Forthcoming in *Synthese*. Please cite the published version when available.

Truth, Understanding, and Normativity in Scientific Models

Lorenzo Spagnesi

Abstract: Scientific models often contain assumptions known not to be true. Despite being false representations, models provide us with a key understanding of phenomena. What is more, the falsehoods that figure in models are in many cases central to them, and there is no available alternative to their use. If falsehoods play such an irreplaceable role in our understanding of phenomena, it would seem that truth is not a key concern of scientific modeling. In this paper, I assess the prospects and challenges of reconciling truth and understanding in scientific modeling. More specifically, I review a thesis recently emerging in the literature, what I shall call the Derivation Thesis (DT), according to which we use models to derive true information. First, I examine different versions of the thesis and develop what I take to be its most promising formulation (what I call the generalized DT). Second, I discuss a serious challenge to the generalized DT. I consider a thought experiment in which an unreliable astrological model gives true explanations by fluke. This scenario challenges the idea that models can provide genuine understanding by generating truths. In response, I argue that genuine scientific models also fulfill a specific normative role that epistemically lucky models lack (what I call the normative generalized DT). I test this hypothesis by analysing how the Ideal Gas Law advances scientific understanding of real gases.

Keywords: truth, understanding, normativity, models, explanation

1 Introduction

A classic view has it that science aims at a true account of phenomena. Many accept that truth, spelled out e.g. as 'truth-likeness', 'approximate truth', 'verisimilitude', or 'empirical adequacy', is what science ultimately seeks to investigate.[1] While truth is a traditional aim of science, both epistemologists and philosophers of science have recently emphasized the centrality of understanding in science (see e.g. Friedman, 1974; Schurz & Lambert, 1994; Kvanvig, 2003; de Regt, 2017; Elgin, 2017; Potochnik, 2017; Rice, 2021). For these authors, science aims at the specific cognitive achievement that goes by the name of 'understanding'. We attempt to scientifically investigate nature not (or not only) to accumulate truths or increase the verisimilitude of our theories, but to *understand* complex phenomena.

Reconciling the aims of truth and understanding in science has proved difficult. A crucial test of such reconciliation is presented by an integral part of contemporary science, i.e. scientific

[1] This is particularly the case for scientific realism (roughly characterized as the view that scientific representations of entities and processes must be—in a sense to be specified—true); see e.g. Psillos (1999). However, anti-realists do not necessarily reject (some notion of) truth as a component of science. Kuhn, for example, considers accuracy to be the "most nearly decisive" criterion in science; see Kuhn (1977, p. 357); van Fraassen's constructive empiricism is not a complete departure from truth, since it requires that statements about observables be empirically adequate (see van Fraassen, 1980). Normative approaches to science, such as Longino's, retain empirical accuracy as a scientific value (albeit a pluralistic or non-epistemic value); see respectively Longino (1991) and Longino (1995).

modeling. Scientific models used in physics, chemistry, biology, economics, geology, etc. often contain idealizations, or assumptions that are known not to be true (e.g. Cartwright, 1983; Godfrey-Smith, 2009; Rohwer and Rice, 2013; Weisberg, 2007a). As noted in the literature, despite being false representations, models provide us with a key understanding of the world (e.g. Rice, 2016, 2021; Elgin, 2017, Potochnik, 2017). What is more, the falsehoods that figure in models are in many cases essential to them, and there is no (in practice or in principle) available alternative to their use (e.g. Elgin, 2017; Potochnik, 2017; Rice, 2018). If falsehoods play such an irreplaceable role in our scientific understanding of phenomena, it would seem that truth is not a central concern of scientific modeling and, arguably, of science as a whole.

In this paper, I evaluate the challenges and prospects of reconciling truth and understanding in scientific modeling. More specifically, I review and assess what I take to be a promising thesis that has emerged in the debate in recent years—what I call the *Derivation Thesis* (or DT). According to this thesis, we use models to derive true information. The latter (but not the former) represent the content of the understanding. To say that models provide understanding is to say that models enable (in ways that may be necessary or hardly replaceable) factive understanding, i.e. understanding containing true information.

In its simplest form, DT requires the derivation of true information as the content of understanding. This version has been most clearly presented by Bokulich (2016) and Lawler (2021).[2] As Bokulich puts it, a model is a "an adequate representation that can succeed in giving genuine physical insight into, and factive understanding of, a phenomenon of interest (Bokulich 2016, p. 274). For Lawler (2021), "felicitous legitimate falsehoods facilitate understanding a phenomena by enabling us to extract information about it" (p. 6876). More specific versions of DT impose additional constraints on the type of information required for understanding. For example, Alexandrova 2008 argues that "models are used as suggestions for developing causal hypotheses that can then be tested by an experiment"; p. 396). Rice—one of the leading defenders of DT— writes that "the goal of system-specific modeling is to provide accurate information about the counterfactual relevance and irrelevance of various contextually salient features within the model's target system" (2016, p. 88). In his 2018 paper, he suggests "that idealized modeling techniques that involve holistic distortion of real-world systems can provide true counterfactual information because many idealized model systems are known to approximate the patterns of behavior of real-world systems" (p. 2812). In 2021, he adds "the model is used to extract information about how the phenomenon counterfactually depends (or fails to depend) on various features of the system by investigating a merely possible system that shows how changing the actual features of the system result in changes in the phenomenon of interest" (p. 4108). In a similar vein, Pincock (2021) states that a model "(i) generates an explanatory generalization and (ii) each idealization in the derivation is partially true so that (iii) there is a wholly true derivation of that explanatory generalization that goes via these underlying truths" (p. 635). As I further show below, despite differences in their accounts, these authors agree that, while the idealized model is not itself part of the understanding, it is used to derive true information. DT—in its various forms—is an increasingly influential view in the literature.

While I think that DT is a compelling thesis (especially for realist-leaning philosophers of science), I also believe that it is, in its simple and specific versions, insufficient to reconcile the tension between truth and understanding in scientific modeling. I will review this thesis in two

[2] Lawler (2021) calls her thesis 'extraction view'. I use a different label, 'Derivation Thesis', to flag that this thesis has emerged in different forms elsewhere. Lawler deliberately leaves the type of true information to be derived unspecified. However, I argue below that this neutrality may be counterproductive.

steps. First, I develop a robust, generalized version of DT. While the simple DT is too permissive about the content of understanding, the specific versions are too narrow in their requirements. Drawing on contemporary accounts of understanding, I argue that the product of a model must be *explanatorily connected* true information about dependence relations. The discussion of what exactly should be derived from idealized models will clarify the requirements for understanding phenomena and offer a useful framework for further assessment.

Second, I discuss a serious challenge to the generalized DT. Inspired by Bird (2007), I consider a thought experiment in which an unreliable astrological model gives true explanations by fluke. This scenario challenges the idea that models provide understanding merely by generating truths (of any kind). I contend that scientific models have a specific use that allows the derivation of genuine understanding—one that epistemically lucky models lack. Contra existing proposals, I argue that this use is best understood as *normative* in character. Scientific models should be used as norms for deriving explanations.

On this reading, the word 'model' does not just stand for a "failure of exact correspondence" (Cartwright, 1983, p. 158), but also means 'standard' or 'term of comparison'.[3] I suggest that the normative function of scientific models lies in their systematic comparison with phenomena. That is, a modeler must establish an evaluative relation between the idealized model and target phenomena (usually captured by data-driven models). This comparison enables scientists to identify deviations and patterns that constitute the key information for deriving explanations. I test this hypothesis with respect to the multifarious ways in which the Ideal Gas Law, the Van der Waals Equation, and Maxwell's Equal Area Rule advance scientific understanding of real gases. Overall, I argue for what I call the normative generalized DT:

> *Normative generalized DT*: a genuine scientific model (i) allows us to derive true explanatory information about dependence relations and (ii) does so by acting as (or being used as) a norm for the systematic comparison of phenomena.

