SCIENTIFIC UNDERSTANDING AND FELICITOUS LEGITIMATE FALSEHOODS

Insa Lawler
University of North Carolina at Greensboro
irlawler@uncg.edu, www.insalawler.com

This is a pre-print of an article published in *Synthese*. The final authenticated version is available online at: https://doi.org/10.1007/s11229-019-02495-0

Abstract

Science is replete with falsehoods that epistemically facilitate understanding *by virtue of being the very falsehoods they are*. In view of this puzzling fact, some have relaxed the truth requirement on understanding. I offer a *factive* view of understanding (i.e., the extraction view) that fully accommodates the puzzling fact in four steps: (i) I argue that the question how these falsehoods are related to the phenomenon to be understood and the question how they figure into the content of understanding it are independent. (ii) I argue that the falsehoods do not figure into the understanding's content by being elements of its periphery or core. (iii) Drawing lessons from case studies, I argue that the falsehoods merely enable understanding. When working with such falsehoods, only the truths we extract from them are elements of the content of our understanding. (iv) I argue that the extraction view is compatible with the thesis that falsehoods can have an *epistemic* value by virtue of being the very falsehoods they are.

Understanding plays a central role in human inquiry, particularly in scientific inquiry. Science aims at understanding phenomena, such as understanding language acquisition.[1] Without loss of generality, understanding requires an epistemic commitment to a systematic account of the phenomenon in question. Paradigmatic instances of systematic accounts of phenomena are explanations or analyses thereof. One might say that a subject understands a phenomenon only if they grasp an explanation or analysis of it. In what follows, I call such analyses or explanations of the phenomenon to be understood the *content* of understanding. The nature of the epistemic commitment involved in understanding is contested. To accommodate the variety of understanding theories, I remain neutral regarding whether it is a form of knowledge or whether it is an epistemic commitment *sui generis*, whether abilities are constitutive of understanding, etc. (For an overview of the debate see, e.g., Gordon 2017.) In what follows, I am

[1] Understanding language acquisition is an example of so-called *objectual* understanding. Scientists also want to understand *why* things are the case. My arguments apply to both forms of understanding. I remain neutral regarding whether understanding why is a form of objectual understanding or *vice versa* (see, e.g., Grimm 2011; Khalifa 2013; Baumberger and Brun 2017).

exclusively concerned with the *content* of understanding. In a nutshell, I argue that the use of prominent falsehoods in science is compatible with a factive view about understanding. I call this view the *extraction view*. A key claim is that these falsehoods can play an *epistemic enabling role* in the process of obtaining understanding but are not elements of the explanations or analyses that constitute the content of understanding.

1 A PUZZLING FACT

Prima facie, understanding is factive: its content can only contain true propositions (or at least approximations to the truth). If I seem to understand a phenomenon but it turns out that my account of the phenomenon contains false propositions, one would say that it looked as if I understood it, but I actually did not. Yet, science is replete with falsehoods that are considered to foster or facilitate understanding rather than prevent it. Examples of such falsehoods range from smoothed curves to heavily idealized models. Due to their utility, Catherine Elgin calls such falsehoods 'felicitous' (2007, p. 39). The falsehoods in question are not just felicitous but they are also considered to be *legitimate*. It is not bad scientific practice to employ such falsehoods. Quite to the contrary, Elgin argues that such falsehoods facilitate understanding not despite but *by virtue* of being the very falsehoods they are: "[...] their divergence from truth or representational accuracy fosters their *epistemic* functioning" (2017, p. 1, my italics). Below, this claim is fleshed out in more detail. Taking it for granted for the moment, we are faced with the following puzzling fact:

A PUZZLING FACT There are legitimate falsehoods that epistemically facilitate understanding of phenomena by virtue of being the very falsehoods they are.

In view of this fact, Elgin maintains that there are just three main options when it comes to the factivity of understanding (2017, p. 15):

> Our predicament is this: We can retain the truth requirement and construe science either [(i)] as cognitively defective or [(ii)] as noncognitive, or [(iii)] we can reject, revise, or relax the truth requirement and remain cognitivists about and devotees of science.

Prima facie, options (i) and (ii) or rejecting *any* truth requirement seem unacceptable. So-called *quasi-factivism* about understanding opts for *relaxing* the truth requirement. The content of one's understanding can contain falsehoods if they concern matters that are *peripheral* to understanding the phenomenon in question (e.g., Kvanvig 2003; Mizrahi 2012) or if most propositions of the content are true (Rice 2016). I call this the *weak* non-factive view of understanding. According to the *strong* non-factive view of understanding (short: non-factivism), we need to *revise* the truth requirement. The content of one's understanding can contain falsehoods that are *central* to understanding the phenomenon (e.g., Elgin 2007, 2017; de Regt 2015, 2017; Potochnik 2017, Rancourt 2017).

There are three main attempts to dissolve the predicament and to resist a non-factive view. According to the non-difference-maker view, felicitous falsehoods turn out to be *harmless*; they point to factors that do not make a significant (explanatory) difference to the phenomenon understood (e.g., Strevens 2008, ch. 8, 2017; Khalifa 2017, ch. 6.3.2; cf. also M. Elgin and Sober 2002). An alternative attempt is to re-consider the *objects* of understanding; what is factively understood is what the respective falsehoods mean or their relation to the phenomenon (e.g., Greco 2014). Yet another attempt is to argue that the value of the falsehoods is *non-epistemic*; the falsehoods are just convenient for practical purposes such as calculation (e.g., Sullivan and Khalifa 2018). All these attempts rebut or substantially weaken the puzzling fact about felicitous falsehoods. According to the non-difference-maker view, the falsehoods do not substantially contribute to the content of understanding, i.e., the systematic account of the phenomenon of interest (which is arguably concerned with the difference-makers). Shifting the object of understanding amounts to denying that the phenomena themselves can be understood based on the falsehoods. And arguing that the value of falsehoods is non-epistemic amounts to denying that understanding is *epistemically* facilitated by the falsehoods.

In this paper, I offer an alternative proposal to dissolve the predicament and to resist a non-factive view of understanding in view of felicitous falsehoods that *fully* accommodates the puzzling fact.[2] I dub this view the *extraction* view. I proceed as follows: In section 2, I characterize felicitous legitimate falsehoods in more detail. A *first* key step in my proposal is to loosen the link between two often conflated things (sect. 3): I argue that the question how these falsehoods are related to the phenomena of interest (relation question) and the question how the falsehoods figure into the content of understanding the phenomena (content question) are independent of each other. One can figure into the content of an epistemic achievement by *enabling* that achievement or by being an *element* of its content. According to quasi-factivism, felicitous legitimate falsehoods figure into the content of understanding by being elements of its *periphery*. According to non-factivism, they figure into it by being elements of its *core*. My *second* key step is to argue that facts from science speak against both positions (sect. 4, 5). In section 6, I turn to my *third* key step. I draw lessons from works by Anna Alexandrova (2008), Christopher Pincock (2014), Alisa Bokulich (2016), and Collin Rice (2016, 2018) to argue that felicitous legitimate falsehoods merely *enable* understanding but are not an element of its content. When working with such

[2] I do not address any other arguments against factivism, such as *arguments from history* against factivism, according to which scientists gained understanding based on theories which turned out to be false (see, e.g., de Regt 2015, 2017). For a discussion, see, e.g., Khalifa 2017, ch. 6.2.

falsehoods only the truths we extract from them are elements of the content of our understanding. Felicitous legitimate falsehoods are not more nor less than fruitful tools to obtain this information. In section 7, I turn the *final step* in my proposal. I argue that the extraction view is perfectly compatible with the thesis that such falsehoods have an *epistemic* value by virtue of being the very falsehoods they are. For instance, they might provide us with epistemic access to relevant information for understanding the phenomenon. So, according to my view, the *process of obtaining understanding* might be rife with falsehoods, but not the understanding obtained.

2 FELICITOUS LEGITIMATE FALSEHOODS

A variety of falsehoods figure into scientific inquiry.[3] Some of these happen by accident. Scientists make mistakes, such as accidentally reporting false numbers, miscalculating, or confusing substances in a chemical experiment. Other falsehoods are what Michael Strevens calls 'deliberate falsehoods' (2017, p. 37). Scientists deliberately employ such falsehoods and do not believe them to be true; they are known to be (slightly or utterly) false. Among these falsehoods are *illegitimate* ones, such as fabricated results or distorted statistical values. Elgin provides a plausible explanation why these are illegitimate. She argues that the propositions that figure into a scientific account of something do not only need to have a tether to the facts; a "tie to evidence is crucial. For evidence supplies the grounding that underwrites epistemic standing" (2017, p. 132). Claims without evidence are not trustworthy. We need acceptance for understanding and evidence for acceptance. Fabricated results are thus not legitimate falsehoods. Yet, there are also a vast number of deliberate falsehoods that are considered to be *legitimate*. Some are the result of *streamlining* or *pruning* collected data, like a smoothed curve of data points gained by ignoring outliers. Some are *generalizations* derived from idealized models, such as the ideal gas law, which implies that the volume and temperature of gases are directly proportional to each other when their pressure is held constant. Some are the *stipulations* that figure into idealized models, such as the stipulation that gas molecules exert no long range forces on each other or that the number of particles is arbitrarily large.[4] In some cases, the divergence from truth is negligible, such as describing nearly round objects as round. We need not worry about such a divergence. Scientists typically have to trade off precision against other values

3 I do not use the term 'idealization' as an umbrella term for these falsehoods because it might be too narrow. Some scientific falsehoods are claimed to be *fictions* rather than idealizations (see, e.g., Godfrey-Smith 2009; Frigg 2010; Bokulich 2012; Sugden 2013). And idealizations are often distinguished from *abstractions* (see, e.g., Jones 2005) or *approximations* (see, e.g., Norton 2012). For more on idealizations see, e.g., Elliott-Graves and Weisberg 2014; Potochnik 2017.

