Mirrors Without Warnings

Roman Frigg[1] and James Nguyen[2]
Forthcoming in *Synthese*

Abstract

Veritism, the position that truth is necessary for epistemic acceptability, seems to be in tension with the observation that much of our best science is not, strictly speaking, true when interpreted literally. This generates a paradox: (i) truth is necessary for epistemic acceptability; (ii) the claims of science have to be taken literally; (iii) much of what science produces is not literally true and yet it is acceptable. We frame Elgin's project in *True Enough* as being motivated by, and offering a particular resolution to, this paradox. We discuss the paradox with a particular focus on scientific models and argue that there is another resolution available which is compatible with retaining veritism: rejecting the idea that scientific models should be interpreted literally.

1. Introduction

Veritism is the position that truth is necessary for epistemic acceptability. Catherine Elgin's *True Enough* provides an extended argument against veritism and in favour of radically rethinking epistemology. She notes that veritism does not sit well with scientific practice because "science unabashedly relies on models, idealizations, and thought experiments that are known not to be true" (p. 1).[3] A clock is modelled as an ideal pendulum; a planet as a point mass; a population as infinite; and so on. A veritistic epistemology must regard such models as mere heuristics that have no place in the final edifice of science. Elgin rejects the dismissal of falsehoods as ultimately dispensable expedients. Rather than being a flaw, the divergence from the truth of many models and idealizations "fosters their epistemic functioning" (*ibid*.). This requires a revision of the central planks of traditional epistemology – belief, knowledge, assertion, and truth – which she aims to replace with notions that are not truth-directed: acceptance, profession, understanding, and what she calls *felicitous falsehoods*.

In this paper we examine the starting point of Elgin's project and ask whether making sense of the observation that science involves the use of models which, if interpreted literally, are false about their target systems really requires far-reaching revisions in epistemology, or whether this insight can be accommodated in another way. We show that the rejection of veritism rests on a literalist understanding of models, and we argue that such an understanding can

[1] r.p.frigg@lse.ac.uk Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science (LSE), London, UK and Centre for Philosophy of Natural and Social Science (CPNSS), LSE, London, UK.
[2] james.nguyen@sas.ac.uk, Department of Philosophy, University College London, London, UK, Institute of Philosophy, School of Advanced Study, University of London, London, UK, and CPNSS, LSE, London, UK.
[3] All references are to Elgin's (2017) unless indicated otherwise.

be resisted. In fact, one can retain veritism and give up literalism, and one can utilise Elgin's own account of scientific representation to this end. In doing so we highlight a so far underappreciated connection between the literature on veritism in the debates over scientific understanding and explanation – there usually referred to as the question of whether explanation and/or understanding is/are 'factive' (see, for example Doyle et al.'s (2018) and the references therein) – and the literature on scientific representation (see our (2016b, 2017a) and the references therein). We hope that by making this explicit we will encourage more work exploring the connections between various different positions across those debates.

We proceed as follows. In Section 2 we review the role that veritism plays in the philosophy of science, with a particular focus on how the notions of *truth, knowledge,* and *belief* are used to characterize the conceptual landscape. In Section 3 we offer a novel way to frame Elgin's project. Rather than taking the prevalence of idealisations in science – construed as falsehoods when interpreted literally – as a premise for an anti-veritist conclusion, we demonstrate that there is a more general paradox underlying her discussion. The paradox is in the form of three propositions, each of which seem prima facie plausible, but which are mutually inconsistent: (i) vertism, (ii) literalism, and (iii) the fact that many parts of science are epistemically acceptable and yet inaccurate when interpreted literally. Construed in this way, anyone who grants (iii) is required to give up on at least one of (i) and (ii), and thus take a stand on questions concerning explanation/understanding and/or scientific representation. We frame Elgin's project as offering one solution to the paradox: rejecting veritism. In Section 4 we discuss the paradox with special focus on scientific models and present an alternative solution: rejecting literalism. We argue that our preferred resolution to the paradox is in fact suggested by Elgin's own account of scientific representation (Chapter 12). In Section 5 we give reasons why the endorsement of non-literalism might be preferable to Elgin's rejection of veritism. Section 6 concludes.

2. Veritism in philosophy of science

Veritism is not only widespread in general epistemology (as Elgin documents in Chapter 2); it is also the background position of much of contemporary philosophy of science, where the notions of truth and belief chart the conceptual landscape in the scientific realism vs. anti-realism debate. According to an influential definition due to Psillos (1999), scientific realism is characterized by three "theses (or stances)" (1999, xvii). The *metaphysical stance* asserts that the world has a mind-independent structure. The *semantic stance* insists that we take scientific theories at face-value, seeing them as literal descriptions of their subject matter that can be true or false. The *epistemic stance* urges us to regard mature and predictively successful theories as (at least approximately) true.[4]

[4] This characterisation of scientific realism is widely used, as is documented in the article on scientific realism in the Stanford Encyclopedia of Philosophy (Chakravartty 2017).

The anti-realist gives up at least one of these commitments. The currently most prominent version of anti-realism is van Fraassen's constructive empiricism (1980). Constructive empiricism shares with realism the commitment to the semantic stance and insists that theories should be taken literally. It differs from realism in that the scope of appropriate belief recommended by the epistemic stance is restricted: we should only believe what a successful scientific theory tells us about the observable world (the theory's observable content), while we should remain agnostic about what it tells us about the unobservable world (the theory's unobservable content). So although both a theory's observable and unobservable content can be true or false, our epistemic commitment only concerns the former, and this commitment still takes the form of belief.

Selective realists take a different strategy. In an attempt to meet the challenge of the so-called "pessimistic meta-induction" they aim to isolate, or "select", those parts of a successful scientific theory that are, in some sense, essential in generating the observational content of the theory (Harker 2012). The hope is that it is these parts that are preserved across theory change: even if a mature and empirically successful theory is given up, whatever it was that ensured it was empirically successful in the first place will be preserved in its successor, and it is this part of the theory it is appropriate to believe. For structural realists this is the "structural content" of the theory (Worrall 1989); for entity realists this is commitment to the basic entities of the theory (Hacking 1983). Regardless, however the essential aspects of a theory are identified, these should go beyond the theory's observational content in order to distinguish selective realism from constructive empiricism.

For our current purposes it is crucial to note that it is still truth and belief that are the operating notions in defining the available positions in the debate. Indeed, different versions of scientific realism and the currently most influential version of anti-realism share a commitment to a literal understanding of scientific theories and a focus on truth and belief. The commitment to Psillos' semantic stance means that all parts of scientific theories are candidates for truth. The debate only turns on whether or not certain parts of scientific theories can be isolated as the appropriate candidates for belief, and on whether or not we can bracket off parts of scientific theories as those that are merely used as working posits.

