, for example, IGL becomes a physical necessity if it is true that gases consist of molecules in motion (which it is).

Although KT gets something fundamentally right about gases, it also makes many idealizing assumptions. For example, it pretends that the molecules in the container will never collide with each other—a patent falsity.[12] So in many cases, explanations by structural necessitation explain by representing empirical laws as physical necessities conditional on assumptions that may not be literally true. This shouldn't be particularly disturbing: at least since the work of Cartwright (1983) we know that even the most fundamental theories in physics are true of the world only with a lot of caveats and hedges.

There is a large recent literature that has explored viewing models *as fictions* (Godfrey-Smith 2009, Contessa 2010, Frigg 2010, Toon 2010, 2012, Levy 2015). For example, it is true in the Bohr model of the atom (but not in reality) that the atom that the electron orbits the nucleus just like it is true in the fiction that Sherlock Holmes lives on 221b Baker Street. On the other hand, it is false in the Bohr model that electrons crash into the nucleus when they do not occupy their stable orbits just as it would be false to say that Sherlock Holmes is a crooked gangster.

The literature on models as fictions has not focused much on model explanations (but see Bokulich (2009)). Still, it may be useful to think of structural necessitation as a representation of empirical laws as *physical necessities in a fiction.* Necessities in fictions should in principle be no more puzzling than actualities in fictions. 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 of course never actually fall to the floor, as Sherlock Holmes does not

---

[11] See Lange (2019) for critiquing several other counterfactual accounts of non-causal explanation for violating the asymmetry constraint.
[12] This is worked out forcefully by Strevens (2008).

exist. Analogously, in KT, 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.

Something analogous can be said of mathematical truths, which are also sometimes employed by explanations by structural necessitation. When Sherlock Holmes claims that 2 + 2 = 4, that's true and when he claims that 2 + 2 = 5, that's false. Sherlock Holmes of course never actually claims any of these things, since he doesn't exist, but we can still make sense of such propositions, just like we can make sense of the proposition that the offspring of homozygous blue and heterozygous brown eyed human beings has a 1 in 4 chance of having blue eyes even though the inheritance of eye colour is much more complicated than what is envisaged by the Mendelian model.

I have been assuming that the regularities explained by models like KT are contingencies. One may object, however, that on certain accounts the laws of nature are not contingencies but necessities themselves, or more correctly that a relation of 'contingent necessitation' holds between the universals described by the law (e.g. Armstrong 1983).

I have three comments about this objection. First, such a view does not sit particularly well with the fact that many laws of nature are in fact not exceptionless universal generalisations but rather ceteris paribus laws that hold only under certain condition and only within certain ranges (Reutlinger et al. 2017). For example, the ideal gas law that we discussed above, is correct only up until a certain temperature and pressure (above which the van der Waals equation is more accurate). Second, on standard accounts of laws of nature as relations of necessitation, the necessitation relation is still *contingent*. That is, on such accounts it would still be contingent that, in our world, a reduction in gas volume necessitates an increase in gas pressure. Hence, even on such accounts, explanations by structural necessitation could still be appealed to in order to explain why the relation has the form that it does have in our world. Third, and relatedly, many have objected that the notion of contingent necessities seems to be rather mysterious (and perhaps self-contradictory). I therefore consider it more fruitful to view empirical regularities and laws as simply contingent relations that can be represented as necessities by the kinds of explanations I'm discussing here.

4 The modal conception and structural necessitation

Broadly speaking, 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).[13] Although this makes the modal conception sound akin to the classical DN model of explanation, Salmon emphasizes that in contrast to the classical DN model

---

[13] Salmon (1984) 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 identifies the necessity of the explanandum's occurrence as the main source of explanation in the 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 in the subsequent decades—until recently (see Section 5). One reason for this lack of popularity may have to do with a criticism raised by Salmon himself against the modal conception.

The issue raised by Salmon concerns statistical and indeterministic events in science (Salmon 1984, 1998). Salmon's main complaint is that the modal conception cannot make sense of such events because they have only a certain probability of occurring and don't *have* to occur (note that the same criticism applies to the DN model). 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. Thus, modal accounts that suppose that *events* are explained by reference to a physically necessary law indeed don't look very plausible in the face of such examples.