4 Another interesting case are falsehoods involved in *hypothetical reasoning*, such as scientific thought experiments or how-possibly models. I do not consider their characteristics in this paper; but see, e.g., Dray 1957; Grüne-Yanoff 2013; Rohwer and Rice 2013; Sugden 2013; Bokulich 2014; van Riel 2015; Verreault-Julien 2019.

like generality. Approximately correct descriptions can also be preferred when they simplify calculation, reasoning, etc.[5] I assume that approximately correct descriptions are compatible with a factive view of understanding. In other cases, the divergence from truth provides a more *distorting* picture of the phenomenon of interest, such as assuming that the number of particles in a fluid is arbitrarily large. For a factive view of understanding, such utterly false falsehoods seem to be especially worrisome. These are the falsehoods I focus on in what follows.

What makes these described distorting falsehoods *legitimate*? One main reason is that they are often *empirically and robustly successful*:

> [T]hese equations [i.e., the Navier-Stokes equations] are extremely *safe* [...]. By 'safety,' here we mean that one can employ these equations quite successfully in engineering context such as pipeline construction, airplane and ship design, and so on. (Batterman and Rice 2014, p. 357, my italics)

> The results [i.e., a match of the model's prediction with the collected data] showed that the observed patterns of foraging behavior were *best predicted* by the optimality model [...]. (Rice 2018, p. 2804, my italics)

> [...] because of systematic, *successful prediction*. [...] the behavior of gases usually agrees with the predictions of the Ideal Gas Law within certain ranges of temperature and pressures. (Mizrahi 2012, p. 251, my italics).

> [...] the behavior of a pendulum over a long timescale can be *derived* to an arbitrary degree of accuracy using one of several singular perturbation techniques. (Wayne 2011, p. 836, my italics)

> This remarkably simple model does, in fact, *recover* a host of features observed in real fluids at these large continuum scales. [...] The result will accurately reproduce many of the macroscopic features of fluid flow. (Batterman and Rice 2014, p. 358, my italics)

So, the empirical success contributes to the required tie to evidence. Because they can provide quite accurate results or recover the features of interest, the idealized equations and the false stipulations involved in models ought to be taken seriously.[6] But importantly, distorting falsehoods need not be empirically successful to be considered legitimate. The *potentiality* of empirical success can be enough. Scientists typically employ distorting falsehoods based on educated guesses about which aspects of a phenomenon they can fruitfully distort. When the model or equation that utilizes

[5] It goes without saying that the notion 'approximately' is context-sensitive. One has to specify a proximity range for the value of interest. A value is significantly different if it is not equivalent to the value of interest or within its specified proximity.

[6] Models are typically not identified as falsehoods for at least two reasons. On the one hand, many models are not something that could be true or false. On the other hand, models typically also involve *accurate* stipulations about their target phenomena.

them cannot accommodate the measured data, it is typically improved or replaced (for the purposes in question). Take the case of the optical Glauber model. It predicts that the collision of two nuclei results in the formation of an ellipsoid based on the idealization that nuclei are perfect spheres of energy. Measurements showed that these collisions result not only in ellipsoids, but other shapes as well. In the light of this observation, the Monte Carlo Glauber model was invented, which does not employ the perfect sphere idealization. So, falsehoods it can be legitimate to propose and theorize with are only *potentially* empirically successful. However, the (potential) empirical success is not *sufficient* for a falsehood to be legitimate.[7] We need what I call an appropriate *tie to the phenomenon*. How is the falsehood related to the phenomenon? Why are we justified in exploring the phenomenon of interest with the falsehood (or something that involves it)?

From now on, I reserve the term 'legitimate falsehoods' for falsehoods that are *potentially* empirically successful and have an appropriate tie to the phenomenon. And I reserve the term 'felicitous legitimate falsehoods' for legitimate falsehoods that are in fact successful. It is these falsehoods that are at center of the debate about factivism about understanding.

But what is an 'appropriate tie' to the phenomenon and what does it imply for the factivity of understanding? I address this issue in what follows.

3 PHENOMENA, FALSEHOODS, UNDERSTANDING: TWO SEPARATE QUESTIONS

The project of establishing an appropriate tie to the phenomenon of interest is crucial. I call one key question it addresses the *relation question*:

RELATION QUESTION How do the falsehoods (or the models that involve them) *relate to* the phenomenon in question?

Answering this question involves answering a couple of closely related questions. One question is whether and, if applicable, how falsehoods or the idealized models that involve them *represent* the phenomenon in question. Many accounts address this question. Some argue that models are not mundane representations but rather akin to *fiction* (e.g., Godfrey-Smith 2009; Frigg 2010; Toon 2012). Others emphasize the *role of the model users*. For instance, Mauricio Suárez argues for an *inferential* account, according to which the model must allow its users to draw inferences regarding the phenomenon (2002). Tarja Knuuttila argues that models are essentially *epistemic artefacts*, which are used to answer the users' questions (2011). And Elgin (e.g., 2007, 2017, ch. 9-12) as well as Roman Frigg and James Nguyen (e.g., 2018) argue that the relation between the falsehoods (or the models) and the phenomenon of interest is that of *representation-as* (following Hughes 1997). For instance, gases are represented

[7] I thank an anonymous reviewer for emphasizing this point.

as ideal gases for a particular purpose. Among other things, for X to represent Y as something, X must *exemplify* features of Y; X functions "[...] as a symbol that makes reference to some of the properties, patterns, or relations it instantiates" (Elgin 2017, p. 184). For instance, a sample card colored with 'grayish green' ink exemplifies that color. Similarly, falsehoods (or models) can exemplify, say, structural properties, such as the direct proportional relation between the volume and temperature of gases when their pressure is held constant.

Another closely related (and sometimes conflated) question is by virtue of what we are *justified* in using such falsehoods. Elay Shech describes this question as follows, focusing on so-called essential idealizations (EI) (2013, p. 1178, my italics):

> We need an account of how our abstract and essentially idealized scientific representations correspond to the concrete systems observed in the world, we need a characterization of EI, and a *justification* for appealing to EIs [...].

The demand for a justification for appealing to falsehoods is also addressed by several scholars (e.g., McMullin 1985; M. Elgin and Sober 2002; Weisberg 2007; Strevens 2008; Bokulich 2011; Batterman and Rice 2014, Rice 2018; Elgin 2017; Potochnik 2017). For instance, as we saw, M. Elgin, Sober, and Strevens argue that we are justified in appealing to falsehoods when they point to non-difference-makers. Rice proposes that the model's system being in the same *universality class* as the system of interest justifies exploring the latter with the former (2018). For Bokulich, the justificatory step is concerned with specifying the model's application domain and showing that the phenomenon of interest falls within that domain (2011, p. 39). Elgin argues that we are justified in using the falsehoods in question precisely because they exemplify properties. An exemplar can exemplify features that are otherwise difficult to discern. When it does, the exemplar affords *epistemic access* to those properties. She argues that felicitous falsehoods give epistemic access to rather obscured features by distorting the phenomenon of interest (2017, ch. 11-12).[8] This justifies using them.

It goes without saying that these and related accounts deserve an in-depth evaluation. But this is not my aim. The questions these accounts address are not directly concerned with the factivity of understanding.[9] The relation question and its project of establishing an appropriate tie between the falsehoods and the phenomenon of interest need to be separated from the project which is at the heart of the debate about the factivity of understanding, namely *the project of establishing how felicitous legitimate falsehoods figure into the content of understanding the phenomena*. I call the key question this project addresses the *content question*:

[8] Frigg and Nguyen's representation-as account differs in at least one important respect. According to them, the properties exemplified by a model are typically *not* the ones that are imputed to the phenomena of interest (2018, p. 217). They propose a detailed account of how the exemplified properties are related to the imputed ones via so-called specification keys (cf. 2018, sect. 5-7).

[9] For more on these and other accounts see, e.g., Suárez 2010; Downes 2011; Frigg and Hartmann 2012.

CONTENT QUESTION How do the falsehoods (or the models that involve them) *figure into the content* of understanding the phenomenon?

This question addresses how we can understand something by employing obviously false assumptions about it. The content question matters when it comes to the factivity debate. The relation question, taken by itself, does not.

It goes without saying that the relation question and the content question can be closely related. One could argue that answers to the content question can piggyback on answers to the relation question. For instance, the non-difference-maker view essentially gives the same answer to the question 'How do the falsehoods relate to the phenomenon of interest?' and 'How do the falsehoods figure into the content of understanding the phenomenon?': They point to things that do *not* make a difference to the phenomenon.[10]

But although these questions are closely related, the first key step towards my extraction view is to argue that the relation question and the content question enjoy a crucial *independence*. They are independent for at least four reasons: *First*, the question whether and how a falsehood represents the phenomenon can be also posed for *unsuccessful* or *illegitimate* falsehoods. The relation question can always be asked. *Second*, by contrast, we only wonder about a falsehood's role for the content of understanding once it is at least empirically successful. For instance, Elgin emphasizes that not every exemplifying falsehood provides understanding. A tether to the facts is required, too (2017, ch. 11). *Third*, addressing the relation question can explain why the falsehoods are empirically successful. Answers to the content question need not address this question. *Fourth*, answering how a falsehood represents (or does not represent) the phenomenon does not *necessitate* an answer as to whether or how the falsehood enters the content of understanding. Successfully representing a phenomenon (or being a fiction thereof, etc.) is not *sufficient* for being an element of a systematic account of that phenomenon. For instance, my blue hat might exemplify the Atlantic ocean's blue on a sunny day. But this is not sufficient for the blue hat playing any role in the content of my understanding the ocean's features.