Elgin's challenge goes right to the heart of this debate. If veritism is to be renounced (and with it a commitment to belief, knowledge, and assertion), then much of the debate over realism and anti-realism has got started on the wrong foot and we have to go back to the drawing board and re-evaluate theories in non-veritist terms. Such a re-evaluation is called for because science is rife with parts which "if interpreted as realistic representations of their referents […] are inaccurate in much the same way that false descriptions of an object are inaccurate" (p. 23). Elgin draws on a wide range of examples to illustrate this point, including the process of "curve smoothing" when turning raw data into a data model (p. 24), *ceteris paribus* laws, such as the law of gravity and Snell's law of refraction (p. 25), stylised facts in economics (p. 26), idealisations in the context of modelling (p. 27), and even Rawls' original position which models political citizens as mutually disinterested (pp. 27-28).

Starkly put then, the problem is the following: realists and anti-realists alike accept some form of the semantic stance, which requires interpreting the successful part of science literally, but many successful parts of science are false thus interpreted, and hence it is not clear whether, and if so how, the veritist approach can accommodate them. This is a serious challenge for anybody interested in scientific realism.

In the remainder of this paper we restrict our focus to model-based science. By this we mean cases where scientists describe a secondary surrogate system, which is then used to reason about the target (see Weisberg 2007).[5] There is little loss of generality in doing so. Thought experiments can be thought of as particular kinds of models (Salis and Frigg forthcoming), and models are usually the sorts of things that are assumed to contain the idealisations that Elgin is concerned about (Cartwright 1983). Our arguments *mutatis mutandis* carry over to other cases such as curve fitting and *ceteris paribus* laws.

3. The core contradiction

The problem Elgin identifies for a philosophical account of science can be presented as a paradox consisting of three individually attractive but jointly inconsistent propositions:

(i) *Veritism*: truth is necessary for epistemic acceptability.
(ii) *Literalism*: the claims of science have to be taken literally.
(iii) Much of what science produces is literally false and yet it is epistemically acceptable.

In both (i) and (iii) we adopt Elgin's use of the term "acceptable", i.e. "to accept that *p* involves being willing to take *p* as a premise, as a basis for action or […] as an epistemic norm or a rule of inference, when one's ends are cognitive" (p. 19). Framing Elgin's project as a paradox in this manner makes it clear how significant the issue is. These three propositions cannot be held together. So anyone who accepts (iii), which we take to be relatively unchallengeable for reasons we discuss below, will have to deny at least one of (i) or (ii). This underlying tension indicates just how important the contemporary discussions concerning whether or not understanding and/or explanation are factive are within the philosophy of science. Moreover, it demonstrates that these discussions are inextricably intertwined with the debates over scientific

[5] In recent years much work has been done attempting to characterize the ontological status of these objects, a discussion of which would take us too far afield here; for a recent review of these endeavours see Gelfert's (2017). For our current purposes it suffices to think of them as abstract objects that represent their targets in a manner that is analogous to the way in which concrete models such as the Phillips-Newlyn machine (Morgan and Boumans 2004) or ball-and-stick models of molecules (Toon 2011) represent their target systems.

representation – a fact that has not yet been fully appreciated in the literature. Our question, then, is to consider how the paradox is to be dissolved.[6]

The "eschatological" response to the paradox would be to give up on (iii). Although much of science seemingly relies on models that, if interpreted literally, are false, this falsity should not be taken seriously because false models are either peripheral or ephemeral: they are either located at the outer edges of our theoretical commitments, or they will be eliminated as science progresses in favour of representations that meet veritist standards.

Elgin dismisses this approach as untenable. Our scientific understanding is built on idealisations, and these cannot be dismissed as peripheral. The ideal gas law is central to our understanding of thermodynamics (p. 15); the Hardy-Weinberg model is central to our understanding of population genetics (p. 61); and so on. Likewise, there are no indications in current science that scientific progress involves the elimination of idealisations, and falsehoods more generally. Idealisation is not the hallmark of primitive science that gets eliminated as research progresses; our best current theories involve them as much as their predecessors did. Hence the elimination of falsehoods like idealisations is "neither necessary nor obviously desirable" (p. 31).

Even if this wasn't the case, and problematic models were to be eliminated in some future science, it remains unclear how this is supposed to help us deal with the success of current mature and predictively accurate parts of science. The realist's epistemic stance doesn't simply advocate that some ideal future science should be considered (approximately) true; the realist advocates that our current mature theories should be thus considered. Elgin's position is explicitly motivated by the requirement that a genuinely useful epistemology should be able to accommodate the success of our *current* science – with all its imperfections (if one were to classify idealisations as imperfections!) – not some hypothesized future science where everything is known about the world to arbitrarily detailed levels of precision (p. 31). Banking on idealisations dropping out of the picture in the future is not only unfounded in what we know about science; it also fails to do justice to the epistemic situation we are currently in.

We agree with Elgin that many of our best current theories contain representations which, if understood literally, are inaccurate in various ways, and that these representations are central to our understanding of the subject matter of these theories. We also agree that an appropriate epistemology and philosophy of science should be equipped to account for their cognitive success. So the eschatological response is untenable, and the contradiction has to be resolved by either renouncing veritism or literalism.

As we have previously seen, Elgin notes that models are inaccurate if they are interpreted as realistic representations of their targets, which provides the starting point for her project. So she accepts literalism and rejects veritism

[6] An anonymous referee notes that Giere's (2009) contribution to Suárez's (2009) and Suárez's (2010) provide related discussions regarding whether (iii) is compatible with scientific realism.

(although as we discuss in Section 5, her own account of scientific representation is compatible with a rejection of literalism). This is the starting point of a programme that aims to rehabilitate the use of falsehoods in science by de-emphasising truth (and its related notions of belief and assertion) and assigning non-factive understanding and felicitous falsehoods centre stage in the epistemology of science.

It bears noting that Elgin does not restrict her focus on the benign cases, which, even though strictly false, are still approximately true. Her claim is more radical: not all false representations are approximately true, and even representations that are radically false have epistemic value.[7] Her examples for such representations are the Hardy-Weinberg equation in population genetics, which assumes that a population is infinite in order to screen off genetic drift, and Rawls' original position, which assumes that agents are mutually disinterested and behind a veil of ignorance (p. 29). These are no exception. One might add Ising's model of ferromagnetism, Kac's model of macroscopic irreversibility in the presence of reversible micro dynamics, classical models of quantum scarring, Schelling's model of social segregation, and Ackerlof's market for lemons to the list.