The structure of the paper is as follows. I begin by presenting the Derivation Thesis in its simple and specific versions as a promising approach to the reconciliation of truth and understanding (section 2.1). I then develop a generalized version of DT that is broad enough without being too permissive (section 2.2). In section 3, I challenge the generalized DT by questioning its sufficiency in providing genuine understanding of phenomena. In section 4, I first review various justification strategies for the information derived from models (4.1) and then develop my own normative take on models in relation to the central case study in the debate (the theory of gas; 4.2). In section 5, I explore the dynamic between the normativity of models and explanation. I conclude in section 6.

2 Reconciling truth and understanding in scientific modeling: prospects and challenges

2.1 The simple Derivation Thesis

There have been several attempts to reconcile truth and understanding in science. Many require that for understanding a phenomenon scientifically, the propositions I hold about it satisfy some truth condition (see e.g. Grimm, 2006; Mizrahi, 2012; Lawler, 2021; Pincock, 2021). It seems

[3] See e.g. one of the definitions in the *Oxford English Dictionary*: "A thing eminently worthy of imitation; a perfect exemplar *of* some excellence. Also: a representative specimen *of* some quality." On the history of the meaning of the word 'model' see Daston (2022).

intuitively the case that scientific understanding of a phenomenon contains true propositions about that phenomenon. For example, to properly understand human behavior, it seems necessary that the propositions I hold about it (drawn from neurology, psychology, sociology, etc.) are true. Conversely, if I find out that my understanding contains false propositions about a phenomenon, I should acknowledge that I have failed to understand that phenomenon. For example, if I use astrology to understand human behavior and I find out that astrology makes false claims, I should acknowledge that my understanding has been poor. The view that truth is a necessary condition for understanding is called *factivism*: understanding is factive, i.e. it contains true propositions about a phenomenon.

In this paper, I assume that factivism about understanding is on the right track, at least when it comes to scientific understanding.[4] But if that is the case, what role should we attribute to scientific modeling? Scientific models often contain idealizations that seem incompatible with factive understanding, yet they are an integral part of science. As noted by Lawler (2021), traditional attempts have usually hinged on undermining the role of falsehoods in understanding. In recent years, a new wave of reconciliation attempts, based on the recognition of the pivotal role of falsehoods, has emerged (see Alexandrova, 2008; Bokulich, 2016; Rice, 2016, 2018, 2021; Lawler, 2021; Pincock, 2021).

Despite their differences, all these attempts share a commitment to what I have called DT. As mentioned, DT comes in a simple and more specific versions. Let's start by focusing on the simple DT. According to the simple DT, idealized models are not themselves part of the content of the understanding but allow true information to be derived. An idealized model can allow the derivation of true predictions, true descriptions, explanatory generalizations, information about dependence relations, counterfactuals, etc. In a formula:

Simple DT: a genuine scientific model allows us to derive true information.

Consider, for example, the following case. I understand something about the behavior of gases using the Ideal Gas Law (hereafter IGL; a classic example in the debate):

$$PV = nRT$$

where P, V, and T are pressure, volume, and absolute temperature respectively, n is the mole of gas, and R is the ideal gas constant. IGL targets the so-called ideal gas: a highly idealized gas composed of dimensionless molecules that do not interact with each other. IGL is false of real gases but allows the derivation of accurate information about their behavior. The view thus sharply separates (1) idealized models from (2) the true propositions derived from them. For example, the latter correspond to the true information about the relation between P and V that is embodied by certain real gases under specific conditions. This information is derived using IGL, but it is not identical with it.[5] (1) and (2) play different roles: (1) plays the instrumental role of allowing true information to be derived; (2) represents the product of the models as tools, i.e. the epistemic 'gain' contained in our understanding.

Some clarification on the meaning of 'derivation' is needed.[6] One might argue that deriving true information from a model simply means isolating truths the model contains or directly

[4] For opposing non-factivist views that reject truth as a necessary condition for understanding see e.g. Elgin (2017), Potochnik (2018), and Doyle et al. (2019).
[5] On modeling and accuracy see also Hubert & Malfatti (2023).
[6] Thanks to an anonymous reviewer for pressing me on this point.

entails—for example, that the IGL, while false, nonetheless contains true counterfactual information, or that it entails truths when applied to real gases. But this minimal reading of 'derivation' is misguided in the present context. DT supporters take the falsity of idealized models at face value. While other approaches are possible,[7] I find this stance justified. First, if we could simply read off true information from idealized models, it would be unclear why idealized models require special treatment in the first place. After all, we can typically isolate true information within non-idealized models or theories. More importantly, it is doubtful that true information can be straightforwardly derived from idealized models. Consider IGL again: this idealized model is only true of a highly distorted ideal gas, and the kind of true information it provides is not trivially transferable to real gases (see e.g., Rice 2016, p. 90),[8] nor is it obvious that real gases fall within the scope of IGL (so that we can directly infer true information from it). Rather, real gases can be interpreted as approximating (or as deviating from) IGL under specific conditions (see e.g., Lawler 2021, p. 6868; Cartwright 1983).[9] As a result, 'derivation' must be understood more strongly in the context of DT: it involves clearly separating the idealized model from the true information about real phenomena that we derive from it.[10]

This view has several merits. First, it divides labor successfully (at least from a factivist perspective). It assigns distinct roles to idealized models and true information, without conflating the means of understanding with its content. Second, the view is flexible enough to allow for a multiplicity of ways of counting as true information and of gathering information (on which the view remains silent). Third, the view is able to accommodate the intuition that idealizations can play a crucial, if not indispensable, role in enabling understanding. It may be the case, for example, that certain derivation methods are irreplaceable to access particularly valuable information. For example, it may be argued that the Hardy-Weinberg Equilibrium (based on the false assumption of an infinite population) plays an indispensable role in understanding population genetics (see Strevens, 2016; Potochnik, 2018; Spagnesi, 2023).

I think the simple DT offers a promising approach to reconciling truth and understanding in scientific modeling. However, in its simple form, I doubt it is sufficiently spelled out for this purpose. One problem is that it is too permissive. For instance, it allows for models that generate isolated true propositions. A model that generates a singular correct prediction would be considered part of our scientific toolbox. However, this seems questionable—it is doubtful that a singular correct prediction contributes meaningfully to our understanding. While theories of understanding differ in significant, if not essential, ways, one recurring feature on which scholars agree is the connectedness (or coherence) of understanding. Unlike knowledge, understanding cannot be isolated. As Zagzebski puts it, "a person can know the individual

[7] For an interesting proposal see Kuorikoski and Ylikoski (2015). They argue that idealized models can inferentially yield true counterfactual information and thereby contribute to factive understanding, even if parts of the models are inaccurate. On this account, derivation (in a strong sense) is not *prima facie* required (pp. 3827–3828). Much, however, hinges, on how 'inferences' are understood here—whether as internal to the model (contained or directly entailed) or rather external to it. Since the authors adopt an extended cognition approach to models (coupled with an inferentialist account of representation) that blurs this distinction, the issue is not easily settled.

[8] It is plausible to think that the idealized model distortions also distort the information we obtain from such models (e.g. the counterfactual information about target phenomena). As Rice argues, models are typically wholes for which there is no obvious decomposition into separate elements (see Rice, 2019). As a result, the separation between idealized and non-idealized parts of the model faces serious mereological concerns.

[9] It is thus important to distinguish between idealized models (such as IGL) and models of actual phenomena (actual gases). See also Rescorla (2018) and Siegel and Craver (2024) for a similar distinction.

[10] This is true of Alexandrova (2008), Rice (2016), Lawler (2021), and Pincock (2021). Bokulich (2016) can also be interpreted this way (see e.g. p. 275), although her commitment to explanatory fictions makes the contours of her view less obviously compatible with the stronger reading of 'derivation'.

propositions that make up some body of knowledge without understanding them. Understanding involves seeing how the parts of that body of knowledge fit together" (2001, p. 244; see also Elgin 2017 and Kvanvig 2003). As I elaborate further below, such connectedness of understanding has been helpfully spelled out in terms of grasping (a system of) explanatory dependence relations (see e.g. Greco 2014; Grimm 2014; Kuorikoski and Ylikoski 2015). Roughly, we genuinely understand a phenomenon when we grasp how the relevant facts fit together, allowing us to account not just for the fact that it has occurred, but for why it has occurred.[11]

As we have seen, the simple DT does not necessarily meet this requirement for understanding. To be sure, it is legitimate to remain neutral on this issue and hold that, as long as models provide us with the right information, there is nothing more to say about DT.[12] However, I believe that only by addressing this concern can DT be a convincing strategy to reconcile truth and understanding. After all, merely deriving truths is not the same as deriving factive (and connected) understanding—it is thus important to precisify DT. Some interpreters recognize this need and propose stricter conditions. For example, Alexandrova (2008) requires "causal hypotheses"; Pincock (2021) emphasizes "explanatory generalizations"; Rice (2016, 2018, 2021) builds a framework based on the derivation of "counterfactual information"; and Potochnik (2018), though working within a non-factivist perspective, invokes information about "causal patterns".