Relatedly, one should keep apart the *epistemic commitment* a scientist has towards a falsehood or model while using it and the epistemic commitment involved in successfully understanding a phenomenon. A scientist might *accept* a falsehood for the sake of exploring a phenomena. But this does not mean that they accept that falsehood for understanding the phenomenon in question.

The extraction view I propose below exploits the independence of the content question from the relation question. It is compatible with different answers to the relation

[10] Some variants of the inferential account conflate these questions when it comes to *explanation* (see, e.g., Kennedy 2012, Jebeile and Kennedy 2015; Fang 2019). They would thus, arguably, also conflate them when it comes to the content of understanding.

question, including answers given by some non-factivists about understanding, such as Elgin's representation-as view. In what follows, I am thus exclusively concerned with the project of establishing how the felicitous legitimate falsehoods figure into the *content* of understanding phenomena. This project can also be phrased in terms of *providing* or *conferring* understanding. Another important step in my positive proposal is to utilize the fact that there are least two substantially different ways to provide, confer, or figure into the content of an epistemic achievement: to *enable* that achievement and to be an *element* of its content. As I argue below, the claim that felicitous legitimate falsehoods play an (epistemic) enabling role for understanding is compatible with factivism about understanding. But the claim that they are elements of its content is not. This stronger claim is what quasi-factivism and non-factivism insist on. In the following two sections, I argue that facts from science speak against it. This is the second key step in my proposal.

4 FELICITOUS LEGITIMATE FALSEHOODS AND QUASI-FACTIVISM

According to quasi-factivism, the systematic account that comprises one's understanding can contain falsehoods as elements, as long as they are in the periphery of the account (cf., e.g., Kvanvig 2003; Mizrahi 2012) or if most propositions in the account are true (Rice 2016). Let us first consider the periphery view. The periphery of an account is made of propositions that are peripheral to the subject matter in question. For instance, that my hat is blue is peripheral to its size. Or think of peripheral vision. Moti Mizrahi illustrates the periphery view with the much discussed example of the ideal gas law, which implies that the volume and temperature of gases are directly proportional to each other when their pressure is held constant.[11]

$$pV = nRT \tag{1}$$

p is a variable for pressure, V for volume, T for temperature, n for the number of gas moles, and R is a constant. Normal gases approximately behave according to this law in particular conditions, such as when the temperature of a gas is not close to its liquefaction point. The ideal gas law is derived from a number of propositions. Some of them are utterly false, such as the assumptions that gas molecules do not collide and that they do not exert any long range forces on each other, etc. (for details see, e.g., Strevens 2008, ch. 8; Rice 2019, sect. 3.1). Mizrahi claims that these falsehoods are only peripheral to understanding the gases' behavior (2012, p. 245):

[11] This law is a combination of Boyle's law, Charles's law, Gay-Lussac's law, and Avogadro's law. There are more refined forms of the law, such as the van der Waals equation, but the differences do not matter for my purposes.

> [...] although the idealizing assumptions [...] are necessary for the derivation of the gas laws, they do not define the behavior of ideal gases. It is the equation $pV = nRT$ that defines the behavior of ideal gases.

So, according to Mizrahi, only the resulting law is an element of the core of one's understanding content. The premises of its derivation basis are peripheral elements. An understanding that contains the ideal gas law is factive insofar as the law is approximately true for gases in particular conditions. It is only quasi-factive "[..] because scientists have to make idealizing assumptions [in the periphery], which may be false, strictly speaking, in order to derive the Ideal Gas Law" (2012, p. 251).

Let us bracket for a moment whether Mizrahi is right that all idealizing assumptions can be located in the derivation of the ideal gas law. Even if that were true, the *first major issue* with the periphery view is that it is not convincing upon closer examination for at least two reasons. *First*, it seems to be simply *false* that the derivation propositions are merely part of the periphery. Any equation or law is examined with respect to a given domain. One needs a domain restriction in order to determine whether a law or equation is true, e.g., whether it is actually the case that $pV = nRT$. One has to determine for each variable what its domain is. The derivation propositions contribute to determining this domain. The ideal gas law is only true for gases whose molecules do not collide, do not exert any long range forces on each other, etc. So, the propositions that Mizrahi locates in the periphery are in fact *central* to the law's application.

Second and relatedly, the law itself does not seem to be an element of the content of scientific understanding. Strictly speaking, the law is not applicable to real-world gases (for a similar point see, e.g., Cartwright 1983, part 3). Gases like methane simply do not consist of molecules that do not collide. Why do scientists nonetheless work with the ideal gas law? It is *not* because they think that the law applies to them. Instead, they use it for *comparisons*; they compare the behavior of a so-called ideal or perfect gas with the measurement results from normal gases. Here is a quote from a textbook (Massey and Ward-Smith 2006, p. 17):

> The assumed properties of a perfect gas are closely matched by those of actual gases in many circumstances, although no actual gas is perfect. The molecules of a perfect gas would behave like tiny, perfectly elastic spheres in random motion, and would influence one another only when they collided. Their total volume would be negligible in comparison with the space in which they moved. From these hypotheses the kinetic theory of gases indicates that, for equilibrium conditions, the absolute pressure p, the volume V occupied by mass m, and the absolute temperature T are related by the expression $pV = nRT$. [...] Most gases, if at temperatures and pressures well away both from the liquid phase and from

> dissociation, obey this relation closely and so their pressure, density and (absolute) temperature may, to a good approximation, be related by [the equation].

So, scientists examine how closely normal gases match the ideal gas law. The law is not applied to the gases, strictly speaking. What follows for the *content* of understanding, say, normal gases' pressure-volume-temperature relation? It does not follow that the law itself is an element of it. Clearly, a physicist does not *commit* to the proposition that normal gases behave according to the ideal gas law. At best, they commit to a comparison, e.g., that normal gases conform to the ideal gas law to a good approximation. This result supports factivism because the comparison is true.

Before I turn to the second major issue with the periphery view, let me observe that Mizrahi seems to echo my interpretation. He claims (2012, p. 247):

> [...] we can attribute to scientists who use the Ideal Gas Law understanding that is factive because the objects of their understanding involve statements of the form 'Gas g behaves approximately like an ideal gas under conditions C' (where C specifies conditions of temperature, pressure, and the like) and 'Gas g deviates from ideal-gas behavior under conditions C*'.

But if that is his view, it is a *factive* rather than a quasi-factive view. As he notes himself, these comparisons are true (2012, p. 248). And for such comparisons the derivation propositions are simply *irrelevant*. They are only relevant for deriving the law but not for the truth of the comparisons. There is thus no need for including the derivation propositions in the periphery of one's understanding account at all. However, I set these considerations aside and continue to interpret Mizrahi as offering a proper quasi-factivist view.

So far, I bracketed the question whether all idealizing assumptions can be located in the derivation basis. The *second major issue* with the periphery view is that they cannot. There are cases where the idealizing assumptions are either *directly involved* in the equations or laws or are otherwise *indispensable* for them. Let us go back to the seemingly innocent ideal gas law. Rice shows in detail that pressure and temperature are statistical variables that are defined within an *idealized* statistical framework (2019, sect. 3.1). So, upon closer examination, there are idealized assumptions involved in the law itself and not only in its derivation basis. Another example illustrates this issue in a more direct way. Take one of the equations comprising the 'Lotka-Volterra' model, which is used for analyzing predator-prey dynamics:

$$\frac{dx}{dt} = \alpha x - \beta x y \tag{2}$$

x is a variable for the number of prey, y for the number of a particular predator population, $\frac{dx}{dt}$ stands for the growth rate of the prey population per unit time t. αx

stands for the *exponential* growth rate the prey population would have if it was not being preyed upon, and βxy stands for the rate of predation upon the prey. It is assumed to be proportional to the rate at which predators and prey meet. So, roughly speaking, the equation states that changes in the prey population's number is the population's growth rate minus the rate at which it is preyed upon. The equation provides approximately correct results for the relation between the size of a prey population (say, rabbits) and the size of its predator population (say, foxes). Together with the equation for the predator population, the Lotka-Volterra model predicts cyclic population developments that are somewhat similar to the population developments observed in real-world predator-prey populations. The equation contains the exponential growth rate idealization α. According to it, the prey population reproduces exponentially when not preyed upon. This is an utterly false falsehood. Real-world prey populations would not have an exponential growth rate even if they were not preyed upon. The periphery strategy cannot be applied to such cases. According to that strategy, only the equation matters but not its derivation propositions. But in this case the falsehood is not hidden in derivation propositions. It is an element of the equation. An advocate of the periphery view could deny that *any* understanding can be gained working with such models. However, the case does not seem to be much different than the ideal gas law case. We have an empirically successful equation. If understanding is gained, the predicament for the periphery view is this: Either the understanding's content contains a central falsehood, which is incompatible with the periphery view. Or, again, only true comparisons are an element of the content, such as 'The population development in recent fox-rabbit population in the US is similar to the cyclic development of ideal predator-prey populations.' If so, there is no need for a periphery *quasi*-factive view. These comparisons are true.

One might argue that the falsehoods contained in the equations or laws or in their derivation propositions could simply be replaced with a so-called *veridical counterpart* (Strevens 2008, p. 300). A veridical counterpart of a falsehood correctly describes the respective properties.[12] Strevens proposes that one would get approximately the same application results if one were to replace the distorting falsehoods with their veridical counterparts (similarly M. Elgin and Sober 2002, p. 448; Rohwer and Rice 2013, p. 338). But even if this strategy worked in some cases, it does not work for all. There are several idealized models that involve *indispensable* falsehoods at their core. Following Rohwer and Rice (2016, p. 1134), I consider a falsehood indispensable

> [...] when removing the [falsehood] has the effect of 'destroying' or 'dismantling' the model; e.g. when the [falsehood] is essential to the model's mathematical representation of the target phenomenon.