The challenge is to account for what the value of such representations is. To meet this challenge Elgin develops an intricate and finely calibrated epistemology that excises the traditional epistemic concepts. It replaces the notion *belief* with that of *acceptance*, the notion of *assertion* with *profession*, the notion of *knowledge* with *understanding*, and it explains the value of false representations in terms of them being "felicitous falsehoods". As noted above, to accept a proposition *p* involves "being willing to take *p* as a premise, as a basis for action or, [...] as an epistemic norm or a rule of inference, when one's ends are cognitive" (p. 19). Relatedly, given the connection between belief and assertion, professing 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" (p. 21). A representation is a felicitous falsehood if it is "an inaccurate representation whose inaccuracy does not undermine its epistemic function" (p. 3). The epistemic function of models can involve affording epistemic access to a representation's object (p. 20), serving as a fruitful working hypotheses (*ibid.*), and, first and foremost, contributing to the understanding that science supplies (p. 1). Understanding does not concern isolated claims, but "a topic, discipline, or subject matter" (p. 43). Understanding is holistic, and understanding a particular matter of fact therefore derives from being able to place this fact into theoretical context. Having an understanding of such a context means to have "an epistemic commitment to a comprehensive, systematically linked body of information that is grounded in fact, is duly responsive to reasons or evidence, and enables nontrivial inference, argument, and perhaps action regarding the topic the information pertains to" (p. 44).[8]

[7] Elgin does not discuss the well-known issues surrounding how to define approximate truth in the first place (Oddie 2016). However, she does seem to assume, and we agree with her, that any notion of approximate truth will not be appropriate to capture the epistemic value of at least some models (when interpreted literally).

[8] Understanding has several dimensions. See Baumberger and Brun's (2017) for an analysis.

Let us illustrate this approach with one of Elgin's own examples. Suppose we use the Hardy–Weinberg model to understand how gene frequencies change during evolutionary processes. The model works by considering an infinite population reproducing at random where two alleles *A* and *a* compete at a single locus. For a particular generation, *i,* we let G^i_{AA} be the proportion of *AA* individuals, and likewise for G^i_{aa} and G^i_{Aa}. We can then define the proportion of alleles in that generation as $p = G^i{}_{AA} + 0.5G^i{}_{aA}$ for *A* and $q = G^i{}_{aa} + 0.5G^i{}_{aA}$ for *a*. Using this one can compute the expected values of genotypes for the following generation:

$$G^{i+1}{}_{AA} = p^2,$$
$$G^{i+1}{}_{aA} = 2pq,$$
$$G^{i+1}{}_{aa} = q^2$$

For this result to hold, it is crucial that the population is infinite (to avoid "genetic drift", i.e. a random event which just so happens to disproportionally affect the frequency of one of the alleles), and also that mating is random, that there is no mutation, that there is no fitness benefit for any genotype, as well as a few other assumptions (for details see Templeton's (2006, Chapter 2)). It's plausible that many of the populations of interest do not exhibit any of these characteristics (and none are infinite!). And yet the model plays a central role in our understanding of gene frequencies in a population and their evolution.

On Elgin's approach this makes the model a "felicitous falsehood":

> The model is no approximation. Populations are not nearly infinite (whatever that might mean). Mating is not nearly random. However indiscriminate actual mating behavior is, physical proximity is required. In the long run, mating only with nearby partners promotes genetic drift. Natural selection and genetic drift are ubiquitous. Migration is widespread. Mutation and random fluctuations are, in real life, unavoidable. Still, to understand the effects of evolution, it is useful to consider what would happen in its absence. By devising and deploying an epistemically felicitous falsehood, biologists find out (p. 29).

Furthermore,

> [i]nasmuch as evolutionary pressures are always present, the model cannot, nor does it pretend to, account for allele distribution generally. It is, however, very useful for some purposes. If population geneticists want to understand how significant an evolutionary factor such as migration is, they need a base rate. They need, that is, to know how alleles would redistribute in its absence (p. 263).

So the model provides understanding concerning the relationships between each of these features and the role they play in determining the allele distribution across a population. Understanding, say, the role a particular historical event (e.g. how a population became isolated into two subpopulations) played in shaping the allele frequencies in a population might require understanding the various interconnections between each of these features. The Hardy–Weinberg model provides the base case to which divergences can be compared. It contributes to how population geneticists can try to isolate the effect of each feature, and thus links them together into a holistic web of understanding. The fact that it contains an infinite number of organisms is vital for the model to generate the results that it does, and yet it also makes the model false. Scientists using the model can accept it for a certain theoretical investigation (to understand the role some other feature played, say) without believing it. When they apply it, they don't assert that it holds of the target but they might profess that it does. And the model's falsity doesn't stop it playing an important epistemic function, even if, when applied to an actual target system, it doesn't provide "knowledge" (at least if interpreted literally as representing the actual population as being infinite). This function contributes to our understanding in the sense that the model illuminates the connections between different effects on allele distributions.

This short sketch cannot do justice to the complexity and depth of Elgin's epistemology, nor can it shed light on its implications for important issues such a holism, nonfactivism, epistemic normativity, and reflective equilibrium. These topics will have to be dealt with in future in-depth discussion of Elgin's epistemology. Our aim here is a different one. We aim to shed light on the motivations for her project, in particular the connection between these far-reaching reconfigurations of epistemology and the underlying assumptions concerning how scientific models, which are false if interpreted literally, represent. To this end we now investigate the third option of resolving the above contradiction, namely to reject literalism.

4. Rejecting literalism

Scientific models, which play an important role in Elgin's rejection of veritivism, are in need of interpretation before they can be said to be false or inaccurate representations. The question then, is how are such models interpreted? In this section we outline Elgin's own account of scientific representation (Chapter 12), and then show how it can be brought to bear in the context of idealised models.[9] We then argue that even by Elgin's own lights, idealised models do not have to be interpreted as false or inaccurate representations of their targets. Thus, somewhat surprisingly, her own account of scientific representation can actually be utilized to motivate a rejection of literalism in a manner which dissolves the core contradiction of the previous section, thereby undermining the main motivation for anti-veritivism.

[9] We note that her account of representation in *True Enough* is continuous with at least some of her earlier work on representation (Elgin 2009, 2010). It is an interesting exegetic question how it relates to her even earlier work, which places less of an emphasis on the idea that science is a model-based practice. Unfortunately space considerations prevent us from discussing the historical development of her thought.

As noted in Section 2, modelling involves introducing a secondary system, the model, which is then used to reason about the target. Models can be the vehicles of surrogative reasoning due to the fact that models represent their target systems. What motivates the core contradiction drawn out of Elgin's work is that models are central to our understanding of their target systems, but nevertheless are not (approximately) "true" or "accurate" representations of their targets.[10] But what does it mean to call a model an inaccurate representation? To answer this question we need to establish what makes a model accurate or inaccurate, and indeed what makes a model a representation in the first place. Considered as "bare" objects, model systems (such as a system of water pipes or a collection of perfect spheres) have no representational properties. They are just objects, either concrete or abstract. The challenge then, is to account for what it takes to turn a model system into a representation of a target, and how this suffices to understand the representational accuracy of a given model.