While these accounts avoid the permissiveness of the simple DT, they tend to be too narrow in their requirements. There seems to be no compelling reason to accept one type of information over another—such as favoring causal generalizations while rejecting counterfactual information—yet what unifies these different kinds of information remains unclear. In the next paragraph, I provide a general account of DT, which I call generalized DT. This generalized version will serve as the basis for my critique.

2.2 The generalized Derivation Thesis

As we have seen in the previous section, connectedness appears central to traditional accounts of understanding. Some proposals highlight various types of information that seem to fit the general idea that understanding allows us to grasp the why of things (causal hypotheses, counterfactual information, explanatory generalizations, etc.). However, this remains somewhat patchy—how can we develop a more coherent framework from this idea?

Influential accounts of the epistemology of understanding, such as Greco's and Grimm's, connects understanding to the identification of explanatory dependence relations. Both emphasize that genuine understanding amounts to knowing (a system of) dependence relations holding between items (see e.g. Greco, 2014; Grimm, 2014). To use Siscoe's helpful formulation, they subscribe to a principle in the vicinity of *Dependence*:

> S understands why p if and only if there is a truth q on which p depends and S has a special cognitive relationship with the fact that a dependence relation holds between p and q. (Siscoe, 2022, p. 782)

[11] For a helpful discussion of different forms of understanding (understanding what, why, and how) see Hubert (2021).
[12] I take this to be Lawler's strategy (2021, pp. 6877-6878).

A model that provides only isolated information (such as a single prediction) does not fit this principle. Understanding p requires knowing a truth q from which p depends, and to which S has epistemic access. No such q is provided by such a model. However, the truths to which IGL leads us are related in a specific way. IGL can be used to derive a general explanation from which several particular predictions depend. What changes is not the content of understanding, but the relation holding between the elements of that content.

How should we spell out this relation? First, I take it to be a dependence relation holding between the elements of a *real target* (the phenomenon at stake). As such, it should not be confused with a purely conceptual relation holding between propositions. A model based on astrological assumptions might show a high degree of conceptual connectedness as a system of propositions, yet we would not say that it informs us of real phenomena. While it is often the case that idealized models present inferential systems of varying complexity from which one can derive truthful proportions, this requirement is neither necessary nor sufficient for a model to produce factive understanding. What matters, at least from a factivist point of view, is that the content of understanding contains dependence relations that are true of phenomena.

Second, the type of relation holding between p and q should be explanatory. In the literature, explanatory relations are sometimes equated with dependence relations. 'Dependence' has a broad meaning that includes different kinds of relations: causation, grounding, composition, substance, essence, supervenience, etc. Following Siscoe (2022), however, it is doubtful that all dependence relations are explanatory. For example, a supervenience relation is not always explanatory, whereas a causal relation usually is. While a thorough analysis of explanatory relations should be left to another occasion, I suggest conceiving of explanatory relations as a specific subset of dependence relations.[13] Two brief but important remarks are in order here. First, while understanding and explanation are conceptually distinct, it is plausible to think that explanation is a key component of understanding in science (Friedman, 1974; Lipton, 2004; Grimm, 2010; Khalifa, 2012; Kuorikoski & Ylikoski 2015; Siegel 2024).[14] Second, and relatedly, the emphasis on explanation is consistent with factivism about understanding. Most accounts of explanation require a truth or accuracy condition for something to count as an explanation (see de Regt & Gijsbers, 2017; and Rice, 2016, 2021). Therefore, the factivist requirement for the content of understanding should apply to the explanations derived from idealized models.

If this is correct, I conclude that a genuinely idealized model must not only allow us to derive true information but also ensure that this information is related in an explanatory way. In a formula:

> *Generalized DT*: a genuine scientific model allows us to derive true explanatory information about dependence relations.

This adjustment prevents the excessive permissiveness of the simple DT, while remaining broad enough to encompass a vast range of types of information about phenomena. Causal and counterfactual information, token explanations and generalizations, as well as other types of

[13] I therefore only partially agree with Strevens' (2008) simple view that there is no understanding without explanation, as it places exclusive emphasis on causal relevance. For a critique of Strevens' narrow conception of explanation see Pincock (2021).

[14] Some think that understanding and explanation should be equated (see Kuorikoski & Ylikoski 2015; Grimm 2010), others that explanation is only a condition of understanding (Siegel 2024). I remain neutral on this issue.

information about dependence relations, can all be included among the contents of understanding—generalized DT is thus a more robust version of DT.

3 The astrological model scenario

The generalized DT holds that scientific understanding contains true, explanatory information about dependence relations, and that idealized models have an instrumental role in obtaining such truths. At this juncture, one might think that the compatibility between understanding and truth in scientific modeling is vindicated. We use idealized models as tools to obtain true information. For example, IGL is not itself part of the content of factive understanding, but a tool that we use to obtain truthful information about the behavior of real gases (e.g. explanatory dependence relations between temperature, pressure, and volume in normal conditions). Despite its plausibility, I now wish to point out some shortcomings of the generalized DT.

Consider the following thought experiment (inspired by Bird, 2007). Let's imagine that:

(1) A community has formed its beliefs about a certain domain (human behavior) using the astrological model AM—an idealized model based on astrology.

(2) By fluke, AM gives true explanations about dependence relations in human behavior (correct explanations, say, about a sample of subjects).

This thought experiment is relevant to the present discussion because it provides a counterexample to the generalized DT. AM is an idealized model that allows a community to obtain true explanatory information about the domain under study. For example, AM happens to produce a plausible scientific explanation (informed by neurological, psychological, sociological, etc. considerations), and such an explanation q explains the true proposition p. (To make this scenario more concrete, one may think that the community associates astrological information with real factors and obtains by chance explanations that happen to be true.) This scenario poses a difficult challenge. For the product of AM and that of a genuine idealized model are exactly of the same kind: they establish the right dependence relation between a true *explanans* and a true *explanandum*. However, this product is obtained by luck, and it is questionable whether it can be considered a genuine model.[15]

Two shortcomings of AM can be highlighted here. First, from AM, the community cannot derive 'knowledge' in any proper sense of the term. It is generally accepted that knowledge is incompatible with epistemic luck (see Bird, 2007; Baumberger et al., 2017). To know something does not just require holding a true belief but also holding it in a justified and reliable manner. The plausible insight underpinning this conclusion is that if I accidently 'know' something, I do not know it. If I accidently know (perhaps via asking a random number generator) that 79 is gold's atomic number, I do not really know that 79 is gold's atomic number.

What about our present focus, i.e. understanding? Can I accidentally understand something? Things are less uncontroversial here, and different cases may require different treatments.[16] Let's stick with the specific case described in the thought experiment. The kind of luck involved in this experiment is known in the literature as 'intervening luck', i.e. luck intervening between our beliefs and the facts (see Pritchard, 2008; Baumberger et al., 2017). Our beliefs about human behavior generated through AM happen to match the facts. They happen to match the

[15] It is more difficult than one might think to identify what exactly makes astrology unscientific (see e.g. Hansson, 2021). This is not, however, to say that there is a debate on whether astrology is a science in the contemporary sense of the term.