These criticisms by Salmon are sound, but do not apply to explanations by structural necessitation. Whereas Salmon and the DN model (and many other accounts of scientific explanation for that matter) are concerned with the explanation of *events*, the explananda of explanations by necessitation are regularities, or more generally, *relations*. And that is in fact in accordance with the way that the mentioned scientific theories explain themselves.

Mendel's model predicts that the ration or probability distribution of flower colour in the aforementioned example will be 3:1. Thus, any given flower in F2 has a 1 in 4 probability of being white. This distribution is indeed necessitated by Mendel's model: provided the 'experiments' are conducted conscientiously, there cannot be any other distribution in F2. The model does not say anything about what the colour of any particular plant has got to be. We may of course give further causal explanations of why a certain colour was expressed by the genes in any *particular* plant (maybe contrary to what was expected). But we shouldn't dismiss an account of explanation because it does not accommodate such piecemeal causal explanations. Particularly not when the scientific model in question doesn't do so either.

5 Lange's explanations by constraint vs. structural necessitation

Recently, Lange (2013, 2017) has revived the modal conception with what he calls "explanation by constraint". Lange's preferred example of an explanation by constraint is inspired by remarks by R. Feynman and concerns the explanation of why various force laws conserve energy. Why do electrical and gravitational forces, despite their different range, different strength and different objects they apply to, all conserve energy? Either this is a mere coincidence, or the forces all conserve energy *because* they obey the principle of conservation of energy. That is, either the forces conserve energy because *each* of the force laws *separately* requires them to do so, for separate reasons, or those two kinds of interactions conserve energy *for the same reason*: because energy conservation requires them to do so (Lange 2017, 49-50).

Another kind of explanation by constraint are mathematical explanations. In Lange's widely discussed example of Mother failing to distribute 23 strawberries evenly among her 3 children because it is impossible to evenly divide 23 by 3. The necessities that do the explanatory work in explanations like these are mathematical necessities. Lange considers mathematical necessities to be 'stronger' than the physical or nomological necessities appealed to in explanations involving conservation laws. Conservation laws, in turn, are stronger than the necessities of 'ordinary' laws of nature (Lange 2017, 51). Lange also speaks of a metaphysical "pyramidal hierarchy" of necessities, whose levels can be determined on the basis of counterfactual conditionals. For example, we can ask whether energy conservation would still have held, if the forces had been different. Since the answer is yes, the necessities of explanations involving conservation laws are nomologically more 'stable' than the necessities at the lower levels.

Explanations by constraint are non-causal, according to Lange. E.g. there is nothing about the constraints set by the conservation laws that would 'cause' the force laws to obey them. Lange therefore also regards Woodward's counterfactual account of causation as unsuited to accommodate explanations by constraint. For example, the conservation laws do not identify circumstances under which the explanandum (i.e., the fact that different forces conserve energy) would have been different (that is, circumstances, in which conservation would *not* hold), nor do they identify *what* would have obtained in the actual explanandum's stead (48-9; 86-8). Explanations by constraints thus do not allow us to answer what-if-things-would-have-been-different questions, which are essential to Woodward's account of explanation.

Explanations by structural necessitation are similar to Lange's explanations by constraints in that they too set bounds on what is possible for the explanandum. E.g. KT rules out certain possible forms which IGL could have taken in world different from ours. There are several differences, though, between these two kinds of explanations.

First, explanations by structural necessitation postulate a *model mechanism* to achieve the representation of empirical regularities as necessities. Explanations by constraint, on the other hand, do not postulate anything. This becomes particularly clear in Lange's discussion of Einstein's distinction between constructive and principle theories. Constructive theories are usually characterized as theories that postulate causal mechanisms or models of the phenomena, whereas principle theories don't (Einstein 1993/1932, Balashov and Janssen 2003). Incidentally, the kinetic theory has been used (also by Einstein himself) as a standard example of a constructive theory. Lange, however, focuses squarely on principle theories and challenges the consensus that they do not explain by arguing that principle theories offer explanations by constraint.

Second, explanations by structural necessitation – contrary to Lange's explanations by constraint – are compatible with the explanans causing the explanandum. For example, it can be both true that KT represents IGL as a necessity and that the molecular constitution of gases causes their behaviour.

Third, explanations by structural necessitation reconcile views – contrary to Lange's – that have it that counterfactual dependencies are essential to explanation with views – such as Lange's

– that have it that explanations can be explanatory by virtue showing that the explanandum must be the way it is. Again, on explanations by structural necessitation, there is an counterfactual structure in the model that represents the counterfactual structure of the explanandum as a necessity.