Thinking about models as representations in these terms, "literalism" can be associated with accounts of scientific representation according to which models are "intended copies" of their targets, at least in certain respects (which can include structural ones) and to certain degrees. Among the accounts of representation that enshrine this idea are similarity and structural accounts of representation.[11] Similarity accounts are based on the notion that a model accurately represents its target by being similar to it, where similarity is explained in terms of model and target sharing certain features. Structural accounts are based on the same idea, but further narrow the focus on structural features. *Literalism* is then the claim that models have to be interpreted as sharing features with their targets in order to be accurate representations of those features. *Veritism* is the claim that only models that accurately represent features of their targets are epistemically acceptable. In this context, premise (iii) is the claim that there are many scientific models that are epistemically acceptable despite having features that play an essential role in defining the model but which are nevertheless explicitly not shared with their targets. The population in the Hardy-Weinberg model is infinite, as is the lattice in the Ising model. These features are not shared with the target (which is finite) and yet the models are acceptable. In these terms, retaining veritism and (iii) requires an account of scientific representation that allows for non-literal interpretations of scientific models.

Elgin herself (Chapter 12) provides an account of scientific representation that she takes to be a literalist account, and as result the models she focuses on come out as falsehoods. Since we are going to argue that Elgin's account can be re-interpreted as a non-literalist account that can in fact be utilised to undercut her own motivation for giving up on veritism, it is worth spelling it out in detail here.

---

[10] We use the terms "accurate" and "inaccurate" rather than "truth" or "falsity" when talking about model-based science. This is in line with much of the literature where models are, strictly speaking, not considered truth-bearers since they are non-linguistic. For opposing views see Mäki (2011), who thinks that models should be understood as truth-bears despite being non-linguistic, and Toon (2012) and Levy (2015) who understand models linguistically.

[11] Here is not the place to delve into the details of these accounts. For reviews and discussions see our (2016b, 2017a).

According to her, we should think of scientific representation in terms of "representation-as".[12] A representation *x*, e.g. a model, a picture, a caricature, represents a target *y*, e.g. a magnet, a landscape, a politician, as *z,* e.g. as undergoing a phase transition, as foreboding, as a bulldog. Elgin's account of representation-as has three components: *denotation*, *exemplification*, and *imputation*.

Denotation establishes representation-of: symbol *x* is a representation-of *y* iff *x* denotes *y*. Denotation is the relation that holds between a name and its bearer and between a predicate and the objects in its extension. Likewise "[p]ictures, equations, graphs, charts, and maps represent their subjects by denoting them" (p. 251). This extends to scientific models: what makes models "about" their targets in the first place is that they denote them. It's worth noting here that in order for a model to denote a target, the latter must exist. This raises the problem of models with non-existent targets. Examples of such models can be drawn from the history of science – models of the ether, caloric, or phlogiston – as well as from current investigations where some models are introduced without the aim of representing any system in the world – Norton's Dome (Norton 2003) or *n*-sex models for $n>2$ (Weisberg 2013). The latter are better understood as facilitating reasoning about the concepts of the theories in which they are embedded rather than any system in the world. Elgin's general theory of representation has a way of accommodating such models. Just as we classify paintings depicting non-existent animals or landscapes as griffin-representations or minotaur-representations despite the fact they don't denote anything, we can classify caloric models as caloric-representations or *n*-sex models as *n*-sex-population-representations despite containing more than two sexes of organisms (p. 252). In general, being a *z*-representation does not entail being a representation-of a *z*. These *can* come together. For example, a 2-sex-population-representation can also be a representation-of an actual population. But they can also come apart in different ways. A caloric-representation is not a representation-of caloric (because caloric doesn't exist), and the name "Catherine" is a representation-of a person despite not being a person-representation.

Denotation alone is not sufficient for the sort of relationship that holds between models and their targets, or paintings and their subjects, and so on. In general, what we call "epistemic representations" can be used to (attempt to) learn about their targets. By reasoning about certain features of the representation, one can draw inferences (which may be false) about the target system.[13] That *x* denotes *y* does not, by itself, accommodate this sort of relationship; reasoning about

[12] We note here that 'representation-as' has also been employed by Hughes (1997) and van Fraassen (2008) in discussions of scientific representation. For the latter in particular, it is an open question about how he accommodates discussions of idealisation and representation-as within his broadly structural understanding of representation. Whilst he is sensitive to the role of distortions in model-based science (2008, Chapter 1), he is not explicit about the role that these play in establishing structural relationships between the structure of the phenomena and the structure of the model (2008, Chapter 11).

[13] We include the parenthetical "(attempt to)" and "(which may be false)" to allow that representations may *mis*represent. For more on this distinction see Section 1 of our (2016b).

features of a name or a predicate does not allow someone to draw any inferences about the bearer of that name or predicate.

Moreover, notice that when we do draw inferences using models, or epistemic representations more generally, these inferences do not concern all features of the representation. The fact that the population in the Hardy-Weinberg model is infinite is relevant to which inferences it licenses, the fact that we distinguish between alleles using upper case and lower case letters, rather than some other lexicographic scheme, is not. So, in addition to denotation, we need to account for the facts (a) that certain features of the model systems are the relevant ones and (b) that by reasoning about these features a model user can generate inferences concerning the subject of the representation.

To accommodate these aspects of scientific representation Elgin invokes the notion of *exemplification*, which is a relation that holds between an exemplar and the property or feature that it is an exemplar of. An exemplar "functions as a symbol that makes reference to some of the properties, patterns, or relations it instantiates" (p. 184; see also Goodman 1976 and Elgin 1996). Samples are a simple example of items that represent via exemplification. The pieces of cloth in the tailor's fabric sample book exemplify their colour, shine, and texture, and the olive we try at the market exemplifies its flavour. Following Elgin we use the phrase "exemplified feature" to refer to any property, relation, or pattern that an exemplar instantiates and refers to. There are no restrictions on these features; they can be "static or dynamic, monadic or relational, and may be at any level of generality or abstraction" (p. 185). There are two conditions to be met for an object *x* to exemplify a feature *F*: *x* must instantiate *F*, and *x* must additionally refer to *F*. A crucial aspect of exemplars is that they afford epistemic access to the features they exemplify: "By exemplifying a feature – by highlighting or displaying that feature – an exemplar affords epistemic access to it. Someone who properly interprets the exemplar is in a position to recognize the feature in question" (p. 260). We can see this in action in the way in which fabric samples exemplify their textures (amongst other features). The sample instantiates its texture, and it does so in such a way that the texture is highlighted and referred back to. Because the sample exemplifies its texture we can use it to reason about other things which also instantiate that feature, e.g. a piece of clothing made out of the type of fabric in question.

Tying these aspects of exemplification together with denotation leads to Elgin's account of representation-as:

> when *x* represents *y* as *z*, *x* is a *z*-representation that *as such* denotes *y*. We are now in a position to cash out the 'as such'. It is *because x* is a *z*-representation that *x* denotes *y* as it does. *x* does not merely happen to denote *y* and happen to be a *z*-representation. Rather, in being a *z*-representation, *x* exemplifies certain properties and imputes those properties or related ones to *y* (p. 260, original emphasis).