[16] For a helpful overview see Baumberger et al. (2017); see also my footnote below on 'environmental luck'.

facts despite AM being a source of information that, more likely than not, would produce false explanations. I take this sense of luck to be incompatible with genuine factive understanding of a phenomenon.[17] Following Grimm (2006), I submit that AM does not give us genuine understanding of the phenomenon. This can be further illustrated with the following example from Pritchard (2009; as presented by Baumberger et al. 2017): imagine that my house has burned down due to faulty wiring causing a short circuit. When I arrive at the scene, I ask a partygoer dressed as a fireman what happened. The partygoer, by sheer luck, guesses that faulty wiring was the cause of the fire. It does not seem the case that the partygoer has genuine understanding of the given explanation, nor does it seem that I acquire understanding by relying on the fake fireman's true explanation.[18]

In other words, even if the dependence relation derived from AM happens to be the correct one, it originates from a flawed source. The issue is not with the derived information (which happens to be correct) but with the unreliable manner in which it is derived. Specifically, the derivation is largely disconnected from the phenomenon it supposed to refer to. If this is correct, merely deriving true explanatory information about dependence relations from a model is insufficient for the model to produce genuine understanding. As a result, the generalized DT fails to deliver what it promises—or better, additional justification is needed to establish the legitimacy of such derivations.

It should be noted that the analysis so far has addressed only the generalized DT. However, in generalizing this view, we may have overlooked key features of specific proposals that could have helped us respond to the AM challenge. Let's take a brief look at two such proposals: a non-factivist one (Potochnik's) and a factivist one (Rice's). Potochnik's view builds on Woodward's (2003) manipulability account of causation and requires that idealized models generate information about causal pattern dependencies. She remains neutral on the metaphysics of causation underpinning this account (p. 34), although we can read 'manipulability patterns' as informing us of objective causal dependence relations. In a factivist context, Rice (2016, 2018, 2021) develops a sophisticated account of how idealized models inform us about counterfactual dependence and independence among features of a system and a phenomenon. His account builds on how a phenomenon changes across modal space and how navigating this space via idealized models provides us with factive understanding (I further discuss Rice's view in section 4.1).

Do these specific accounts, which build on robust theories of dependence relations, help us solve the AM challenge? Not really. Despite the strengths and weaknesses of any given account of dependence relations, specifying the type of dependence does not eliminate the element of luck that can produce it. For example, AM might happen to generate accurate information about causal pattern dependencies or about counterfactuals between features of a system and a phenomenon. We could, for instance, infer such truths by exploiting a lucky correlation between manipulability patterns or counterfactual information and astrological data—without

[17] For a similar position, see e.g. Grimm (2006). Grimm argues against Kvanvig's thesis that understanding is not a species of knowledge since it is compatible with luck (Kvanvig 2003; see also Hills 2015 and Baumberger 2011). Rohwer (2014) develops a compelling case that understanding is compatible with luck if we integrate multiple sources of information rather than relying on a single unreliable one. The case I discuss, however, is limited to the classic scenario of single-source information.

[18] For an analogous example see also Grimm (2006, p. 525). However, compare this case to one in which I ask a real fireman, despite being surrounded by partygoers dressed as firemen. This type of luck, known as 'environmental luck', arises not from the relation between beliefs and the facts but from the surrounding conditions in which a piece of information is obtained. This might be a stronger case for understanding not being as species of knowledge. For a discussion see Baumberger et al. (2011).

genuinely grasping the relevant relations at stake. But this result is unsurprising: the details of the preferred account of dependence relations postulated in DT cannot dictate how we generate the information about these relations—no matter how sophisticated the account). More broadly, I submit that any product-based approach to evaluating model derivation can be undermined by formulating thought experiments similar to the one discussed.[19] In our case, we can reformulate the experiment to yield not only true information but also any kind of explanatorily connected true information.

If specifying the type of dependence relation is not a promising strategy for addressing the AM challenge, the discussion so far suggests an alternative: what seems to matter in avoiding these counterexamples is not what we derive from idealized models but how we derive it. Genuine understanding requires not just the right kind of dependence relation but proper *justification*—or, as I will argue, a proper grounding in the relation between models and phenomena. While some previous accounts have touched on this idea, I aim to make it central. In the next section, I explore ways to justify the information derived from idealized models.

4. The normativity of scientific models

4.1 Justification strategies

In the previous section, we saw that even sophisticated accounts of DT face a serious challenge. While they allow us to derive true explanatory information about dependence relations, they fall short of producing genuine understanding of phenomena. In other words, scenarios like the one described challenge the very instrumental value of idealized models. Even if idealized models provide the right content for understanding, they are arguably unfit as derivation tools. A tool that barely gets the job done is hardly a tool at all—certainly not a good one. Thus, idealized models complying with the generalized sophisticated DT view may still fail to be effective tools for deriving understanding.

How can we address this challenge? Similarly to debates about knowledge, I suggest that what is lacking here is a justification for the information derived from idealized models. The merely enabling role of idealized models is insufficient to provide genuine understanding. This insufficiency arises not merely because their instrumentality is not always successful, but more fundamentally because the mere derivation of the right content for understanding undermines the acquisition of genuine understanding. As a result, the information derived by idealized models needs to be justified. We need to show that the derivation of information is a good one. This justification concerns *how* the content of understanding is derived.

Even at this preliminary stage, it is clear that such a justification is not easy to achieve. In similar contexts, one can appeal to the 'correctness' of what enables the derivation of accurate information. For example, a true theory from which true consequences derived seems a strong basis for justification. However, as we will discuss later, this option is blocked by the very nature of idealized models: by definition, they contain falsehoods. A second and crucial difficulty is that, since we are attempting to justify the products of tools, we cannot rely on these very products, or further products, to achieve this justification. I take this to follow from what we have established at the end of the previous section: any product-based (or instrumental) justification of model derivation can be undermined by scenarios involving luck. Thus, appealing to the products of models to justify the information we derive from them risks

[19] On this point see also Bird (2007, p. 72).

either (i) circularity, or, if it involves appealing to further products, (ii) an infinite regress of justification. As we will see, even accounts that have raised the justification question fail to provide a convincing answer. Let me now briefly explore three possible options for justifying the information that we derive from idealized models.[20]

(1) *Internalism*. A first option to justify the information that we derive from models is to appeal to some of their internal feature(s), e.g. their being interconnected or their appeal to some basic beliefs. However, this is not a promising strategy in this context for at least two reasons. First, because models vary greatly in their content and construction rules. The existence of a wide variety of propositional and non-propositional models makes it difficult to identify an invariant feature or set of features that may justify them.[21] Additionally, an explanation of how true explanatory information is derived from a model that appeals *only* to internal features of the model would fail to account how a model relates to phenomena and the explanatory account thereof. To be sure, it is conceptually possible to derive an explanation of a phenomenon from a model without there being a direct relation between the model and the phenomenon (as AM shows). However, it is plausible to assume that a good way of deriving information about a phenomenon is by relating to it or being responsive to it (as many assume; see e.g. Rice 2016; Lawler 2021; Pincock 2021). The next two options explore variations on this approach.

(2) *Veritism*. A standard way of justifying a method (in this case, a derivation method) is to appeal to the truth of its premises. If the method is based on true premises, it is also warranted that the claims we extract from it are true. The latter is clearly a non-starter in the present context since scientific models are characterized by containing falsehoods. However, Rice has developed an interesting proposal that broadly aligns with this strategy. According to him, it is not necessary for models to accurately represent phenomena, in part or in total, to produce valuable information; other links can suffice. The one he emphasizes is that the idealized models and the real target belong to the same 'universality class': as he puts it, "this entails that the model and the real-world system(s) will display similar patterns of macroscale behavior even if the model drastically distorts the entities, relationships, and processes of its target system(s)" (2016, p. 94). Belonging to the same universality class is thus the 'true' premise required to justify the derivation of information.

Rice's proposal is ingenious, but it struggles to address our justification question. Much, of course, depends on what one means by 'universality class'. If interpreted metaphysically, as referring to something akin to a 'universal', this view would be difficult to defend, as it would introduce demanding ontological commitments. However, Rice defends a weaker and more flexible conception of universality classes. He defines them as sets of similar behaviors (see e.g. Rice 2018, p. 2000). But such a minimal definition poses problems as well. As Pincock (2021, p. 632) notes, the real target and the idealization may behave similarly for objectively different reasons. The only way to determine whether they belong to the same universality class is if the model generates true counterfactual information. If this is correct, this strategy does not offer a justification that is independent of the model output. Indeed, it is circularly based on the correctness of the output that it should justify. In a similar vein, Bokulich (2016, p. 274) argues that some fictions qualify as genuine because they are "credentialed"—i.e., they are accepted by the scientific community as productive of factive understanding. But this form of

[20] I focus here on accounts that are compatible with factivism.