[12] There is more than one veridical counterpart for every distorting falsehood. For instance, the collision frequency of gas molecules can be described to varying degrees of accuracy.

At least given our scientific resources and mathematics, it is claimed to be impossible in these cases to get similar (or even any) application results based on veridical counterparts. Several case studies illustrate that models featuring indispensable falsehoods are ubiquitous in science (e.g., Wayne 2011; Bokulich 2011; Kennedy 2012; Batterman 2009, Batterman and Rice 2014, Rice 2018, 2019). In this paper, I do not evaluate these studies. I assume that they are empirically accurate. Let me just give you two brief illustrating examples. One popular example are idealized models that employ the so-called *thermodynamic limit* (e.g., Batterman 2002; Batterman and Rice 2014, Rice 2018). Roughly speaking, it assumes that the number of particles of the system as well as its volume are arbitrary large. Models that employ the thermoydynamic limit are, for instance, so-called phase-transition models. Phase transitions are abrupt changes of the qualitative macroscopic properties of a system or substance, such as water's freezing into ice, the transition from liquid to gas, or the magnetization of iron. The thermodynamic limit is said to be indispensable because the phenomenon of a phase transition cannot be *produced* with a model that assumes finite particles. By employing finite systems, say, systems based on statistical mechanics, we cannot model phase transitions.[13] So, it seems that we cannot explore the nature of phase transitions without the thermodynamic limit.

Another popular example are so-called optimality or optimization models (cf. Rice 2012, Rice 2018, sect. 3.2; but also M. Elgin and Sober 2002, pp. 446-448; Potochnik 2007, 2009, 2010). The basic goal of such models is to analyze why particular phenotypic traits occur by determining optimal strategies for obtaining, say, net energy intake, given a set of limiting factors such as the costs of finding food. As Rice shows, such models do not only involve numerous idealizations but idealizations that are *entrenched* with each other (which renders them indispensable) (2018, 2019). One of Rice's cases is an optimality model of the foraging behavior of arctic wintering eiders (a kind of duck). The patterns that result from the model's application are quite similar to the patterns observed for the real-world eiders' foraging behavior. So, there is the desired empirical success. This model stipulates, among other things, that the eider population is arbitrarily large, that there is no intergenerational overlap, that the selection pressure in the eider population remains constant, and that natural selection is the only evolutionary factor that matters for the phenotypic trait's evolution (cf. Rice 2018, sect. 3.2). Importantly, "[...] the contributions made by the idealizations cannot be quarantined from the contributions made by the accurate parts of the model" (Rice 2018, p. 2808). Only the entrenched contributions allow us to produce the matching behaviors. Thereby, the idealizations become indispensable.

[13] Some argue that the thermodynamic limit is dispensable (e.g., Butterfield 2011; Norton 2012). For some useful discussion see, e.g., Shech 2013, 2017. I take for granted that it is necessary.

For none of these cases and the previous ones does the periphery view work. One cannot put the distorting falsehoods into the periphery. If understanding is gained, an understanding's content contains either the falsehoods as elements or true comparisons utilizing them. So, cases of indispensable falsehoods either speak in favor of non-factivism or factivism, but not in favor of the periphery view.

To sum up, on the one hand, the periphery view seems to be doubly mistaken: (i) the derivation propositions are not merely in the periphery and (ii) the law is not an element of an understanding's content. On the other hand, the periphery view cannot account for cases featuring falsehoods in the equations or indispensable falsehoods.

There is another quasi-factive view, namely Rice's view according to which only *most* propositions of our understanding's content need to be true.[14] But, upon closer examination, Rice's analysis of *felicitious legitimate falsehoods* does not demand a quasi-factive view. Rice considers two idealized models. For each of them, he argues something along the following lines (2016, p. 88, my italics):

> 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. [...] This information about counterfactual relevance and irrelevance then leads the modeler to acquire *factive* scientific understanding of why the phenomenon occurred in the target system. This understanding is constituted by a set of *true* beliefs about the counterfactual relevance and irrelevance of various contextually salient factors [...].

So, according to Rice, true information is obtained working with felicitous legitimate falsehoods. The only ground for counting his view as quasi-factive is that the *accurate* information "[...] is incorporated into a larger body of *mostly true* scientific knowledge about the phenomenon of interest" (2016, p. 88, my italics). However, the issue of integrating the accurate information into something else is not concerned with the role of the falsehoods for the content of one's understanding. So, from the perspective of felicitous legitimate falsehoods, Rice's view is compatible with factivism. It goes without saying that there might be other grounds for arguing that the larger account contains some falsehoods. But I am exclusively concerned with the argument from felicitous falsehoods against factivism in this paper.

The overall upshot is that if felicitous legitimate falsehoods were elements of an understanding's content, they would not be elements of its periphery. They would rather be elements of its core. In what follows, I argue that facts from science also speak against this claim.

---

[14] Rice considers his view to be a form of factivism (2016). However, I use the term 'factivism' exclusively for views that demand that *all* propositions be true.

5 FELICITOUS LEGITIMATE FALSEHOODS AND NON-FACTIVISM

Non-factivism does not take issue with central or indispensable falsehoods. For instance, according to Elgin's view, such falsehoods can be a central element of the account that constitutes the content of scientific understanding: "Felicitous falsehoods, on my theory, are integral to tenable accounts. They are not merely causal antecedents" (2017, p. 97). For instance, "[...] ecological models of population size embody an understanding of the dynamics of population regulation" (2017, p. 42; similary de Regt and Gijsbers 2017; Rancourt 2017). However, non-factivism is not the view that the content of one's understanding can *only* contain false propositions. As Elgin puts it, "[...] understanding somehow answers to facts" (2017, p. 37) – in whatever way precisely. It is just not true that "all or most of the [central] propositional commitments that comprise a genuine understanding are true" (2017, p. 37). Non-factivism is mainly argued for by pointing to the fact that felicitous legitimate falsehoods are frequently employed in science, that they play an important role for scientific progress, and that many idealizations cannot be de-idealized without loss of empirical success. Elgin sums up (2017, p. 35):

> As we have seen, many epistemically estimable accounts contain models and idealizations that do not even purport to be true.

Indeed, it is true that scientists frequently employ models and idealizations that do not even purport to be true or could not be replaced with veridical counterparts. However, this fact is *not sufficient* for concluding that the falsehoods are elements of the content of the understanding obtained by working with them. As I have argued before, the content question enjoys independence from the relation question. We need a positive argument for why the fact that scientific accounts involve falsehoods warrants that these are elements of the core of one's understanding's account.

To argue against non-factivism, one could try to diminish the role of the falsehoods for the scientific accounts or for the understanding obtained. But this is not my approach. I embrace Elgin's claim that "[...] the ideal gas law figures in the understanding provided by thermodynamics [...]" (2017, p. 61). However, the crucial question is *how* they figure into our understanding. Figuring into an epistemic achievement does not mean one is an *element* of its content. As I emphasized before, one could also figure into it by playing an *enabling* role. In the next section, I argue for such an enabling view. In this section, I provide arguments against the claim that the falsehoods figure into one's understanding by being an element of its content by pointing to three facts from scientific practice that conflict with this claim.

The *first* fact from scientific practice is that scientists themselves are aware of the fact that the felicitous legitimate falsehoods are false; they are *deliberate* falsehoods. Recall the textbook quote. It was clear that the ideal gas law is concerned with a 'perfect' or 'ideal' gas and that normal gases are *not* such gases. Pincock describes models that involve the falsehood that the ocean is infinitely deep (2014). This falsehood is also not considered to be true. Relatedly, as I have pointed out before, the falsehoods are typically *not applicable* to the phenomena of interest. The ideal gas law applies to 'ideal' gases and optimality models to 'optimal' organisms. It does not seem to be plausible that a scientist's understanding of a phenomenon is comprised of an account containing something that is *known to be false* of it and *known to be not applicable* to it. After all, understanding is an epistemic *commitment* to the account. This argument from deliberate falsehoods is strengthened by the fact that if the deliberate falsehoods were elements of one's understanding's account, any understanding subject would face a *clash of contradictory epistemic commitments*. The felicitous deliberate falsehoods with which we are concerned in this paper are not just utterly false, but they also *contradict* known properties of the real-world entities. An understanding subject would have to accept at the same time that the ocean is infinitely deep and not infinitely deep, that the particles of the phenomenon of interest are both arbitrarily large and finitely large, etc. In view of these plain contradictions, Shech concludes with respect to the Aharonov-Bohm (AB) effect that "[...] [the infinity idealization is] not essential for the explanation and understanding of the manifestation of the concrete, physical AB effect" (2017, p. 10). In general, explaining or understanding phenomena does not seem to be compatible with contradictory propositions about the phenomena. It is also not clear to me how Elgin's account can cope with these contradictions. According to her, "[a] tenable [understanding] account is a tapestry of interconnected commitments that collectively constitute an understanding of a domain" (2017, p. 16). It seems difficult to imagine that such plain contradictions can fit into a complex of interconnected epistemic commitments, especially in light of the fact that Elgin additionally claims that these commitments involve the disposition to *act* on the content one commits to (cf. 2017, ch. 2).

One might object that deliberate falsehoods are *professed* in scientific practice and thereby become an element of the account. What is professing? According to Elgin, "[...] to profess that *p* is [...] to make *p* available to function as a premise or rule of inference in a given context for a given cognitive purpose" (2017, p. 21). It might be reasonable to assume that scientists profess or accept a falsehood for the sake of working with it (see also Potochnik 2017, ch. 4.1.1). They act as if it were true. However, this objection is not convincing. Even if something is professed, it does not *thereby* become an element of someone's account of a phenomenon. Take a different case: In *reductio ad absurdum* arguments, the claim to be defended is *negated* as a supposition. The very point of such arguments is that the claim of interest should be accepted because its negation would be untenable. The negation is professed but not integrated into one's account of the phenomenon. Professing that *p*, taken by itself, thus does not require *integrating p* into one's account.