This account applies to models, and it accounts for (a) and (b) above. What makes a feature of a model relevant when it comes to generating inferences about its target system is that the model exemplifies that feature. And what explains how a model can be used to generate an inference about its target is that a model user gains epistemic access to the exemplified features, and then imputes those features to the target system by taking the target to be in the extension of the exemplified feature. In Elgin's words: "models exemplify features they share with their targets and impute those features to their targets. Where they are successful, they afford epistemic access to aspects of their targets that we might otherwise miss" (p. 262).

Understood in this way, we can see why Elgin's account of representation-as paves the way for a literal interpretation of scientific models: models are understood as representing their targets as having certain features that the models themselves exemplify (and therefore instantiate). In this sense, despite Elgin's arguments against resemblance views of representation, her account could be seen as a developed version of it, and she notes this possibility herself:

> for *x* to exemplify a property of *y*, *x* must share that property with *y*. So *x* and *y* must be alike in respect of that property. It might seem, then, that resemblance in particular respects is what is required to connect a representation with its referent [...]. There is a grain of truth here. If exemplification is the vehicle for representation-as, the representation and its object resemble one another in respect of the exemplified properties (p. 261-62).

It is this interpretation of her account that underlies a literalist interpretation of models according to which a model that has a relevant property that is not shared with the target is a falsehood.

However, even when discussing this account, Elgin is sensitive to the idea that target systems needn't instantiate the *exact* feature exemplified by the model in order for the model to be successful. In fact, the definition of representation-as quoted above states that *x* imputes exemplified features *or related ones* to *y*, and she emphasizes that "or relates ones" is crucial (p. 260). In benign cases the features instantiated in the target are not off by much and the divergence is negligible (p. 261). In other cases the discrepancy is more significant. Elgin provides the following example:

> A caricature that exaggerates the size of its subject's nose need not impute an enormous nose to its subject. By exemplifying the size of the nose, it focuses attention, thereby orienting its audience to the way the subject's nose dominates his face or, through a chain of reference, the way his nosiness dominates his character (p. 260).

In this example the caricature exemplifies the feature of having a large nose. But it's explicitly not this feature that is imputed to the target. Rather the feature is connected, via a "chain of reference", to a "related one", nosiness as a character trait, and it's this trait, not the feature literally exemplified by the caricature, which is imputed onto the target.

The same happens with scientific models, which can exemplify features that are not themselves imputed to the target. The question then is whether or not we can further explicate what it means for a model that exemplifies some feature *F* to impute a feature G, which is related to F, to the target. Relations can be multifarious and so there is no general account of what "related ones" means. However, it would seem important to give the possibility of imputing a related property G a clearly demarcated space in an account of representation. For this reason we incorporate into our DEKI account the requirement that scientific representations need to come with *keys* which explicitly specify the connection between the feature exemplified by a model and the features to be imputed to a target system. A key essentially is a rule of association that correlates every exemplified feature F of the model with a feature G that is imputed to the target. We can then say that model M is z-representation of target T iff the following conditions hold (Frigg and Nguyen 2018, p. 220):[14]

1. M denotes T.
2. M exemplifies z-properties $F_1, \ldots, F_n$.
3. M comes with key K associating the set $\{F_1, \ldots, F_n\}$ with a set of properties $\{G_1, \ldots, G_m\}$.
4. M imputes at least one of the properties $G_1, \ldots, G_m$ to T.

In some cases these keys might be the identity (which would amount to a literal interpretation of a model, i.e. a model which exemplifies some feature F and represents its target as having F), but in other cases the imputed features can diverge from the exemplified features, which would be a non-literal way of interpreting the model. Non-literal interpretations are common. Consider a map of the world. It exemplifies a distance of (roughly) 40cm between the points labelled "Chicago" and "London". The cartographic key the map comes with includes a scale – 1 mm : 10 miles – which allows us to translate a feature exemplified by the map to a feature of the world: that there is a distance of (roughly) 4000 miles between Chicago and London. Someone who imputes to the world the feature that Chicago and London are 40cm apart simply doesn't understand how a map works.

The importance of keys can also be seen with another every-day object: a car's wing mirror. Mirrors provide epistemic representations. In fact mirrors are paradigmatic examples of such representations. By reasoning about the features of the mirror image we can reason about the features of the object shown in the mirror. In the case of car's wing mirror we can learn about the cars behind us, and about the traffic situation to the side and rear more generally. The image denotes the relevant area, and exemplifies various features: whether or not there is a car in the mirror, the distance between the car in the mirror and the side of door panel in the mirror, and so on. But a wing mirror is a convex mirror, and unlike the straight mirrors that one usually finds in bathrooms, such mirrors make objects appear smaller than they are. If one looks in the wing mirror and infers that the car behind is as far away as it appears in the mirror, one makes a fatal error. The car is in fact much closer!

[14] For more on the DEKI account and the relationship it shares with Goodman and Elgin's account of representation see our (2017b).

Car manufacturers are worried that drivers make the interpretative mistake to think that the car mirror works just like their bathroom mirror and therefore often deliver mirrors engraved with the warning "objects in mirror are closer than they appear". On our account of scientific representation this warning works as a *key*. The key tells us to translate the exemplified distance features in the mirror image into another distance feature to be imputed to the actual situation on the road. In fact, the mirror's warning explicitly councils the interpreter *against* a literal interpretation of the epistemic representation. And in doing so allows for a translation between the features exemplified by the representation and the features imputed to the target. Once this is understood the image is no longer a felicitous falsehood, and it can be understood as a literally accurate representation of the target (but only if interpreted non-literally).

Models need to be keyed up in the same way. Consider the case of a scale model of a boat being used to represent the forces an actual boat faces when moving at sea. The exemplified feature in this instance is the resistance the model boat faces when dragged through a tank of water. But this doesn't translate into the water resistance faced by the actual boat in a straightforward manner. The resistance of the model and the resistance of the real boat stand in a complicated non-linear relationship because smaller objects encounter disproportionate effects due to the viscosity of the fluid. The exact form of the key is often highly non-trivial and emerges as the result of a thoroughgoing study of the situation. Other models work with limit keys. The Ising model of a ferromagnetic substance instantiates infinity (in the sense that it's an infinitely extended lattice). But we don't impute this feature to the magnet; we impute being large enough for boundary effects to be negligible. When using the Hardy-Weinberg model we needn't impute being an infinite population mating at random to the target, we can impute being large enough for effects of genetic drift to be negligible, which implies that we should expect the genotype distribution to be around (but not necessarily exactly) the one given by the equations in Section 3. When using Schelling's model to target social segregation we don't impute being set on a grid with no cost of moving to, say, residential patterns in Chicago; we impute the claim that even a relatively "low" preferences regarding how many of one's neighbours one wants to be similar to oneself can yield global segregation. And so on. Determining how to move from features exemplified by models to features of their target systems can be a significant task, and should not go unrecognized in an account of scientific representation, or indeed scientific practice generally.