[21] For example, as we have seen above, an astrological system of beliefs can show a great degree of interconnectedness, provide true explanations, and yet fail to be a scientific model. On the challenges of a unified account of models see Weisberg (2007b).

justification is, again, instrumental, i.e. based on the model's output. As we have seen, such a justification cannot successfully justify the information derived from idealized models.

(3) *Non-representationalism*. The difficulty with the previous strategy suggests that the problem of justifying models may stem from the representational function that is attributed to them. Philosophers such as Strevens and Pincock have defended non-representationalist justifications of idealized models. On Strevens' influential account, idealizations are not meant to represent facts. Rather, they signal that some factor is (causally) irrelevant to the purpose of explanation. For example, that the molecules of a gas have no interaction is not the representation of a feature of the gas but rather flags that such interaction is causally irrelevant to the target of explanation (see Strevens, 2008).[22] Pincock (2021) objects that Strevens' solution is limited to a narrow conception of causation and explanation. He proposes the following solution: "each idealization in the derivation is partially true so that … there is a wholly true derivation of that explanatory generalization that goes via these underlying truths" (p. 635).[23] In other words, an idealized model indicates a "commitment" to derive an explanation through true claims entailed by it. A true derivation may not be available at present (after all, it is only a commitment—we may not be able, for example, to derive a true explanation of genotypic frequency without assuming an infinite population) and may rely on various (not necessarily causal) explanatory strategies.

I believe that non-representationalism is on the right track and that there is much to be learned from the approaches of Strevens and Pincock. Their views account for the relation between models and phenomena (*contra* internalism) while avoiding the pitfalls of veritism. However, I submit that neither Strevens' nor Pincock's accounts manage to justify the information derived from models. On closer inspection, both views center their justification strategy around the derivation of further truths—causally irrelevant facts or partial truths entailed by the model. The key idea seems to be that models can (at least potentially) be turned into a set of factivist claims. But this is to *double down* on the instrumental value of models. Models exist to signal such a (potential) reduction—they are ultimately dispensable as tools for obtaining the relevant information. However, as I have suggested, this strategy poses a problem for DT. What we aim to justify is precisely that models provide the right information, yet appealing to the derivation of further derived information leads to an infinite regress of justification. If this is correct, it remains unclear how this approach can provide genuine understanding rather than merely enabling the right content of understanding.

4.2 Scientific models *qua* norms

I think there is a better way to justify the information derived from idealized models. As we have seen, it remains unclear how such a justification can be obtained by appealing to the internal or representational features of models. A more promising route seems to lie in the non-representational relation between models and phenomena. What still needs to be established is

[22] One may object that Strevens (2008) defends a representational account of models, given that, in his view, the latter inform us about difference-making features of a real-world target. However, my classification of Strevens' view within the non-representationalist camp is intended to highlight the special role he assigns to idealizations. For instance, in his discussion of the Hardy-Weinberg Equilibrium (HWE), Strevens (2019) emphasizes that it is an idealized concept that holds only at the limit and merely signals the irrelevance of population size. What is representational, then, is not HWE itself but rather the non-idealized models through which scientists derive HWE. Strictly speaking, HWE serves a different function: it flags a truth about causal irrelevance. If it represents anything, it would represent an infinite population—precisely what Strevens seeks to avoid (2019, p. 1725).

[23] This solution is based on Yablo's notion of 'partial truth': "p is partially true when for some r, p entails r, r is true and the subject matter of r is part of the subject matter of p" (Pincock, 2021, p. 635).

the specific non-representational justification of the information derived from models. I suggest that such a justification is *normative* in character. A model is to be used, as the word suggests, as a 'norm' or 'standard'. A 'norm' or 'standard' does not tell what reality is—it does not directly describe or explain real features of the world. It rather gives us a term of comparison for inquiry.[24] More specifically, I contend that a scientific model gives us a term of comparison for the explanation of phenomena. Explanations are derived by the systematic comparison of phenomena with the model.[25] Let me clarify some key features of the normative use of models first.

As some things are used as standards for other things, models are compared with the phenomena for which an explanatory account is sought.[26] But what is 'comparison' in this context? I take comparison to be an evaluative relation established by the modeler between the idealized model and target phenomena (often described by data-driven models). The comparison can concern any property, feature or aspect of the phenomena under study and is usually performed by calculating the deviation between the model and actual measurements. For example, as I will expand below, we can compare IGL with data-driven models of actual gases with respect to the property of compressibility. This relation allows to identify deviations between the model and the target phenomena. Such deviations (and patterns thereof) are key to deriving explanatory information.

Two remarks are important here. First, the acts of comparing the idealized model with the target phenomena must be as systematic and complete as possible. Given the inevitable limitations of experimentation, systematic comparison is best described as a 'commitment'.[27] Second, while the model will differ from the target phenomena, the magnitude of such deviations can vary. In some case the deviations will be minimal. In other cases, they will be substantive. What matters is that the explanations derived from the idealized model are based on these deviations. As we will see below, the deviations between IGL and data-driven models can be used to derive the true and explanatory information that some actual gases, under certain conditions, embody the dependence relation described by IGL; and second, that specific factors—such as molecular size and intermolecular forces—account for the divergence between IGL and real gases.

I contend that the normativity of idealized models allows us to justify the true explanatory information derived from them. For by using a model as a norm, we do not just derive true information but do so in a way that it is responsive to phenomena. This responsiveness consists in the *act* of systematically comparing a model with phenomena.[28] I take the act of systematic

[24] For a similar emphasis on 'comparison' (but within a different framework) see Frigg (2010, pp. 263-264).
[25] In elaborating this approach, I take my cue from several sources: first, Kant's account of regulative ideas and their prescriptive function in inquiry is a historical precursor of this view (for an analysis see Spagnesi, 2023); second, my approach resonates with Woody's functional approach to explanation (2015), which has a similar emphasis on explanation as an activity and on the normativity of models; third, and more generally, I understand my normative approach to be in tune with the recent 'zetetic turn' in epistemology, i.e. a turn from doxastic attitudes to the processes of inquiry that produce them.
[26] Think of Polykleitos' *Doryphoros* as the 'canon' of the proportions of the human body in sculpture. Daston (2022) reconstructs the original meaning of rule as 'model' or 'canon'.
[27] I borrow this term from Pincock (2021).
[28] As a result, the normative use of models cannot be undermined by a product-based objection. Of course, one can isolate understanding how as a product of reflection on the activity of systematic comparison from understanding how as the activity of comparison. However, I take the former to presuppose the latter since the activity makes possible the reflection on it.

comparison to be a reliable procedure—one that is incompatible with luck.[29] For to say that a systematic comparison is accidental is a contradiction in terms. Of course, this is not to say that any comparison of models with phenomena will bring about the right explanations. Nor is it to deny that the right explanations may be accidentally derived even once a comparison has been performed. But it is to commit oneself to the systematic investigation of the relation between a model and phenomena—thus ruling out lucky derivations of true information (at least in ideal conditions).[30]

One concern that needs to be addressed before developing this view further is that, for scientists to compare a model with its real target, they need to know what is true of the phenomenon. However, scientific modeling often takes place in contexts where such information is lacking—or if it were available, there would be no reason to generate a model in the first place.[31] On closer inspection, however, this concern is misguided. While it is true that models are, in many cases, our primary means of accessing relevant information (as the generalized DT asserts), this does not mean that all information about a phenomenon is derived solely from idealized or highly idealized models. Take the cases of IGL or the Hardy-Weinberg Equilibrium. We can compare IGL with empirical data, observed phenomena, and, more generally, non-idealized, data-driven models of real gases (i.e., gases measured in experimental conditions). Similarly, the Hardy-Weinberg Equilibrium can be compared with datasets on finite biological populations. Such comparisons allow us to extract additional relevant information, which, as we saw above, may be accessible only through the use of idealized models.[32]

I have suggested the derivation of factive understanding is justified if it happens via the systematic comparison with phenomena. Consider our thought experiment. In using AM, no comparison of the model with phenomena is performed.[33] What about IGL? In IGL, the derivation of explanations is obtained by the systematic comparison of the model with phenomena and the resulting identification of patterns—or so I contend. To back up my hypothesis, I need to delve into more details of the model.