Fourth, Craver and Povich (2017) have argued that one kind of Lange's explanation by constraint, namely distinctively mathematical explanations, fails to observe explanatory asymmetry. For example, Mother's *success* to distribute her strawberries evenly among her 3 children and the fact that 23 is not divisible by 3 explains why Mother doesn't have 23 strawberries. But this 'reversal' of the explanans and the explanandum (as compared to Mother's *failure*) shouldn't count as an explanation (Lange (2018) agrees with this). In so far as this objection is sound, my account of structural necessitation successfully addresses a central problem of explanation which Lange's modal account (or at least one subspecies of it) doesn't.[14]

Finally, with regards to Lange's pyramidal hierarchy, we can say that explanations by structural necessitation are probably best placed on rungs lower than conservation principles. The transfer of energy in KT, for example, must abide by them. On the other hand, explanations by structural necessitation should be placed on a higher rung of necessity than empirical regularities, especially when empirical regularities are themselves not to be conceived as necessities (as I suggested in Section 3.2).

The fact that explanation by necessity employ necessities that are not as "strong" as the necessities of conservation principles, however, does not mean that explanations by constraint would be less explanatory. For example, in contrast to the conservation laws, KT offers a *mechanism for why* IGL must take the form that it does take: the pressure of a gas must increase (not decrease) when the gas volume is reduced because the frequency of molecular collisions will be increased, given that the intermolecular distances will decrease (not increase) when the gas volume is reduced. Thus, I think we're justified to say not only that KT offers a deeper explanation than IGL (contra Woodward), but also that KT offers us a deeper explanation of IGL than energy conservation.[15]

We mentioned in Section 3 that mathematics plays a role in some explanations by structural necessitation, such as in Daltonian model of chemical compounds and the Mendelian model of inheritance. Recall that part of the explanation provided by the Daltonian model of chemical compounds is that "fractional" multiple proportions are impossible, because atoms are presumed

---

[14] In his reply to Craver and Povich (2017), Lange (2018) argues that matters of fact appealed to in explanations with the right directionality are "understood to be constitutive of the physical task or arrangement at issue", whereas that is not the case for explanations with the wrong directionality. Povich (forthcoming) finds Lange' reply wanting. An assessment of this dispute is beyond the scope of this paper.

[15] By "deeper" explanation I of course do not mean what Woodward (2003) and Hitchcock and Woodward (2003) mean by it, namely explanations that answer a *wider* range of what-if-things-had-been-different questions, which in their account is conditional on generalizations that exhibit greater degrees of invariance under interventions.

indivisible. Part of the explanation provided by the Mendelian model is that certain proportions of phenotypes (e.g. 2:2) are ruled out by the possible combinations of recessive and dominant alleles. Mathematics certainly plays a role in these explanations. But are the explanations distinctively mathematical?

For Lange, distinctively mathematical explanations are explanations that explain the world not by virtue of the causes or causal powers that they describe, but rather by virtue of the mathematics itself. For example, in the famous Mother example (mentioned above) one might view the fact that Mother has three children and the fact that she has 23 strawberries as "causes" for the explanandum, namely her failure to distribute the strawberries evenly (Lange demonstrates that prominent causal accounts indeed could give such a verdict). And yet, what explains the explanandum has got nothing to do with any causes or "the world's network of causal relations", as Lange (2013) puts it, but rather with the simple mathematical fact that the number 23 doesn't divide by the number 3 evenly. That's not only true, of course, for the particular example of Mother, but for "any collection of twenty-three things so divided", as Lange puts it. So the particular causes don't matter to the explanation. The same is true for another oft-mentioned example for a genuinely and distinctively mathematical explanation, namely why it's not possible to cross the seven bridges of Königsberg exactly once.

In contrast to Lange's distinctively mathematical explanations, the putative causes in explanations by structural necessitation *do* matter to the explanation. In the Daltonian model, if the postulated atoms had been presumed divisible, the law of multiple proportions would not have been explained: it would then have been an open possibility for elements to combine in non-integer multiples (which they do not). In the Mendelian model, likewise, if the alleles had been presumed mutable during meiosis, the laws of inheritance would not have been explained: it would have been possible for *other* ratios than the ones predicted by the model (e.g. 3:1 in F2) to obtain (which is not the case). Hence, although mathematics is certainly employed in these explanations, the characteristics of the postulated entities (or putative causes) – contrary to Lange's distinctively mathematical explanations – make all the difference to the success or failure to the explanation of the explanantia by structural necessitation.