There is one big difference between wing mirrors and models: models, even though they require a key, rarely, if ever, come with warnings written on their sleeves. Models are, as it were, mirrors without warnings! It may be a matter of some compunction that models usually don't come with explicit keys, and one might wish that scientists were more explicit about them. However, not being made explicit should not be mistaken for being absent. No one with any understanding of mechanics thinks that a Newtonian model imputes "being a point particle" to the sun; nor need an electromagnetic model impute backward causation to a moving charge. The key is often implicit in a scientific practice, and students learn how to interpret models on the job. Yet, an implicit key is still a key.

It now transpires how the introduction of the key serves to allow for non-literal interpretations of scientific models, and thus how it dissolves the core contradiction. Yes, scientific models are idealised representations of their target systems. And yes, if interpreted literally they deliver falsehoods which are in tension with the idea that they offer us understanding of their subject matter, at least if understanding is construed factively. This puts pressure on an epistemology and philosophy of science that takes truth as a necessary condition for epistemic acceptability. But rather than give up on veritism, one can instead give up on the idea that models have to be interpreted literally. Just because a model exemplifies an infinite number of individuals, this feature does not have to be interpreted as holding in the actual population. Just because the idealised pendulum exemplifies being subject to no air resistance or friction, it needn't be taken to represent the actual pendulum in the grandfather clock in your office as being subject to no air resistance or friction. Rather, it represents the clock mechanism as being such that the effect of those aspects is negligible for ordinary time-keeping. In this way one can salvage veritism at the expense of literalism.

An upshot of this discussion is to emphasise that the position we have arrived at is itself motivated by something very close to Elgin's own account of scientific representation. In order for her account of scientific representation to motivate her anti-veritism, it better be the case that the sorts of models she uses to motivate such a position, the Hardy-Weinberg model and so on, come out as falsehoods (albeit felicitous ones). And this requires that they be interpreted literally. To the best of our knowledge, Elgin doesn't motivate why this should be the case (as we have seen, it doesn't follow from her account of scientific representation). Her account of representation thus leaves two ways of accommodating the use of idealisations in science: either reject literalism but retain veritism, or accept literalism and reject veritism. Elgin herself gives us no reason to prefer the latter approach. This leaves a lacuna in her account that deserves to be addressed, and as we argue in the next section, there are good reasons to prefer the former option.

The roots of non-literal thinking go back at least to Aristotle's discussion of analogy, and non-literalism has become a prevalent doctrine in philosophy of science when the protagonists of the Vienna circle insisted that theoretical terms had to be defined in terms of observables. In the more recent literature on models Hesse (1963) construes models as *analogies* and Black (1962) talks about models as *metaphors* (for a discussion of both views see Bailer-Jones's (2002)). Giere (2004), Teller (2001) and Weisberg (2013) characterise the model-world relation in terms of similarity, and it has become increasingly popular to think of scientific models as analogous to works of fiction; see e.g.

Bokulich's (2009, 2012, 2016), Frigg's (2010), Toon's (2010), and Levy's (2015), as well as the essays collected in Suárez's (2009) and Woods' (2010).[15]

While these approaches have the potential to go some way to accommodating model-target mismatches, it not clear how far they actually go. Hesse explicates analogy in terms of shared features (1963, 8). Black's theory of metaphor still insists that a successful model must be "isomorphic with its domain of application" (1962, p. 238). Weisberg is explicit that similarity amounts to sharing properties (Weisberg 2013, Chapter 8), while Giere and Teller offer no explicit analysis of similarity. The various discussions scientific models in fictional terms intertwine the functional and ontological status of models, with no clear sense of non-literalism emerging from those who focus on model functions. Bokulich's (2016) discussion of how non-verdical fictional models provide understanding suggests that at least one reason to think that models are fictions is to interpret them literally; and Levy's (2015) analysis of how fictional models relate to the world is based on the notion of partial truth, the idea that a statement is partially true "if it is true when evaluated only relative to a subset of the circumstances that make up its subject matter" (*ibid*., p. 792). So upon closer inspection these accounts offer less to the non-literalist than she would have hoped for.

Proponents of these accounts can insist that the accounts have at least the *potential* to be developed in a non-literal way. The similarity account can do this, for instance, by explicating similarity at the level of the features themselves, rather than requiring that models and their targets share the exact same features (Khosrowi forthcoming). We agree, but the fact remains that in the current form none of them provides anything like a full-fledged account of how non-literal interpretations of models can accommodate the sort of idealisations Elgin uses to motivate her non-veritism. So at least in its current instantiations, non-literalism remains little more than promissory note. What is needed is an explicit formulation of a non-literal account of representation that can account for the kinds of examples that we encounter in contemporary model based science. We do not maintain that DEKI has invented non-literalism; but we do maintain that it is the only bona-fide attempt in the current discussion about scientific representation to develop an account that is upfront and explicit about its non-literal commitments and assigns these a systematic place in its theoretical architecture.[16, 17]

---

[15] We are grateful to an anonymous referee for encouraging us to think about discussions of models and fiction in this way.

[16] Another account of scientific representation, primarily associated with Suárez (2004, 2015) could also be brought to bear on this issue. This "inferential account" focuses on the inferences about target systems by scientific models, and these inferences don't have to be underpinned by a literal understanding of the models. According to this account the idealised models we're concerned with here could be used in way that instructs informed agents to only draw inferences with true conclusions regarding the target from them (retaining veritism), but do so by connecting them to the explicit model-target mismatches (giving up on literalism). To the best of our knowledge there has been no explicit investigation yet about how this could be done in the inferential framework, but it does seem like a fruitful avenue of research.

Finally then, before moving on, it is worth briefly recapping how what we have said here relates to the discussion of realism vs. anti-realism in Section 2. The crucial thing to note is that our account of non-literal interpretations cuts across any of the distinctions drawn in that debate. The account enjoys considerable latitude both in terms of the scope and the nature of the keys, and can thereby accommodate realist and anti-realist positions alike. As regards scope, an anti-realist of the constructive empiricist stripe who accepts the semantic stance but denies the epistemic one, could adopt DEKI with a literalist key and only believe in claims about observable features of the target system. An anti-realist who also denies the semantic stance could adopt DEKI and only key-up observables features to be imputed to targets. A selective realist could only believe, or more radically only key-up, the particular kind of features they are selective about. The realist may have no restrictions whatsoever about the type of features they consider in their *F*s and *G*s. As regards the nature of keys, the use of non-identity keys in no way implies an anti-realist position. Observable as well as unobservable parts of the model can be keyed-up with a non-identity key and yet one can take the resulting claims at face value. So adopting DEKI does not prejudice what position one has to take in the realism vs. anti-realism debate.

We thus hope to have provided an account of scientific representation that is neutral across realist/anti-realism spectrum, and moreover can be utilised by those who are sensitive to the fact that our best scientific models offer idealised representations of many aspects or features of the external world.

5. Veritism vs. Literalism

So far we have discussed two "packages", two ways of dealing with the paradox outlined in Section 3. Elgin's package is to renounce veritism and retain literalism; ours is to renounce literalism and retain veritism.[18] This raises the obvious question whether there are reasons to prefer either of the packages. Unfortunately there are no easy answers: both options have widespread implications across the philosophy of science and epistemology. They're wholesale views about scientific practice and the nature of knowledge and understanding, and choosing between them seems to us to be the sort of thing which depends on how one values a package as a whole, rather than the sort of thing which can be broken down into neat little reasons to prefer one to the other.