It is often said (rather vaguely) that, under certain conditions, some real gases follow IGL closely.[34] But as we know, real gases are made up of (i) molecules that take up volume and (ii) molecules that interact with each other through forces. How exactly (and to what extent) can we say that ideal gases approximate the behavior of real gases? A good indicator to understand the relation between IGL and real gases is the compressibility or compressor factor (Z). Z is

[29] For a classic formulation of reliabilism see Goldman (1979). From the latter, however, I only take the minimal claim the justification is given by a process (i.e. an activity), not by a belief or set of beliefs. I remain neutral on the other aspects of the theory and on the best way to understand it (e.g. whether 'virtue reliabilism' fares better).
[30] Of course, no justification strategy is without objections. For example, it is questionable whether this solution is able to address evil demon scenarios (without further specification). My modest contention is that it addresses the problem raised in the thought experiment above with respect to scientific models that produce true information. I leave a more complete defense to another occasion.
[31] I thank an anonymous reviewer for pressing me on this point.
[32] On the relation between models and experiments see e.g. Morgan (2006).
[33] This is postulated in the scenario: we derive a true explanation from AM by luck—just as we guess a true account of Comanche history, or identify the true cause of a fire by asking a fake fireman. Of course, outside the scenario, we could systematically compare AM with phenomena, as we can do with any idealized model. In that case, AM would be used as a norm for comparison. And if such a comparison were to yield true explanations (which it plausibly doesn't), AM would also count as a successful model. Thanks to an anomynous reviewer for raising this worry.
[34] This is especially true of monoatomic gases, such as helium and neon. However, also polyatomic gases, such as oxygen or carbon dioxide, follow the ideal gas to a certain extent under certain conditions (see discussion below).

the ratio of the molar volume of a gas to the molar volume of an ideal gas at the same temperature and pressure:

$$Z = PV / nRT$$

For a gas with ideal behavior, Z=1. As shown in *Figure 1* (left graph), the compression factor changes significantly for different gases at different pressures.

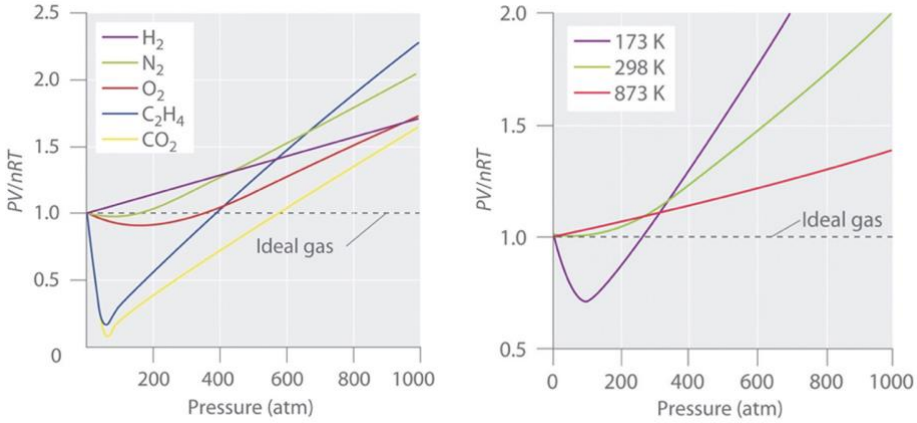

Fig. 1 Two graphs of the compressibility factor (Z) versus pressure (P). The graph on the left shows Z vs. P for different gases at 272 K. The graph on the right shows Z vs. P for nitrogen at three different temperatures. Copyright: LibreTexts shared under CC BY-NC-SA 4.0

This graph shows the compression factor Z for different pressures at 273 K for various gases (H_2, N_2, O_2, C_2H_4, and CO_2). Since for the ideal gas Z=1 over the whole range of pressures, it can be seen that only a few real gases (H_2, O_2, N_2) approximate the behavior of the ideal gas and only over a limited pressure range (about <400). At low pressures, most gases have Z<1, while at higher pressures, Z>1. These deviations are caused by the larger molar volumes of real gases than those predicted by IGL. IGL assumes dimensionless molecules moving in empty space. Real gases take up more space simply because their molecules do have volume. The effect is more pronounced at higher pressures, where gases are more compressed and the available empty space decreases.

To calculate the intermolecular forces for a given gas, one needs to look at the compression factor of a gas at different temperatures. Figure 1 (right graph) shows the compression factor of nitrogen (N_2) at various temperatures. The compression factor of N_2 at T=298 K and 873 K below 200 bar is relatively close to the expected ideal gas value (Z=1). However, Z values deviate from 1 at lower temperatures (173 K). At this temperature, Z is significantly <1. This deviation is due to the attractive forces acting between the molecules of real gases. If the pressure is constant, this results in a reduced volume, which explains Z<1. The effect is more pronounced at lower temperatures due to the relatively lower kinetic energy of the molecules.

In order to assess both deviations (volume and intermolecular forces), the behavior of the ideal gas described by IGL acts as a term of comparison for the investigation of the phenomena—specifically, the molar volume of the ideal gas is the *denominator* of the compression factor Z. The latter has the normative use of being the standard against which empirical results must be compared systematically, i.e. with respect to all possible phenomena of interest (this is represented by the continuity of the lines in the graphs above). The systematic comparison of empirical results to IGL, in turn, allows the identification of deviations. In this case, the

systematic comparison with phenomena (either real or potential) allows the identification of the following deviations (among others):

(a) The deviation of Z of H_2, O_2, N_2 from IGL is relatively small over a limited pressure range (0-400 bar).

(b) The deviation of Z of N_2 at T=298K and 873K from IGL is relatively small over a limited pressure range (0-200 bar);

(c) The deviation of Z of H_2, O_2, N_2 from IGL is relatively significant at high pressure (>400 bar);

(d) The deviation of Z of C_2H_4 and CO_2 from IGL is relatively significant over a large pressure range;

(e) The deviation of Z of N_2 at T=298K and 873K from IGL is relatively significant at high pressure (>200 bar);

(f) The deviation of Z of N_2 at T=173 from IGL is relatively significant over a large pressure range.

Note that none of the claims above are 'representational'. Each claim records a difference between IGL and the behavior of real gases. In the 'close' cases (points a and b), the derivation of representational claims about dependence relations is straightforward. Under certain conditions and for some gases, the deviation of the behavior of a gas from that of an ideal gas is so negligible that one can use IGL to derive accurate explanations. For example, one can say that, over the limited pressure range 0-400 bar, P and V are inversely proportional if n and T are held constant in the real gases H_2, O_2, N_2. Under these conditions, this dependence relation is both accurate and explanatory. As such, it represents an apt content of factive understanding (and should therefore be distinguished from the idealized dependence relation between P and V in an ideal gas).

In the 'far' cases (c through f), large deviations from IGL indicate the need for further explanations. In a well-known model such as IGL, further explanations are already available. We know that these deviations can be explained by the effects of the volume and interaction of the molecules of real gases. For example, we can say that, at pressure higher than 400 bar, the relation between P and T in the real gases H_2, O_2, N_2 deviates from IGL. This is due to the relatively larger molar volumes of these gases at high pressure, which arise from intermolecular forces and the finite volumes of their molecules. The latter is an accurate and explanatory dependence relation (a causal explanation) that figures in the content of the understanding. Importantly, we can develop new models of gases that incorporate this information, leading to more accurate results or information about different dependence relations (see section 5). Less known cases require complex interpretation and various experimentation techniques to obtain comparable results.