6 Explanatory demarcation: beyond the dichotomy

Traditionally, it has been assumed that the explanantia in scientific explanations must be true, or at least approximately true (Hempel 1965). Yet more recently philosophers have eased up to the idea that the dichotomy of "either true, or no explanation" may be mistaken (Bokulich 2008, Strevens 2008, Batterman 2009, Bokulich 2009, 2011).[16] I think there are at least three categories of models with respect to explanation: not only are there (i) models that get the relevant causes roughly right

[16] There is also a sizable literature that explores the understanding scientists can gain about reality with highly idealized models (Elgin 2004, 2007, Rohwer and Rice 2013, de Regt 2015, Reutlinger et al. 2018, Doyle et al. 2019).

and that explain their targets and (ii) models that neither get the causes even roughly right nor explain, but also (iii) models that don't get the relevant causes right, but which still explain their targets. How does my account draw the lines between these three categories?

The demarcation of category (ii) and category (iii) has caused trouble for some accounts of model explanation, such as Bokulich's. As Belot and Jansson (2010) have argued, Bokulich's account cannot distinguish cases of genuine explanations from "cases where we intuitively have a prediction but do not have an explanation", such as the Ptolemaic, geocentric, model of the solar system (83–84; see below).[17] On my account, in contrast, the demarcation of category (ii) and category (iii) is straightforward: models of the second category don't explain because they don't manage to represent empirical regularities as necessities, whereas models of the third category do.

Let's consider one of the phenomena that the Ptolemaic model failed to explain: the 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 model, this fact is accommodated arbitrarily by *stipulating* that the centre of the epicycle on which the inner planets supposedly move would be fixed on a line connecting the sun and Earth. But this does not *explain* why the inner planets don't ever move far away from the sun; the model (by virtue of the stipulation) just represents the fact *that* they don't. There is nothing in the Ptolemaic model of planets moving on epicycles that would constrain this stipulation: the only reason for it derives from the phenomena themselves (which is why the model is often also described as ad hoc).

In contrast, in the Copernican model, the inner planets *cannot possibly* move away from the sun beyond a certain angle: the inner planets' orbits are encompassed by the Earth's orbit around the sun (see Figure 1), which explains *why* maximum elongation *has got to occur*. In contrast to the Ptolemaic model, this follows from the basic constituents and structure of the geometric model (in particular: the order of the planetary orbits and the sun).

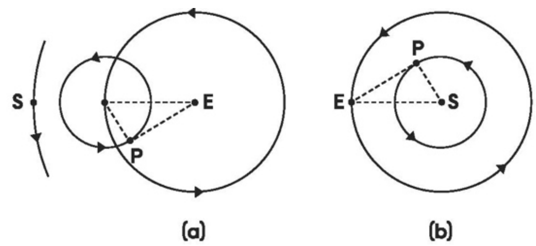

Figure 1: maximum elongation as accounted for in the Ptolemaic and the Copernican model (a and b, respectively). From Kuhn (1957).

[17] See Bokulich (2012) for a reply and Schindler (2014) for a critical assessment of this reply.

Now let's move on to category (iii) models. Consider for example the caloric theory of heat (CT), in which heat is conceived of as a *substance*, called caloric. Caloric particles are mutually repulsive, but attract (and are attracted by) matter particles (Figure 2).[18]

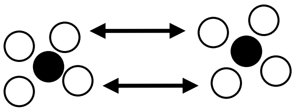

Figure 2: Caloric particles (white) repel each other but are attracted by matter particles (black).

Interestingly for our purposes, also on CT, IGL can be represented as a necessity by the mechanism postulated by CT and by means of the force laws employed (by e.g. Laplace). For example, when the volume of a gas container is reduced, the pressure has got to rise on CT, because the mutual repulsion of caloric particles will be stronger the closer the particles are to each other. Likewise, if the temperature of the gas is increased, there will have to be an increase in pressure, because an increase in temperature would result in an increase in the amount of caloric in the container. The increase of caloric particles, in turn, would increase the net amount of repulsion between the caloric particles, which on CT would be the cause of the pressure of gases.[19] So it seems that CT, despite its patent falsity, can explain IGL by way of structural necessitation just as KT can.