Having said that, we do want to provide some motivations for why the package we have offered should be considered attractive. First, as demonstrated by our simple examples of epistemic representations above, namely the map and wing mirror in Section 4, and indeed Elgin's own example of a caricature, the value of

---

[17] Additionally, as discussed in Section 5, Strevens (2008) provides a discussion of how idealisations work, which, whilst not couched in these terms, could be understood in terms of providing a non-literal account of how they represent.

[18] We also note the possibility of rejecting both veritivism and literalism. To the best of our knowledge no one in the debate does this.

many of these representations is not accounted for by a literal interpretation.[19] Such examples can be multiplied with ease. The colour of a line on a tube map doesn't represent the tube line itself as being coloured (it, non-literally, represents which particular line it is) anymore than the litmus paper turning red represents the solution it is dipped in as red (it, non-literally, represents it as acidic). Moreover the introduction of a *key* into an account of these representations illuminates the role that things like disciplinary practices and interpretational schemes (cartographic traditions, artistic styles, and so on), play in our use of such objects to represent the world. By focusing our attention on the keys, and the practices in which they are embedded (in addition to, rather than in place of, the objects doing the representing themselves), we gain a better understanding of how these epistemic representations work.

Secondly, the same observation applies in the context of scientific modelling: we think that our non-literalism plus veritism package better accounts for scientific practice. As argued above, keys seem to be operating implicitly in many cases of scientific modelling. Engineers who use scale models to study the resistance of a ship when dragged through water will not assume that the model represents the real ship literally, i.e. as having the same resistance, not even approximately. They will use complicated scaling relations to derive the resistance of the real ship from the resistance of the scale model, and these relations constitute the key of the model. Similarly, no one infers from the use of the two-dimensional Ising model in statistical mechanics that samples of K_2NF_4 or RB_2MnF_4 are two-dimensional, or consist of an infinite number of atoms. Rather they will infer that both substances have strong horizontal and weak vertical interactions, and that boundary effects will be small compared to the contributions of the "mid grid" atoms (Baxter 1982).[20]

Third, our package has the benefit of allowing us to retain much of the philosophy of science (and epistemology) based on veritism. One of the reasons why Elgin's project is so impressive is its broad scope and wide-reaching implications. As discussed above, the standard way of understanding the scientific realism vs. anti-realism debate is framed in terms of truth, knowledge, and belief. Adopting Elgin's package would require reframing the debate itself. And this isn't restricted to that particular debate. In the philosophy of science veritist assumptions play central roles in our understanding of various positions in, for example, debates concerning the nature of confirmation, induction, underdetermination, and the role of theoretical virtues in theory choice. And the point is not restricted to philosophy of science. As Elgin herself points out, "philosophy valorises truth" (p. 1), and in epistemology "the conviction that truth is vital is virtually axiomatic" and "veritistic commitments run deep. Abandoning them requires radical revisions" (p. 9). Because of how central veritism is to

[19] It's essential that the wing mirror gives information about the distance of the vehicle in the back, and a map must tell us how far apart places are. Under a literal interpretation, Elgin's caricature comes out as a felicitous falsehood: it generates understanding but not factive information about its subject. It seems more natural to say that the caricature is simply an accurate representation of the subject's character.

[20] Nguyen (forthcoming) also discusses how giving up on literalism can help account for the prevalence of the use of so-called 'toy models' across the sciences.

much of philosophy (of science and beyond), to see Elgin's project through would require radical reformulation of philosophical positions across the board. Such radical revisions seem justified only if there are compelling positive reasons that such a reformulation is unavoidable. If there were a necessity for a revolution we would welcome it. But rebels need causes. And, as we argue above, Elgin's discussion of representation and idealisation in science leaves open an alternative option: reject literalism instead. This option doesn't require the same sort of radical philosophical reformulation, and thus at least prima facie, seems to be the more economical way of proceeding.[21]

Fourthly, and finally, in arguing that Elgin's project highlights a paradox that should worry anyone who accepts that idealisations play a significant epistemic role across contemporary (and historical) science, we have highlighted that the debates concerning understanding/explanation and scientific representation have to be addressed in tandem. And, as should be expected given this connection, there are echoes of our non-literalist suggestion already present in the former, in particular in Strevens' (2008, Chapter 8; 2017) treatment of idealisation. Obviously this is not the place to embark on a detailed discussion of his "kairetic" account of explanation, but it is worth drawing attention to the fact that his approach, understood as a non-literal way of accounting for the epistemic value of idealised scientific models, provides evidence that non-literalism is an attractive position.

Strevens argues that idealisations, such as the assumption that there are no interactions between particles in a gas which is used in the derivation of the relationship between pressure and volume as stated in Boyle's Law, should be interpreted non-literally. More precisely, he suggests that when a model contains a parameter that is set to zero, it needn't be interpreted as representing the feature in the target system corresponding to that parameter to be zero, but rather it can be interpreted as representing that feature as being irrelevant to the behaviour represented by the model (within a certain range of scenarios). So when we interpret the "there are no interactions" feature of the model as representing the feature of the target that "the interactions aren't difference makers with respect to the relationship between volume and pressure" we arrive at an accurate representation (at least in the appropriate regimes, i.e. where the volume of the container is large enough). Thus, we have saved veritivism at the expense of literalism.

Whilst this indicates the sorts of possibilities allowed by non-literalism, it does bear noting that there is an important difference between the DEKI approach and Strevens' treatment. The latter requires that the idealised aspect of the model – the parameter that is set to some extremal, obviously false if interpreted literally, value – represents the corresponding feature of the target system as not being a difference-maker. Understood through the lens of DEKI, this is a *particular* kind of non-literal key. There is, however, a question whether this approach captures all idealised representations. Streven's account can only be

[21] Having said that we are inclined to see the revolutionary aspect of Elgin's project as a positive for its own sake, at least in the sense that it widens the logical space of philosophy.

applied in cases where the idealisation takes the form of setting a parameter to an extremal value. There are, however, cases of idealisation in which this is not the case. The distortions present in the mirrors with warnings, maps, and e.g. the use of model ships discussed above, are not of this form, and it is not obvious how Strevens' account of idealisation would deal with them. The DEKI framework allows for different keys in these cases, and, possibly, keys according to which the idealisations can play a more positive representational role rather than just flagging features as not being difference-makers.[22] Regardless of the details of how one goes about dealing with these sorts of idealisations, it should be clear how the DEKI account of scientific representation provides a more general and systemic framework in which to investigate how non-literal keys can be utilised to save veritism.