Note also that although models typically involve 'close' cases, this is not a necessary condition for models to produce explanations. If compared with the behavior of real cases, a model may simply give rise to deviations (of any magnitude) that call for explanations. There is no qualitative difference between what I have called 'close' and 'far' cases. In both cases, a comparison is made between the model and the real case(s) to identify deviations from which explanations can be derived (through interpretation). Finally, it should be emphasized that, while the systematic comparison of a model with phenomena may shed light on token deviations (some of which may be especially informative, such as critical points), it will

typically identify *patterns* of deviations.[35] Such patterns will facilitate the elaboration of explanations with varying degrees of generality and specificity.

This latter point is crucial for understanding how my proposal differs from traditional representationalist accounts. It is possible to argue that even according to representational accounts of idealized models, a comparison is made—presumably a systematic comparison with phenomena to assess accuracy. From this point of view, my normative account might not seem fundamentally different. However, in comparisons with norms, what matters is not the similarity relation between the model and the phenomenon, but rather the deviations between the two. These deviations are open to interpretation and serve as the source of information about the phenomenon. For this reason, the comparison is best described as an ongoing commitment. One might further object that, if this is the case, then anything could be used as a model (since no similarity between model and phenomenon is required). But I contend that this is actually a strength of the proposed account. As the variety of existing models shows, anything can potentially serve as a model—what matters is that it functions as a norm for deriving true explanatory information. This does not lead to excessive permissiveness, because while anything can be a model, only those that generate accurate and explanatory information through their normative role belong in our scientific toolbox.

On the proposed account, we derive genuine understanding from idealized not only if the resulting information is true and explanatory, but also if such information is normatively obtained. I formulate my normative version of generalized DT more precisely as follows:

> *Normative generalized DT*: a genuine scientific model (i) allows us to derive true explanatory information about dependence relations; and (ii) it does so while acting as (or being used as) a norm for the systematic comparison of phenomena.

In other words, a model contributes to understanding not only through the content it yields, but also through the way that content is derived. This justification ensures that our understanding is not accidental, but responsive to the phenomena themselves. It succeeds because it does not rest on the model's output, but on the specific role the model plays *vis-à-vis* the phenomena.

5 The dynamic between the normative function of models and explanations

It may sound strange that scientific models play a normative role. After all, scientific models are part and parcel of scientific inquiry, which is a descriptive enterprise. Moreover, it seems that a model really does contribute to our understanding of phenomena. Nothing I have said so far contradicts these claims. There is little doubt that science is a descriptive endeavor and that models contributes to our understanding of phenomena. Such a contribution, however, is neither purely instrumental nor representational. On the proposed view, models are to be used as norms for the derivation of genuine understanding of phenomena.

The normativity of models is to be understood as directly responsible for the accurate description of phenomena. This key aspect of scientific modeling can be best appreciated by highlighting the dynamic between modeling and explanation. As a norm, an idealized model is a reference point from which we can derive explanations—it is never fully realized or representationally successful. For example, there is simply no real gas of which IGL is true or accurate *simpliciter*, nor is there any actual population described by the Hardy-Weinberg

[35] For an emphasis on the role of patterns in modeling see Potochnik (2018).

Equilibrium. On the contrary, the comparison of the model with real cases results in the identification of deviations that are conducive to explanation: (1) when an idealized model provides relatively accurate results over a certain range, such results can always be made more precise with respect to different dimensions of investigation (i.e. different ranges, questions, or aims); (2) when real cases deviate significantly from the idealized model, more experimentation and interpretation are required. In both cases, the 'space' left between the model and the phenomena calls for new explanations.

This dynamic can be illustrated by further exploring the theory of gases. There are multiple ways to model the behavior of gases to give more accurate results than those of IGL. For example, the van der Waals Equation (VWE) accurately accounts for the effects of the volume and the interaction of molecules:

$$(P + an2/V2)(V–nb) = nRT$$

Compared to IGL, the equation includes (i) a correction for the pressure value (an2/V2), which accounts for the decrease in pressure due to intermolecular attraction; and (ii) a correction for the volume (V–nb), which gives a more accurate measure of the empty space available for gas molecules; a and b are values specific to each gas and depend on temperature and volume. As it is often noted, the increase in accuracy determines a decrease in the generality of the equation. VWE is much more specific than IGL, although it still provides a framework for obtaining accurate results over ranges not covered by IGL (see e.g. Epstein 1937). VWE is thus a model that covers some of the explanatory 'space' left by IGL by producing more accurate predictions and explanations.

While VWE provides more accurate results, it also gives rise to deviations when compared to phenomena. Real gases behave noticeably differently from the predictions of VWE below the critical temperature (in this case, the end point of the phase equilibrium where liquid and vapor coexist). As shown in Figure 2, for T<Tc, P and V are related non-monotonically.

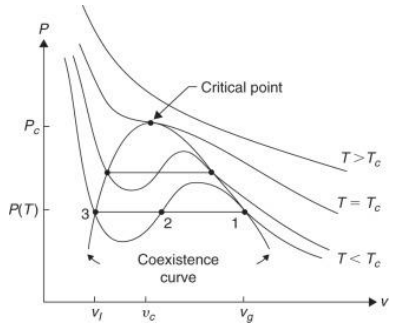

Fig. 2 The isotherms of a van der Waals system. Reprinted from Pathria & Beale 2011, Copyright (2011), with permission from Elsevier

Over a certain range of molar volume v we find an 'unphysical' region (where $(\partial P/\partial V)>0$), which must be corrected by using Maxwell's so-called Equal Area Rule (EAR). The latter allows the construction of an isotherm that signals the transition from the gaseous state with molar volume Vg to the liquid state with molar volume Vl at constant pressure P(T). Between Vl and Vg (coexistence curve), the system is in a partly liquid, partly gaseous state with an infinite compression factor (since $\Delta V \neq 0$ but $\Delta P=0$). This is a singular behavior that cannot be explained using VWE and requires EAR as an additional modeling technique.

The transition from IGL to VWE and the implementation of EAR illustrate the dynamic between modeling and explaining. Each model allows us to derive explanations but also leaves space for other models to take on additional explanatory functions. In such a way, the normative use of models is fully integrated with the descriptive purpose of science. Different models can also take on different explanatory functions with respect to the multifarious aims of science. While the example I have offered remains within the coordinates of a given explanatory framework, there is nothing to prevent models from pursuing orthogonal or hardly reconcilable explanatory frameworks. For this reason, I take my view to be compatible (at least in principle) not only with the internal dynamic of single theoretical framework, but also with a multiplicity of different frameworks, each of which pursues different dimensions, questions, or aims of science.[36]

One may object that this dynamic merely reflects the uncontroversial point that any explanation derived from an idealized model is always partial: by answering some questions, it also raises others.[37] While I acknowledge that this dynamic describes an undeniable feature of scientific inquiry, the proposed view clarifies the motivation behind it. Consider again a model like AM. Even though this model provides an accurate explanation of phenomena, it does not itself initiate further questions about them. If we investigate AM's explanations, it is not because of AM but because we independently raise new questions about that phenomenon. Representational accounts of models also struggle to explain this dynamic. While any representational model inevitably leaves some aspects of a phenomenon unexplained, it is unclear why a representationally accurate model of a given dependence relation should invite further questions about it—we already have the correct account of the object of inquiry.

This view, like non-representational accounts, highlights that for any model, there is always more to explore. However, because this signaling is based on identifying deviations, it motivates a particularly rich and varied research program. The deviations between the model and the investigated object outline a space of inquiry that is, in principle, never fully closed and can be explored from different perspectives and goals. For example, we might refine certain details of a dependence relation, question its grounds, or analyze it from different standpoints. By contrast, in the non-representationalist accounts discussed earlier, the 'more to explore' is exhausted once the model is (potentially) replaced by further truths derived from it. In short, the proposed view clarifies the non-obvious motivation behind an otherwise unquestionable fact about scientific inquiry, that models always raise further questions about phenomena.