The *second* fact from scientific practice supports the first argument. It is not uncommon to use several idealized models for exploring the same phenomenon. Michael Weisberg calls such cases 'multiple-models idealizations' (2007). For instance, apart from the ideal gas law, there is the van der Waals equation and the virial equation for modeling gases' pressure-volume-temperature relation. All of them are employed in science (including the heavily idealized ideal gas law). Not uncommonly, these models are *contradictory* idealized models. A common example are models used for analyzing water. Some construe water as a continuous fluid and others construe it as composed of discrete particles. All these models are employed to examine the features of water, although it is *impossible* for the employed distorting idealizations to be true at the same time. Water cannot be a continuous fluid *and* composed of discrete particles. While these are cases where scientists *switch* between models, there are also cases where a model *contains* incompatible sub-models. For instance, some multi-scalar models model fluids differently at different density scales. Elgin argues that multiple-models idealization is not an issue because "[e]ach model exemplifies different features and affords epistemic access to different aspects of the target" (2017, p. 267).[15] I agree that this modeling practice is legitimate. However, if so, it clearly does not seem to be plausible that all central falsehoods involved in these multiple models are elements of one's understanding of the phenomena. This might lead to an even bigger clash of commitments. Scientists neither believe that water is a continuous fluid nor that it is composed of discrete particles. Assuming that they commit to both falsehoods in their understanding's accounts *in addition* to believing that water consists of molecules seems to be absurd. Even if one distinguishes the objects of understanding relative to different aspects of the target phenomena examined, this would not change my argument. The false assumptions about the nature of water are not restricted to *aspects* of water but are concerned with its general properties.

But we need not just wonder about scientists' commitments from the armchair. Here is a quote from an essay on optimality models that appeared in *Nature* and was written by two influential biologists (Parker and Smith 1990, p. 27, my italics):

> [...] If, by this review, we could lay rest to the idea that the application of optimization theory requires either that we assume, or that we attempt to prove, that organism are optimal, we would be well satisfied. [...]
>
> We distinguish between general models and specific models, though in reality they form part of a continuum. General models have a *heuristic* function; they give qualitative insights into the range and forms of solution for some common

[15] For other explanations why scientists use multiple models for examining the same phenomenon see, e.g., Morrison 1999; Bailer-Jones 2003; Weisberg 2007; Elliott-Graves and Weisberg 2014.

> biological problem. [...] Specific models are designed to be applied quantitatively to specific species, and include parameters that can readily be measured. They are often modified (more complex) versions of some general model, devised specifically for comparison with a particular set of observations. If the predictions of the model match the biological observations, we may hope that we have made *correct* assumptions about the nature of adaptions.

Biologists do not assume that organisms behave optimally. This assumption just serves to construct and test models. Importantly, if these are empirically successful (e.g., matching predictions), it is still *not* concluded that the organisms behave optimally. Instead, true propositions about the adaption mechanism are (hoped to be) *inferred*. We return back to this point in the next section.

Independent of the 'clash of epistemic commitments' objection, there is another general objection. The *third* fact from scientific practice is that scientists *appreciate* the limited empirical success of felicitous legitimate falsehoods. As I mentioned throughout the paper, they only *approximately* recover behaviors or only deliver approximately correct predictions. Take the ideal gas law. As Strevens emphasizes, Boyle's law "[...] does not hold exactly for any gas" (2008, p. 305).[16] Rice claims that "[...] many idealized model systems are known to *approximate* the patterns of behavior of real-world systems [...]" (2018, p. 2812, my italics). Scientists are aware of these facts. They typically do not gloss over the fact that the predictions are only approximately accurate. Their understanding of the phenomena based on the falsehoods is thus expected to involve an *appreciation* of the approximation. And this expectation is not disappointed. Recall the textbook quote. It was emphasized that the ideal gas law can only be considered to be a good approximation. In a paper on the foraging behavior of honeybees, two optimality models are tested. The model that is considered to be best is one where "[...] the predictions from the efficiency model do not differ significantly from the observations [...]" (Schmid-Hempel et al. 1985, p. 63). The model is not claimed to perfectly recover or predict the phenomenon. This suggests that the falsehoods themselves are not elements of the account an understanding subject commits her- or himself to. And, again, if a true comparison like 'Phenomenon X behaves approximately like falsehood F' is part of the commitments, this speaks in favor of factivism and not non-factivism.

Let me sum up: Non-factivism emphasizes that felicitous legitimate falsehoods are ubiquitous in science. The upshot of my analysis is that they are nonetheless not elements of the core (or, for that matter, the periphery) of an understanding's content. The fact that scientists know that they are utterly false, not applicable to the

[16] This should not be surprising. It is precisely because idealized models involve distorting falsehoods that they do not provide us with accurate results. For instance, the effect of collisions between gas molecules has *some* effect on the behavior of the gas.

phenomena, and only approximately match the observations, and that including them into an understanding's account would lead to contradictions, strongly suggests that the falsehoods are not something an understanding subject commits to. But how then do they figure into understanding? In what follows, I answer this question.

6 FELICITOUS LEGITIMATE FALSEHOODS AND FACTIVISM

Given the importance of felicitous legitimate falsehoods for scientific progress, it seems clear that they *figure into* scientific understanding, and that, at least in some cases, they are indispensable for it. I promised a *factive* view that can account for these and other central assumptions. In particular, I claimed that my view can account for the assumptions that (i) felicitous legitimate falsehoods facilitate understanding of phenomena, (ii) by virtue of being the very falsehoods they are, (iii) and in an epistemically relevant way, and (iv) that they might be might be indispensable for understanding. In what follows, I show how my view achieves this.

6.1 *Four key steps*

As I mentioned before, one key step towards my view was to separate the relation question (How do the falsehoods relate to the phenomenon?) from the content question (How do the falsehoods figure into the content of understanding the phenomenon?). This separation opens up promising options for factivists about understanding. By appealing to indispensable falsehoods, quasi-factivists and non-factivists about understanding typically argue that the link between veridical representation and understanding should be loosened (see, e.g., Elgin 2017; Rice 2018; de Regt 2015), agreeing with Andrew Wayne's verdict that "the close link hitherto assumed between successful [i.e., verdical] representation and explanation should be loosened" (2011, p. 831; see also, e.g., Batterman 2009, Batterman and Rice 2014, Rice 2016, Rice 2018; Kennedy 2012). But due to the independence of the relation question from the content question, factivists can agree that this link should be loosened; they only need to *additionally* argue that the close link hitherto assumed between representation and understanding should be loosened.[17] And this is what my factive view does, as I show below.

My second key step was to argue against the views that felicitous legitimate falsehoods figure into an understanding's content by being elements of its periphery and core, respectively.

Now I turn to my third key step. It is twofold. Its first part is to exploit the distinction between the *process of obtaining understanding* and the *obtained understanding*: Facilitating understanding or figuring into it is not exhausted by being an element of

[17] This argumentative step is also at the heart of Rice's quasi-factivist view (2016). He argues that accurately representing the real-world target system in question is not required for understanding.

it. By playing an enabling role for understanding, one also facilitates or figures into to it. Enabling understanding takes place in the *process of obtaining* understanding. Being an element of the content (also) concerns the obtained understanding, i.e., the *product* of that process. It goes without saying that there could be more or less understanding of a phenomenon. But any 'gradable' account of understanding needs to specify a minimal threshold for when *some* understanding is obtained.

Importantly, enabling something is *independent* of being an element of it. For instance, my teaching can figure into my students' successfully written exams by training skills that enable them to write those exams, but my teaching is not an element of the successfully written exams. My teaching figures into the success but is not a constitutive part of it. Likewise, something can figure into obtaining an understanding's content without being an element of its content. According to the view I defend, this is precisely the role felicitous legitimate falsehoods play. In a nutshell, they function as tools that help us to build systematic accounts of the phenomena of interest; they help us to *extract* relevant information. I thus dub it 'the extraction view'. This accounts for (i). Below, I go into its details.

The second part of my third key step is to argue that at least in current scientific practice this extracted information turns out to be *accurate* information. This is the *factive* part of my view. It draws from several case studies, as I show below.

In the subsequent section, I turn the *final step* in my proposal. I argue that the extraction view is perfectly compatible with (ii)-(iv), namely that the falsehoods have this enabling role by virtue of being the very falsehoods they are, that they facilitate understanding in an epistemically relevant way, and that they might be indispensable. The process-product distinction is crucial here. I argue that the falsehoods might be ineliminable in the process of obtaining understanding but not for the understanding obtained. This is fully compatible with factivism about understanding, which only concerns the factivity of the *content* of understanding.[18]

6.2 *The extraction view*

A central claim of the extraction view is that felicitous legitimate falsehoods facilitate understanding a phenomena by enabling us to extract information about it. For instance, one works with the ideal gas law to extract information about normal gases. The extraction view decouples representation and understanding. It is not by virtue of representing the phenomena (or by being a relevant fiction, etc.) that felicitous legitimate falsehoods contribute to the content of understanding that phenomenon. What is contained in a scientific model and what a scientist commits to when attempt-

[18] As Elgin emphasized in conversation, one might separately discuss whether the falsehoods figure into the *justification* of the understanding's contents. I do not consider this possibility here. But in light of my arguments against non-factivism, I do not expect that this is the case.

ing to understand a phenomenon based on it thus need not coincide. The extraction view also decouples the justification that renders the falsehood legitimate from the obtained understanding. For instance, being in the same universality class as the phenomenon to be understood might justify using an idealized model to analyze the phenomenon (cf., e.g., Batterman and Rice 2014, Rice 2018). But, according to the extraction view, only the information acquired about the shared properties is an element of an understanding's account.