I'm happy to embrace this consequence of my account in the spirit of an *explanatory liberalism*, some form of which seems unavoidable. Again, it's generally accepted that KT explains IGL despite its many idealizations and literally false assumptions. Likewise, the Copernican model explains planetary motions even though it is literally speaking false that the planets move on circles around the sun (they move on ellipses). Mendelianism is in many ways highly idealized too: even the stereotypical example of eye colour in humans does not follow the simple scheme of dominant vs. recessive trait but in reality requires the study of several genes by molecular genetics. Finally, Dalton's atomism looks extremely simplified when compared to what we know about atoms and chemistry. In particular, it's false that atoms are indivisible. It would be highly revisionist to pertain that none of these models is explanatory because they don't correctly represent reality.

It must be emphasized that explanatory liberalism does have to result in explanatory anarchism. Again, on my account there are models that clearly fail to explain the phenomena, such as the Ptolemaic model: contrary to the models mentioned in the previous paragraph, it fails to represent the relevant empirical regularities as necessities. Furthermore, even in cases where entirely wrong models manage to represent empirical regularities as necessities (such as CT), there

[18] See e.g. Chang (2003) for more details.
[19] 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).

are good reasons to prefer the explanations provided by "truer" models. Those reasons have to do with *theoretical virtuous*. 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 (Howson 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. Thus, by virtue of its wider explanatory scope and greater fertility, KT is a better explanation of IGL than CT.

Explanatory scope and fertility are of course only two out of a number of theoretical virtues on the basis of which scientists can assess their theories and models (Kuhn 1977, Schindler 2018). Simplicity and mathematical tractability are also an important considerations, ruling out more fantastical 'models' with a much more demanding metaphysics than what's required by e.g. postulating particles and their interactions.

In sum, there are good reasons for preferring explanations provided by models that at least approximate the true causes to models that do not, even when both manage to represent empirical regularities as necessities.

7 Conclusion

In this paper I argued that some explanations in science are explanations by structural necessitation. These explanations explain by representing empirical regularities and the associated counterfactuals as necessities in a model. I have argued that explanations by structural necessitation are not subject to early criticisms of the modal conception of explanation and they are distinct from Lange's explanations by constraint.

References

Armstrong, David Malet. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

Balashov, Y and M Janssen. 2003. Presentism and relativity. *The British Journal for the Philosophy of Science*, **54** (2): 327-346.

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

Belot, Gordon and Lina 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, Alisa. 2008. Can Classical Structures Explain Quantum Phenomena? *The British Journal for the Philosophy of Science*, **59** (2): 217-235.