6 Conclusion

In this paper, I have argued that a genuine scientific model (i) allows us to derive true explanatory information about dependence relations; and (ii) it does so while acting as (or being used as) a norm for the systematic comparison of phenomena. This is not to say that models in science always fulfil (i) and (ii). Some models may just be 'black boxes' for the derivation of information—they may produce the right kind of information, but we do not know how such information is derived. Not knowing how true information is derived is different from knowing that it is the product of sheer luck (as in the AM case). As a result, such models may have a legitimate place in our scientific toolbox. Conversely, there may be genuine models that are systematically compared to phenomena but do not produce explanations. These models are

[36] I leave the development of this point to another occasion.
[37] I am grateful to an anonymous reviewer for raising this objection.

simply not the successful ones. They may be poorly constructed (or superseded by better ones), or they may be waiting to be fine-tuned to deliver true explanations. Of course, to fully explore the complications and challenges of the proposed approach to modeling would require further conceptual development and analysis of case studies beyond the tried-and-true example of the theory of gas. My aim in this paper was to explore some of the challenges for the Derivation Thesis and to outline a possible way of reconciling truth and understanding in scientific modeling.

Acknowledgements

Earlier versions of this paper were presented in Lisbon, Trier, Düsseldorf, London, and Oviedo. I am grateful to the audiences, especially to Kristina Engelhard, David Hommen, Andreas Hüttemann, Roman Frigg, Pietro Gori, Alba Padrós, Lorenzo Sala, Lorenzo Sartori, Gerhard Schurz, Giulia Terzian, and Oscar Westerblad, for their stimulating questions on various aspects of this work. The three referees for this journal provided insightful and constructive comments that helped improve the paper. Any errors are my own.

References

Alexandrova, A. (2008). Making models count. *Philosophy of Science*, 75(3), 383–404.

Baumberger C., C. Beisbart, and G. Brun (2017). "What is understanding? An overview of recent debates in epistemology and philosophy of science." In S. Grimm, C. Baumberger, and S. Ammon (Eds.), *Explaining Understanding. New Perspectives from Epistemology and Philosophy of Science* (pp. 1–34). Routledge.

Baumberger, C. (2011). Types of understanding: Their nature and their relation to knowledge. *Conceptus*, 40, 67–88.

Bird, A. (2007). What is scientific progress? *Noûs*, 41(1), 64–89.

Bokulich, A. (2009). "Explanatory fictions." In M. Suárez (Ed.), *Fictions in science: Philosophical essays on modeling and idealization* (pp. 91–109). Routledge.

Bokulich, A. (2016). Fiction as a vehicle of truth: Moving beyond the ontic conception. *The Monist*, 99, 260–279.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Daston, L. (2022). *Rules: A short of what we live by*. Princeton University Press.

de Regt H.W. (2017). *Understanding scientific understanding*. Oxford University Press.

de Regt, H.W., and V. Gijsbers (2017). How false theories can yield genuine understanding. In S. Grimm, C. Baumberger, and S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 50–75). Routledge.

Doyle Y., S. Egan, N. Graham, and K. Khalifa (2019). Non-factive understanding: A statement and defense. *Journal for General Philosophy of Science*, 50, 345–365.

Elgin, C. Z. (2017). *True enough*. MIT Press.

Epstein, P.S. (1937). *Textbook of thermodynamics*. John Wiley & Sons.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, 71, 5–19.

Frigg, R. (2010). Models and fiction. *Synthese*, 172(2), 251–268

Godfrey-Smith, P. (2009) Models and fictions in science. *Philosophical Studies*, 143, 101–116.

Goldman, A.I. (1979). What Is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge: New studies in epistemology* (pp. 1–25). Reidel.

Greco, J. (2014). Episteme: knowledge and understanding. In K. Timpe and C. Boyd (Eds.), *Virtues and their vices*. Oxford University Press.

Grimm, S. R. (2006). Is understanding a species of knowledge? *British Journal for the Philosophy of Science*, 57, 515–35.

Grimm, S. R. (2010). The goal of explanation. *Studies in the History and Philosophy of Science*, 41, 337–344.

Grimm, S. (2014). Understanding as knowledge of causes. In A. Fairweather (Ed.), *Virtue epistemology naturalized: Bridges between virtue epistemology and philosophy of science* (pp. 329–345). Springer.

Hansson, S. O. (2021). Science and pseudo-science. *Stanford Encyclopedia of Philosophy.*

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

Hills, A. (2015). Understanding why. *Noûs*, 50(4), 661–688.

Hubert, M. (2021). Understanding physics: 'What?', 'why?', and 'how?'. *European Journal for Philosophy of Science*, 11(3), 1–36.

Hubert, M. and Malfatti, F. (2023). Towards ideal understanding. *Ergo*, 10(22), 578–611.

Kaplan, D., and C. Craver (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78(4), 601–27.

Khalifa K. (2012). Inaugurating understanding or repackaging explanation? *Philosophy of Science*, 79, 15–37.

Kuorikoski, J., and Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, 192(12), 3817–3837.

Kvanvig, J. L. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge University Press.

Kuhn, T.S. (1977). *The essential tension. Selected studies in scientific tradition and change*. University of Chicago Press.

Lawler, I. (2021). Scientific understanding and felicitous legitimate falsehoods. *Synthese*, 198, 6859–6887.

Lipton, P. (2004). *Inference to the best explanation*. Routledge.

Longino, H.E. (1991). Multiplying subjects and the diffusion of power. *The Journal of Philosophy*, 88(11), 666–674.

Longino H.E. (1995). Gender, politics, and the theoretical virtues. *Synthese*, 104(3), 383–397.

Mizrahi, M. (2012). Idealizations and scientific understanding. *Philosophical Studies*, 160(2), 237–252.

Morgan, M.S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology*, 12(2), 317–329.

Pathria, P.K. and Beale P. D. (2011). *Statistical mechanics*. Butterworth-Heinemann.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Pincock, C. (2021). A defense of truth as a necessary condition on scientific explanation. *Erkenntnis*, 88(2), 621–640.

Pritchard, D. (2008). Knowing the answer, understanding and epistemic value. *Grazer Philosophische Studien,* 77, 325–39.

Pritchard, D. (2009). Knowledge, understanding and epistemic value. *Royal Institute of Philosophy Supplements*, 64, 19–43.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. Routledge.

Rescorla, M. (2018). An interventionist approach to psychological explanation. *Synthese*, 195(5),1909–1940.

Rice, C. (2016). Factive scientific understanding without accurate representation. *Biology & Philosophy*, 31(1), 81–102.

Rice, C. (2018). Idealized models, holistic distortions, and universality. *Synthese*, 195(6), 2795–2819.

Rice, C. (2019). Models don't decompose that way: A holistic view of idealized models. *The British Journal for the Philosophy of Science*, 70(1), 179–208.

Rice, C. (2021). Understanding realism. *Synthese*, 198, 4097–4121.

Rohwer, Y. (2014). Lucky understanding without knowledge. *Synthese*, 191, 945–959.

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

Schurz, G. and K. Lampert (1994). Outline of a theory of scientific understanding. *Synthese*, 101, 65–120.

Siegel, G. (2024). Scientific understanding as narrative intelligibility. *Philosophical Studies*, 181(10), 2843–2866.

Siegel, G., and Craver, C.F. (2024). Phenomenological laws and mechanistic explanations. *Philosophy of Science*, 91(1), 132–150.

Siscoe, R.W. (2022). Grounding, understanding, and explanation. *Pacific Philosophical Quarterly*, 103, 791–815.

Spagnesi, L. (2023). Regulative idealization: A Kantian approach to idealized models. *Studies in History and Philosophy of Science*, 99, 1–9.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Strevens, M. (2016). How idealizations provide understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 37–49). Routledge.

Strevens, M. (2019). The structure of asymptotic idealization. *Synthese*, 196, 1713–1731.

van Fraassen, B. (1980). *The scientific image*. Clarendon Press.

Weisberg, M. (2007a). Who is a modeler? *The British Journal for the Philosophy of Science*, 58(2), 207–233.

Weisberg M. (2007b). Three kinds of idealization. *The Journal of Philosophy*, 104(12), 639–659.

Woodward, J. (2003). *Making things happen*. Oxford University Press.

Woody, A.I. (2015). Re-orienting discussions of scientific explanation: A functional perspective. *Studies in History and Philosophy of Science*, 52, 79–87.

Zagzebski, L. (2001). Recovering understanding. In M. Steup (Ed.), *Knowledge, truth, and duty: essays on epistemic justification, responsibility, and virtue* (pp. 235–251). Oxford University Press.