The arguments against quasi-factivism and non-factivism speak for such an extraction view. The fact that scientists are aware that the felicitous legitimate falsehoods are utterly false and only provide approximately matching results speaks in favor of the fact that they do not take them at their face value but extract information from them due to their empirical success. The fact that scientists work with incompatible idealized models for the same phenomenon (e.g., different water models) strongly suggests that such models do not provide either individually or taken together an account of the phenomenon to which the scientists commit. Instead they extract information about different aspects of the phenomenon using the models.

The extraction view draws lessons from other views, although it differs from them. For instance, the extraction view shares the decoupling of representation and understanding with Knuuttila's epistemic artefact view of idealized models (2011, Knuuttila and Boon 2011), according to which the construction of idealized models is best understood as the construction of a hypothetical entity that facilitates the *exploration* of the phenomenon. However, an important difference is that the extraction view is not to committed to *any* particular view about the representational nature of the falsehoods or the models that involve them. It is *neutral* regarding the relation question. For this reason, the extraction view also does not collapse into an inferential representation account like Suárez (e.g., 2002), although it shares the idea that the model users draw inferences from working with the model. Moreover, as I address below, a couple of authors proposed that models can be used to extract particular kinds of information. For instance, Rice proposed that idealized models enable us to extract *modal* information about the phenomena of interest (Rice 2018). The extraction view I propose embraces these views. But it is neither limited to a particular kind of information that is extracted nor to idealized models.

In order to use the extraction view to defend the *factivity* of understanding, one needs to argue that *accurate* information is extracted. My strategy is to utilize several case studies of scientific falsehoods (e.g., Alexandrova 2008; Pincock 2014; Bokulich 2016; Rice 2016, 2018) and other facts about scientific practice to show that the information extracted from the falsehoods or models that facilitates understanding is indeed accurate.[19] Let me start by going back to the quote from *Nature*:

> If the predictions of the model match the biological observations, we may hope that we have made *correct* assumptions about the nature of adaptions.

Scientists make educated guesses when constructing falsehoods or idealized models. When these are empirically successful, this is meaningful. It seems that some guesses were correct. But, as we have seen, it does not seem plausible to infer from the empirical success that the model qua model captures the phenomenon correctly. It contains elements known to be false. And this is also not the inference drawn by scientists. Instead, they infer that they have made *some* correct assumptions. The empirical success suggests that the falsehoods "[...] capture something that is right," as Reiss puts it (2012, p. 49). Bokulich advocates a view along these lines for so-called scientific fictions. She states (2016, p. 261): "fictions can be an effective means by which we can come to understand truths that would otherwise be difficult to grasp." According to the factive extraction view, it is this something that matters for the content of understanding. When we work with deliberate falsehoods, we want to extract accurate information about the phenomenon of interest. These pieces of information are elements of an understanding's account of the phenomenon, but not the falsehoods themselves. Felicitous legitimate falsehoods are tools for arriving at the relevant information. Elgin states that "[t]he ideal gas law accounts for the behavior of gases by describing the behavior of a model gas [...]" (2017, p. 61). The extraction view would state more carefully: The ideal gas law captures something that is true of normal gases, namely that the volume and temperature are, approximately, directly proportional when the pressure is held constant. This true information accounts for the property of normal gases.

The case studies reveal that there is a variety of extraction methods or results when extracting true information using falsehoods. In this paper, I do not add any novel extraction methods. The extraction view is not limited to any specific ones, but allows for a variety of *legitimate* extraction methods. Extracting information should not be arbitrary. The legitimacy of the described extraction methods derives its status from the justification to appeal to the falsehoods in the respective accounts (i.e., from the project of establishing the tie to the phenomenon). I assume that this justification link typically suffices for establishing the legitimacy of the method.[20] But let me turn to concrete methods. In some cases, the understanding's content is *closely related* to the falsehoods. In the case of the ideal gas law, one can understand the normal

[19] Note that this result is not specific to understanding. It also matters for debates about model explanation, etc.

[20] Recall that the justification question need not collapse into the relation question.

gases' pressure-temperature-volume relation by factoring into account how much the falsehoods in the law's derivation basis differ from reality. For instance, understanding that the molecules' collisions slightly affect the pressure-temperature-volume relation partially explains why the normal gases' measures differ from the ideal gas law's predictions. In such a case, one can construct a veridical counterpart of the gas law's derivation (for a detailed case study see Strevens 2008, ch. 8; for a critical view see Rice 2019, sect. 3.1). Pincock illustrates another case of a close relationship between the falsehoods and the understanding's content. He argues that there are cases where we can *mathematically transform the original equations* that contain the falsehoods (in his case the assumption that the ocean is infinitely deep) to equations that don't (2014). Alexandra proposes a similar approach for economic models (2008). She suggests constructing *open formulas* from felicitous idealized models that can be applied to real-world cases.[21] In other cases, the result we obtain from working with a falsehood is a hypothesis that needs to be refined. For instance, in the case of the Lotka-Volterra model the fit of measured data and the model's prediction is only a rather loose one. The measured population development is only roughly cyclic-like. Such an empirical success, at best, gives rise to the hypothesis that there are cyclical-like (dependency) relationships between predator and prey populations. However, such a rough hypothesis does not seem enough for understanding the predator-prey relations. And indeed biologists did not leave the model unchanged. Instead, they refined and de-idealized it to obtain a better fit with the measured data.

In all the cases so far, the extracted information was closely related to the falsehoods. But there are also other extraction methods. For instance, in some cases one can infer information based on the fact that an idealized model as a *whole* is empirically successful. According to Kennedy, using idealized models can generate so-called comparison cases that allow model users to identify explanatory 'difference-making' factors (2012, p. 331). According to Bokulich (2016) and Rice (2016, 2018), one can extract *counterfactual* information from working with models. Take the lattice gas models. According to Rice, such models "[...] can be used to *extract* various kinds of modal information that can be used to provide explanations of the behaviors of real fluids" (2018, p. 2802, my italics). For instance, '[l]attice gas models can show us how patterns of fluid flow counterfactually depend on the fluid's pressure, density, and viscosity" (2018, p. 2810). It is arguably this modal information that is an element of understanding fluid flow. Or recall the eider model. It examines the foraging behavior of eiders. Eiders dive to catch their food (typically blue mussels). What is peculiar about their foraging behavior is that the eiders dive most frequently at times that are

[21] Contra Alexandrova, the information we extract from models is typically not the *starting* point of working with the model but the end. It is because of the model's empirical success that we have a good reason to believe that the extracted information holds true for real-world phenomena.

the *least* profitable from a preying point of view, namely when the currents are getting stronger. Instead, one would expect them to dive when the currents are less strong. So, the explanandum of interest is, say, 'Why do the eiders dive most frequently at times that are the *least* profitable?' The result of applying the eider model leads to the hypothesis that the foraging pattern seems to be an adaptive response to a tradeoff between short-term energetics in dive cycles, longer term digestive constraints, and the cyclical nature of tidal currents (cf. Rice 2018, p. 2805). Rice writes explicitly that (2018, p. 2808, my italics)

> [...] the mathematical techniques involved in optimality modeling *allow biologists to extract modal information* about how the explanandum counterfactually depends on population-level constraints and tradeoffs among fitness enhancing variables and why most of the physical features of the target system(s) are irrelevant to the target explanandum.

It is these results one commits to in understanding the eiders' behavior.

Extracted information can also be information about *which factors are negligible* for the phenomenon of interest, as suggested by Strevens (2008, 2017; see also M. Elgin and Sober 2002; Khalifa 2017, ch. 6.3.2).[22] An illustrative example is given by Elgin herself (2017, p. 263, my italics):

> Familiar gas models represent gas molecules as lacking attractive force. If the results of our calculations are confirmed when we read them back into the target, we have reason to think that intermolecular forces do not play a significant role in gas dynamics. Knowing as we do that every material object attracts every other one, we do not conclude that there is no attraction. Rather, we *conclude* that for the sort of understanding we currently seek [...] intermolecular attraction is negligible.

According to the extraction view, what is an element of the content of understanding normal gases' volume-temperature relation is only the *true* proposition that

[22] The extraction view is nonetheless substantially different from the non-difference-maker view of scientific falsehoods in at least two respects. (a) According to the extraction view, there is no need for the falsehoods to be rendered harmless. They do *not* need to point to non-difference makers. It only matters that they are legitimate and that they can be used to extract relevant information. (b) The non-difference-maker view is crucially limited. Strevens assumes that the non-idealized parts of a model describe the difference-makers for the explanatory target and the idealized parts point to non-difference makers (Strevens 2008, p. 318):

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

This account cannot cope with cases where the idealizations are *indispensable* for extracting the explanatory information. In such cases, the non-idealized parts do not contain all the difference-makers for the phenomenon to be explained. Instead, one extracts information about the relevant difference-makers from the falsehoods. The account can also not cope with cases of falsehoods that *distort difference-makers*. For examples of both cases see, e.g., Rice 2018. The extraction view can easily cope with these cases.

intermolecular attraction is negligible for the pressure-volume-temperature relation, not the known-to-be-false assumption that gas molecules do not collide.

To sum up: When felicitous falsehoods are empirically and robustly successful and legitimate, we conclude that they capture something that is *true* of real-world phenomena. Then, we can extract what is captured and the latter fosters our understanding of the phenomenon of interest. The extracted information is an element of the proposition that comprise a scientific understanding of the phenomena. As we have seen, there is a variety of information that can be extracted from felicitous legitimate falsehoods. Sometimes one gains a schematic content and sometimes information about counterfactual facts. So, there can be a close relationship between the falsehoods (or the model) and the extracted information; but there doesn't have to be. In the next section, I address how the extraction view can do justice to the *epistemic value* of the falsehoods. But before I turn to the epistemic addition to the extraction view, I want to address some challenges for and limits of the extraction view.