———. 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, Nancy. 1983. *How the laws of physics lie*. Oxford: Oxford University Press.
Chang, Hasok. 2003. Preservative realism and its discontents: Revisiting caloric. *Philosophy of science*, **70** (5): 902-912.
Colombo, Matteo, Stephan Hartmann, and Robert Van Iersel. 2015. Models, mechanisms, and coherence. *The British Journal for the Philosophy of Science*, **66** (1): 181-212.
Contessa, Gabriele. 2010. Scientific models and fictional objects. *Synthese*, **172** (2): 215-229.
Craver, Carl F and Mark Povich. 2017. The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A*, **63**: 31-38.
da Costa, Newton CA and Steven French. 2003. *Science and partial truth: A unitary approach to models and scientific reasoning*: Oxford University Press.
de Regt, Henk 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.
Doyle, Y, S Egan, N Graham, and K Khalifa. 2019. Non-factive understanding: A statement and a defense. *Journal of General Philosophy of Science*, **50** (3): 345-365.
Einstein, Albert. 1993/1932. *The world as I see it*, The World as I See It: Open Road Media.
Elgin, Catherine Z. 2004. True enough. *Philosophical issues*, **14** (1): 113-131.
———. 2007. Understanding and the facts. *Philosophical Studies*, **132** (1): 33-42.
Frigg, Roman. 2010. Models and fiction. *Synthese*, **172** (2): 251-268.
Godfrey-Smith, Peter. 2009. Models and fictions in science. *Philosophical Studies*, **143** (1): 101-116.
Grüne-Yanoff, Till. 2013. Appraising models nonrepresentationally. *Philosophy of Science*, **80** (5): 850-861.
Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
Hitchcock, Christopher and James Woodward. 2003. Explanatory generalizations, part II: Plumbing explanatory depth. *Nous*, **37** (2): 181-199.
Holton, Gerald James and Stephen G Brush. 2001. *Physics, the human adventure: From Copernicus to Einstein and beyond*. New Brunswick: Rutgers University Press.
Howson, Colin. 1976. *Method and appraisal in the physical sciences: The critical background to modern science, 1800-1905*. Cambridge: Cambridge University Press.
Jansson, Lina. 2015. Explanatory asymmetries: Laws of nature rehabilitated. *The Journal of Philosophy*, **112** (11): 577-599.
Khalifa, Kareem, Jared Millson, and Mark Risjord. forthcoming. Inference, explanation, and asymmetry. *Synthese*, **OnlineFirst**.
Kitcher, Philip and Wesley Salmon. 1987. Van Fraassen on explanation. *Journal of Philosophy*, **84** (6): 315-330.
Kostić, Daniel. 2020. General theory of topological explanations and explanatory asymmetry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **375** (1796): 20190321.
Kuhn, Thomas S. 1957. *The Copernican revolution: planetary astronomy in the development of western thought*. Harvard: Harvard University Press.
———. 1977. Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension*, Chicago: University of Chicago Press, 320-333.
Lange, Marc. 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.
———. 2018. A reply to Craver and Povich on the directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A*, **67**: 85-88. http://www.sciencedirect.com/science/article/pii/S0039368117302418.
———. 2019. Asymmetry as a challenge to counterfactual accounts of non-causal explanation. *Synthese*. https://doi.org/10.1007/s11229-019-02317-3.
Levy, Arnon. 2015. Modeling without models. *Philosophical studies*, **172** (3): 781-798.
Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science*, **67** (1): 1-25.
Mellor, David H. 1976. Probable explanation. *Australasian Journal of Philosophy*, **54** (3): 231-241.
Needham, Paul. 2004. Has Daltonian atomism provided chemistry with any explanations? *Philosophy of Science*, **71** (5): 1038-1047.
Povich, Mark. forthcoming. Modality and constitution in distinctively mathematical explanations. *European Journal for Philosophy of Science*.
Reutlinger, Alexander. 2016. Is there a monist theory of causal and noncausal explanations? The counterfactual theory of scientific explanation. *Philosophy of Science*, **83** (5): 733-745.
Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. 2018. Understanding (With) Toy Models. *British Journal for the Philosophy of Science*, **69** (4): 1069–1099.
Reutlinger, Alexander and Juha Saatsi. 2018. *Explanation beyond causation: philosophical perspectives on non-causal explanations*: Oxford University Press.
Reutlinger, Alexander, Gerhard Schurz, and Andreas Hüttemann. 2017. Ceteris paribus laws. *Stanford encyclopedia of philosophy*, edited by Edward N. Zatla, <https://plato.stanford.edu/archives/spr2017/entries/ceteris-paribus/>.
Rohwer, Yasha and Collin Rice. 2013. Hypothetical pattern idealization and explanatory models. *Philosophy of Science*, **80** (3): 334-355.
Saatsi, Juha and Mark Pexton. 2013. Reassessing Woodward's account of explanation: Regularities, counterfactuals, and noncausal explanations. *Philosophy of Science*, **80** (5): 613-624.
Salmon, Wesley. 1984. *Scientific Explanation and Causal Structure of the World*. Princeton: Princeton University Press.
———. 1998. *Causality and explanation*. New York: Oxford University Press.
Salmon, Wesley 1989. Four Decades of Explanation. *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, **13**: 3–219.
Schindler, Samuel. 2014. Explanatory fictions—for real? *Synthese*, **191** (8): 1741-1755.
———. 2018. *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: Cambridge University Press.
Strevens, Michael. 2008. *Depth: an account of scientific explanation*. Cambridge, Mass.: Harvard University Press.
Toon, Adam. 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.
von Wright, Georg Henrik. 1971. *Explanation and understanding*. Ithaca: Cornell University Press.
Votsis, Ioannis and Gerhard 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, James. 2003. *Making things happen: a theory of causal explanation*. Oxford: Oxford University Press.
———. 2014. Scientific Explanation. *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/.