6.3 *Challenges for and limits of the extraction view*

The extraction view faces some challenges and limitations. *First*, the extraction view needs to explain why scientists often *explicitly* work or reason with the falsehoods but do not always with extracted information (as Elgin emphasizes (2017, ch. 2). I think that there are three main reasons: (i) Falsehoods are often employed for their *practical value*. They are easier to handle in calculation and theoretical reasoning and they might be needed to solve the model's equations. If one is *aware* that and in which respects they are inaccurate, there can be more value than harm in using the falsehoods. (ii) Such cases do not seem to be too much different from using deliberate falsehoods in our everyday life. Take Elgin's example of Fiona who accepts for the purposes of travel planning that Amherst is 90 miles from Boston (2017, ch. 2). Obviously, Fiona does not accept that Amherst is *precisely* 90 miles from Boston. A most natural description of what she accepts is that Amherst is *approximately* 90 miles from Boston. When it comes to numbers we often streamline. We often say that it is 3:00 pm, although it is 3:02 pm, and so on. One might object that science – in contrast to everyday life – is a matter of precision. We can be sloppy in everyday life, but not in science. However, the crucial similarity is not the sloppiness. It is the fact that the subjects are *fully aware* of the fact that the claims are not true, strictly speaking. (iii) We use the falsehoods for their *epistemic value*, as I explain below. For instance, the falsehoods (or the models that involve them) might provide us with important epistemic access to the relevant information.

Second, I offered no *knockdown* argument for the claim that felicitous legitimate falsehoods contribute to the understanding's content by enabling one to extract its elements. But I think it is the best view to accommodate all the facts about understanding and felicitous legitimate falsehoods, as I have argued above.

Third, the extraction view does not offer any *guarantee* that only *true* information is extracted from the falsehoods; nor that understanding subjects exclusively commit to true propositions. None of the extraction methods that I mentioned guarantees that. But guaranteeing the latter is not the aim of my view. I only want to account for the factivity of understanding in view of felicitous legitimate falsehoods. I also do not think that it is a problem that the extraction does not guarantee that only true information is extracted from the falsehoods. On the one hand, this problem is not specific to the extraction view. Except for deductive reasoning based on true premises, there are no inference methods that guarantee true results. On the other hand, I can arguably shift the burden of proof. I am not aware of any case studies of felicitous legitimate falsehoods where false information that was *extracted* from models was considered to be an element of one's understanding's account. So, the burden of proof is to provide a convincing counterexample to the extraction view.

Fourth, the extraction view does not offer any story about how to construct the *whole* systematic account (e.g., the complete explanation) from working with a falsehood or model. As we have seen, we typically only extract *elements* of it but rarely (if at all) the whole systematic account. This is arguably a cost to pay for decoupling representation from understanding. But I do not think it is too a high a cost to pay. What counts as a sufficiently systematic account of a phenomena for understanding it is a project of its own. If the systematic account is supposed to be an explanation, we need a theory of explanation. Something analogous is true for any other systematic account that is considered acceptable for understanding phenomena.

7 THE EPISTEMIC VALUE OF FELICITOUS LEGITIMATE FALSEHOODS

According to the extraction view, felicitous legitimate falsehoods are compatible with a factive view of understanding. Now I turn to the fourth key step in my proposal, namely to account for the *epistemic value* of these falsehoods and to show that they are no cognitive crutches. One needs to show that the falsehoods facilitate understanding by virtue of being the very falsehoods they are, that they do it in an epistemically relevant way, and that they might be indispensable. Elgin doubts that factivists (what she calls veritists) can account for these desiderata (2017, pp. 28-29):

> If he is feeling charitable, the veritist might grant that such devices could play a heuristic role in the presentation or application of the theory. But he would insist that for the account to be epistemically acceptable, they must be excisable with no loss of cognitive content or epistemic justification. Such a position is likely to implausibly discredit much of our best science. [...] Veritists evidently have to simply deny that accounts that ineliminably deploy devices like the

> Hardy-Weinberg model or the mutually indifferent deliberators behind the veil of ignorance are epistemically acceptable.

Elgin is not alone with her concerns. Rice argues that "[t]o simply reject these highly idealized models as 'nonexplanatory' would be to render incomprehensible much of what contemporary science has purported to explain" (2018, p. 2809). Daniela Bailer-Jones emphasizes that "[i]f a model is only an heuristic device, it need not tell us anything about how things really are concerning the phenomenon modelled" (2003, p. 60). And Julian Reiss writes (2013, p. 282):

> The problem is [...] that if that is all models do, the mystery is not resolved. Why do economists build complex, mathematically sophisticated models rather than, say, resort to creativity and intuition, crystal balls, hypothesis-generating algorithms or consciousness-enhancing drugs? All of these sources of inspiration would be a lot easier to come by, and some of these would be more fun, than doing the hard work of constructing and solving a model. To warrant their existence, models must do more than to provide hypotheses. They must have some genuine epistemic benefit.

However, these concerns can be addressed by combining the extraction view with a plausible answer to the *separate* question of how the falsehoods (or the models that involve them) *relate* to the real-world phenomena. We need an answer to the relation question to answer why models are better than crystal balls or why some models might be even indispensable in science. The extraction view only tells us something about the relationship between the falsehoods and the content of understanding. Importantly, answers to the relation question can go hand in hand with the extraction view. Let me illustrate this by showing that Elgin's exemplification view is perfectly compatible with the extraction view. Recall that, according to her view, felicitous falsehoods provide us with *epistemic access* to the phenomena of interest by *exemplifying* some of their relevant features. She also emphasizes that they provide us with access to crucial information that would often difficult to get otherwise. This is why they facilitate understanding by being the very falsehoods they are – in an epistemically relevant way. Such falsehoods could not be replaced by crystal balls or hypothesis-generating algorithms because they – unlike these balls or algorithms – provide epistemic access to information. This is their epistemic value. Such a combined view thus does not render the felicitous legitimate falsehoods *epistemically* idle. Moreover, that they enable us to extract the information explains the judgment that models are explanatory. We consider them to be explanatory because we explain with propositions that are extracted from the models. An extraction view is also not committed to the claim that falsehoods should be avoided in science. Valuable tools are kept and not removed. Without idealizations it would be notably more difficult to gain access to information that is required for understanding (for a similar point see, e.g., Craver 2014; Potochnik 2017; Strevens 2017). And that is also why "[...] scientific (and other disciplinary) communities are within their epistemic rights to deploy them" (Elgin 2017, ch. 5).

But what about cases where the falsehoods are considered to be *indispensable*? In these cases, they give us an epistemic access that could perhaps not be gained otherwise. As Rice claims (2018, pp. 2809-2810; similarly Morrison 2009, p. 128):

> [...] these modeling techniques are often [...] ineliminable because they allow scientific modelers to extract the desired explanatory information that would otherwise be inaccessible. [...] without these [...] techniques [they] would no longer be able to provide certain scientific explanations.

Batterman puts it as follows (2009, p. 428):

> [...] some idealizations are explanatorily ineliminable. That is to say, I argue that the full understanding of certain phenomena cannot be obtained through a completely detailed, nonidealized representation.

However, the only thesis that these claims justify is a thesis that is compatible with factivism, namely that the falsehoods are indispensible in the *process of obtaining understanding*. The claim that they are indispensible for the *content* of understanding is a *non sequitur*. It is perfectly fine to endorse the genuine-epistemic-access stance on models and felicitous legitimate falsehoods without thereby endorsing the claim that falsehoods are elements of the content of the *obtained* understanding. Being essential for providing epistemic access to crucial explanatory information is not the same as being explanatorily ineliminable. Consider a different case. Imagine that there is just one method for detecting whether some rash is a disease of a particular kind. Without this method we couldn't know whether a person has that disease. In other words, we would not have epistemic access to this information otherwise. But even if so, the explanation of the fact that this person has the disease need not involve the method. Similarly, an understanding that makes use of information that could only be gained by means of falsehoods is not required to contain any of the falsehoods among the propositions the understanding subject commits her- or himself to.

8 CONCLUDING REMARKS

Understanding a phenomenon requires an epistemic commitment to a systematic account of it. In view of the fact that felicitous legitimate falsehoods figure into our scientific understanding, Elgin maintains that retaining the truth requirement for understanding amounts to construing science as cognitively defective or as non-cognitive. I have tried to show that this is not the case. A factive view of understanding can fully accommodate the epistemic importance of scientific falsehoods. By distorting their target objects, felicitous legitimate falsehoods might provide us with epistemic access to true information that is difficult or even impossible to discern otherwise. That is why their falsity is a virtue and not a vice. Science is not cognitively defective or non-cognitive in employing such tools. But felicitous falsehoods are not more nor less than tools. Our understanding of the phenomena is grounded in the true information we extract from the falsehoods. This does not render the falsehoods themselves elements of the understanding's content. To use a Wittgensteinian image, one might need a ladder to climb up; without it one does not reach the top. But after climbing, the ladder need not be taken along.[23] Similarly, there might be no understanding regarding some phenomena without falsehoods. But the obtained understanding does not contain those falsehoods as propositions one is committed to. According to the factive view I proposed, science does not need to get rid of felicitous legitimate falsehoods. They will presumably continue to be a crucial element of the advancement of science.

Acknowledgements
Discussions with colleagues and advisors contributed to shaping the view that I defend in this article. I'm grateful to (in alphabetical order) Christoph Baumberger, Jochen Briesen and his students, Henk de Regt, Finnur Dellsén, Anna-Maria Asunta Eder, Catherine Elgin, Benjamin Feintzeig, Roman Frigg, Philipp Haueis, Christoph Jäger, Marie I. Kaiser, Kareem Khalifa, Federica Malfatti, Christian Nimtz, Thomas Spitzley, Michael Strevens, Emily Sullivan, Raphael van Riel, Kate Vredenburgh, and the participants of Thomas Spitzley's and Christian Nimtz's research groups. I also thank the audiences in Aarhus, Atlanta, Barcelona, Bochum, Bordeaux, Exeter, Ghent, Greensboro, Innsbruck, Pärnu, and Seattle, as well as three anonymous reviewers for their constructive criticisms and suggestions.

I gratefully acknowledge that part of my research for this article was funded by the Volkswagen Foundation for the project 'A Study in Explanatory Power', by the German Academic Exchange Service (DAAD) for a research stay at the New York University (2015-2016), by the OeAD for an Ernst Mach Scholarship, and by an Emmy Noether Grant from the German Research Council (DFG), Reference No. BR 5210/1-1.

[23] This metaphor is meant to be compatible with 're-using the ladder to reach the top,' i.e., with re-using the falsehoods to rebuild one's understanding or to help others understanding the phenomena. I thank an anonymous reviewer for emphasizing this point.

REFERENCES

Alexandrova, A. (2008). Making Models Count. *Philosophy of Science*, 75(3):383–404.

Bailer-Jones, D. (2003). When Scientific Models Represent. *International Studies in the Philosophy of Science*, 17(1):59–74.

Batterman, R. (2002). *The Devil In The Details: Asymptotic Reasoning In Explanation, Reduction, And Emergence*. Oxford University Press.

Batterman, R. (2009). Idealization and Modeling. *Synthese*, 169:427–446.

Batterman, R. and Rice, C. (2014). Minimal Model Explanations. *Philosophy of Science*, 81(3):349–376.

Baumberger, C. and Brun, G. (2017). Dimensions of Objectual Understanding. In Grimm, S. R., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, pages 165–189. Routledge.

Bokulich, A. (2011). How Scientific Models Can Explain. *Synthese*, 180:33–45.

Bokulich, A. (2012). Distinguishing Explanatory from Nonexplanatory Fictions. *Philosophy of Science*, 79(5):33–45.

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

Bokulich, A. (2016). Fiction As a Vehicle for Truth: Moving Beyond the Ontic Conception. *The Monist*, 99(3):260–279.

Butterfield, J. (2011). Less Is Different: Emergence and Reduction Reconciled. *Foundations of Physics*, 41(6):1065–1135.

Cartwright, N. (1983). *How The Law of Physics Lie*. Clarendon Press.

Craver, C. (2014). The Ontic Account of Scientific Explanation. In Kaiser, M., Scholz, O., Plenge, D., and Hüttemann, A., editors, *Explanation in the Special Sciences. The Case of Biology and History*, pages 27–52. Springer, Berlin.

de Regt, H. (2015). Scientific Understanding: Truth or Dare? *Synthese*, 192(12):3781–3797.

de Regt, H. and Gijsbers, V. (2017). How False Theories Can Yield Genuine Understanding. In Grimm, S. R., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, pages 50–74. Routledge.

Downes, S. (2011). Scientific Models. *Philosophy Compass*, 6(11):757–764.

Dray, W. (1957). *Laws and Explanations in History*. Oxford University Press.

Elgin, C. (2007). Understanding and The Facts. *Philosophical Studies*, 132(1):33–42.

Elgin, C. (2017). *True Enough*. MIT Press.

Elgin, M. and Sober, E. (2002). Cartwright on Explanation and Idealization. *Erkenntnis*, 57(3):441–450.

Elliott-Graves, A. and Weisberg, M. (2014). Idealization. *Philosophy Compass*, 9(3):176–185.

Fang, W. (2019). An Inferential Account of Model Explanation. *Philosophia*, 47(1):99–116.

Frigg, R. (2010). Models and Fictions. *Synthese*, 172(1):251–268.

Frigg, R. and Hartmann, S. (2012). Models in Science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2012 edition.

Frigg, R. and Nguyen, J. (2018). The Turn of the Valve: Representing with Material Models. *European Journal for Philosophy of Science*, 8(2):205–224.

Godfrey-Smith, P. (2009). Models and Fictions in Science. *Philosophical Studies*, 143:101–116.

Gordon, E. (2017). Understanding in Epistemology. *Internet Encyclopedia of Philosophy*, page https://www.iep.utm.edu/understa/.

Greco, J. (2014). Episteme: Knowledge and Understanding. In Timpe, K. and Boyd, C., editors, *Virtues and Their Vices*, pages 287–302. Oxford University Press, Oxford.

Grimm, S. (2011). Understanding. In Pritchard, D. and Bernecker, S., editors, *The Routledge Companion to Epistemology*, pages 84–94. Routledge.

Grüne-Yanoff, T. (2013). Genuineness Resolved: A Reply to Reiss' Purported Paradox. *Journal of Economic Methodology*, 20(3):255–261.

Hughes, R. (1997). Models and Representation. *Philosophy of Science*, 64:325–336.

Jebeile, J. and Kennedy, A. (2015). Explaining with Models: The Role of Idealizations. *International Studies in the Philosophy of Science*, 29(4):383–392.

Jones, M. (2005). Idealization and Abstraction: A Framework. In Jones, M. and Cartwright, N., editors, *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences*, pages 173–217. Rodopi.

Kennedy, A. (2012). A Non Representationalist View of Model Explanation. *Studies in History and Philosophy of Science*, 43(2):326–332.

Khalifa, K. (2013). Is Understanding Explanatory or Objectual? *Synthese*, 190(6):1153–1171.

Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge University Press.

Knuuttila, T. (2011). Modelling and Representing: An Artefactual Approach to Model-Based Representation. *Studies in History and Philosophy of Science, Part A*, 42(2):262–271.

Knuuttila, T. and Boon, M. (2011). How Do Models Give Us Knowledge? The Case of Carnot's Ideal Heat Engine. *European Journal for Philosophy of Science*, 1:309–334.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.

Massey, B. and Ward-Smith, A. (2006). *Mechanics of Fluids*. CRC Press.

McMullin, E. (1985). Galilean Idealization. *Studies in History and Philosophy of Science*, 16:247–273.

Mizrahi, M. (2012). Idealizations and Scientific Understanding. *Philosophical Studies*, 160(2):237–252.

Morrison, M. (1999). *Models and Mediators*. Cambridge University Press.

Morrison, M. (2009). Understanding in Physics and Biology. In de Regt, H., Leonelli, S., and Eigner, K., editors, *Scientific Understanding: Philosophical Perspectives*, pages 123–145. University of Pittsburgh Press.

Norton, J. (2012). Approximations and Idealizations: Why the Difference Matters. *Philosophy of Science*, 79:207–232.

Parker, G. and Smith, M. (1990). Optimality Theory in Evolutionary Biology. *Nature*, 348(6296):27–33.

Pincock, C. (2014). How to Avoid Inconsistent Idealizations. *Synthese*, 191:2957–2972.

Potochnik, A. (2007). Optimality Modeling and Explanatory Generality. *Philosophy of Science*, 74(5):680–6915.

Potochnik, A. (2009). Optimality Modeling in a Suboptimal World. *Biology and Philosophy*, 24(2):183–197.

Potochnik, A. (2010). Explanatory Independence and Epistemic Interdependence: A Case Study of the Optimality Approach. *The British Journal for the Philosophy of Science*, 61(1):213–233.

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

Rancourt, B. (2017). Better Understanding Through Falsehood. *Pacific Philosophical Quarterly*, 98:382–405.

Reiss, J. (2012). The Explanation Paradox. *Journal of Economic Methodology*, 19(1):43–62.

Reiss, J. (2013). The Explanation Paradox Redux. *Journal of Economic Methodology*, 20(3):280–292.

Rice, C. (2012). Optimality Explanations: A Plea for an Alternative Approach. *Biology & Philosophy*, 27:685–703.

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.

Rohwer, Y. and Rice, C. (2013). Hypothetical Pattern Idealization and Explanatory Models. *Philosophy of Science*, 80(3):334–355.

Rohwer, Y. and Rice, C. (2016). How are Models and Explanations Related? *Erkenntnis*, 81:1127–1148.

Schmid-Hempel, P., Kacelnik, A., and Houston, A. (1985). Honeybees Maximize Efficiency By Not Filling Their Crop. *Beahavioral Ecology and Sociobiology*, 17:61–66.

Shech, E. (2013). What Is the Paradox of Phase Transitions? *Philosophy of Science*, 80(5):1170–1181.

Shech, E. (2017). Idealizations, Essential Self-Adjointness, and Minimal Model Explanation in the Aharonov-Bohm Effect. *Synthese*, online first.

Strevens, M. (2008). *Depth. An Account of Scientific Explanation*. Harvard University Press.

Strevens, M. (2017). How Idealizations Provide Understanding. In Grimm, S. R., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, pages 37–48. Routledge.

Suárez, M. (2002). An Inferential Conception of Scientific Representation. *Philosophy of Scienc*, 71(5):767–779.

Suárez, M. (2010). Scientific Representation. *Philosophy Compass*, 5(1):91–101.

Sugden, R. (2013). How Fictional Accounts Can Explain. *Journal of Economic Methodology*, 20(3):237–243,.

Sullivan, E. and Khalifa, K. (2018). Idealizations and Understanding: Much Ado About Nothing? *Australian Journal of Philosophy*, forthcoming.

Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction, and Scientific Representation*. Palgrave-Macmillan.

van Riel, R. (2015). The Content of Model-Based Information. *Synthese*, 192(12):3839–3858.

Verreault-Julien, P. (2019). How could models possibly provide how-possibly explanations? *Studies in History and Philosophy of Science Part A*, 73:22–33.

Wayne, A. (2011). Expanding the Scope of Explanatory Idealization. *Philosophy of Science*, 78(5):830–841.

Weisberg, M. (2007). Three Kinds of Idealization. *The Journal of Philosophy*, 104(12):639–659.