Now, we are not arguing that these observations are knockdown arguments for why one should reject literalism rather than veritism when faced with the core contradiction of Section 3. But taken together we hope that they are persuasive, and, perhaps more importantly, that they indicate what's at stake in such a decision. Regardless of which package prevails, Elgin's work illuminates a central question concerning how scientific models contribute to our epistemological goals.

6. Conclusion

We have argued that Elgin's project highlights a paradox that should worry anyone who accepts that idealisations, which are false and sometimes radically so if interpreted literally, play a significant epistemic role across science. This paradox brings together the questions of how to understand the role of truth in explanation and understanding, and of how to interpret such idealisations in the first place. This shows that these two debates have to be addressed in tandem. We demonstrated how Elgin, herself a significant contributor to both debates, provides an account of scientific representation that is compatible with literalism, thereby providing the means with which to undercut her own argument against veritism. In addition, we provided some arguments for why literalism might be preferred to a rejection of veritism. Most importantly, what we hope to have done is demonstrate where the various fault lines and connections lie in the debates in question, and draw attention to how significant they are for the philosophy of science.

Acknowledgements

Thanks to Catherine Elgin for many illuminating discussions over the years, and for her encouragement and support for our project. Jochen Briesen, Catherine Elgin, and two anonymous referees offered comments on earlier versions, which

[22] For example, within the framework we can allow for the possibility that there is a functional dependence between the idealised aspect of the model and the relevant feature of the target. In the case of limits, for instance, the DEKI account can capture the idea that idealising a feature of a model by ε will distort the corresponding feature in the target by δ.

helped improve our arguments. Thanks also to the organisers and participants in the workshop on *True Enough* at the University of Innsbruck and an audience at the University of Oxford for helpful discussions. JN thanks the History and Philosophy of Science program at the University of Notre Dame, where this paper was started, and the Jacobsen Trust, the British Academy (via a Rising Star Engagement Award), and the Jeffery Rubinoff Sculpture Park for support in finishing it.

References

Bailer-Jones, D. M. (2002). Models, Metaphors and Analogies. In P. Machamer, & M. Silberstein (Eds.), *The Blackwell Guide to the Philosophy of Science* (pp. 108-127). Oxford: Blackwell.

Baumberger, C., & Brun, G. (2017). Dimensions of Objectual Understanding. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science* (pp. 165-189). New York: Routledge.

Baxter, R. J. (1982). *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Black, M. (1962). *Models and Metaphors. Studies in Language and Philosophy*. Ithaca, New York: Cornell University Press.

Bokulich, A. (2009). Explanatory fictions. In M. Suárez (Ed.), *Fictions in Science. Philosophical Essays on Modelling and Idealization* (pp. 91-109). London and New York: Routledge.

Bokulich, A. (2012). Distinguishing Explanatory from Nonexplanatory Fictions. *Philosophy of Science, 79*(5), 725-737.

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

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Chakravartty, A. (2017). "Scientific Realism", The Stanford Encyclopedia of Philosophy (Summer 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.

Doyle, Y., Egan, S., Graham, N., & Khalifa, K. (2018). Non-Factive Understanding: A Statement and Defense (forthcoming). *Journal for General Philosophy of Science, Special Issue on Normative and Naturalistic Approaches to Understanding*.

Elgin, C. Z. (1996). *Considered Judgement*. Princeton: Princeton University Press.

Elgin, C. Z. (2009). Exemplification, Idealization, and Scientific Understanding. In M. Suárez (Ed.), *Fictions in Science. Philosophical Essays on Modeling and Idealization* (pp. 77-90). New York and London: Routledge.

Elgin, C. Z. (2010). Telling instances. In R. Frigg, & M. C. Hunter (Eds.), *Beyond Mimesis and Convention: Representation in Art and Science* (pp. 1-18). Berlin and New York: Springer

Elgin, C. Z. (2017). *True Enough*. Cambridge, Mass: MIT Press.

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

Frigg, R., & Nguyen, J. (2016a). The Fiction View of Models Reloaded. *The Monist, 99*, 225-242.

Frigg, R. & Nguyen, J. (2016b). "Scientific Representation", The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2018/entries/scientific-representation/

Frigg, R., & Nguyen, J. (2017a). Models and representation. In L. Magnani, & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 49-102). Cham: Springer.

Frigg, R., & Nguyen, J. (2017b). Scientific represenation is represenation as. In H.-K. Chao, & R. Julian (Eds.), *Philosophy of Science in Practice: Nancy Cartwright and the Nature of Scientific Reasoning* (pp. 149-179). Cham: Springer.

Frigg, R., & Nguyen, J. (2018). The turn of the valve: representing with material models. *European Journal for Philosophy of Science, 8*(2), 205-224.

Gelfert, A. (2017). The Ontology of Models. In L. Magnani, & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 5-23). Heidelberg/New York: Springer.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(4), 742–752.

Giere, R. N. (2009). Why scientific models should not be regarded as works of fiction. In M. Suárez (Ed.), *Fictions in Science. Philosophical Essays on Modelling and Idealization* (pp. 248-258). London: Routledge.

Goodman, N. (1976). *Languages of Art* (2nd ed ed.). Indianapolis and Cambridge: Hackett.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Harker, D. (2012). How to Split a Theory: Defending Selective Realism and Convergence without Proximity. *British Journal for the Philosophy of Science, 64*(1), 79-106.

Hesse, M. (1963). *Models and Analogies in Science*. London: Sheed and Ward.

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science, 64*(Supplement), S325-S336.

Khosrowi, D. (forthcoming). Getting Serious about Shared Features. *British Journal for the Philosophy of Science*.

Levy, A. (2015). Modeling without models. *Philosophical Studies, 152*(3), 781-798.

Maki, U. (2011). Models and the locus of their truth. *Synthese, 180*(1), 47-63.

Morgan, M., & Boumans, M. (2004). Secrets hidden by two-dimensionality: the economy as a hydraulic machine. In S. de Chadarevian, & N. Hopwood (Eds.), *Models: The Third Dimension of Science*. Stanford: Stanford University Press.

Nguyen, J. (forthcoming). It's Not a Game: Accurate Representation with Toy Models, *British Journal for the Philosophy of Science*.

Norton, J. (2003). Causation as Folk Science. *Philosophers' Imprint, 3*(4), 1-22.

Oddie, Graham (2006), "Truthlikeness", The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2016/entries/truthlikeness/

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*: Routledge.

Salis, F. and Frigg, R. (forthcoming). Capturing the Scientific Imagination. In P. Godfrey-Smith, & A. Levy (Eds.), *The Scientific Imagination*. New York: Oxford University Press.
Strevens, M. (2008). *Depth*. Cambridge, Mass: Harvard University Press.
Strevens, M. (2017). How idealizations provide understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*. New York: Routledge.
Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*(5), 767-779.
Suárez, M. (Ed.). (2009). *Fictions in Science. Philosophical Essays on Modelling and Idealization*. London and New York: Routledge.
Suárez, M. (2010). Fictions, Inference, and Realism. In J. Woods (Ed.), *Fictions and Models: New Essays* (pp. 225-246). Munich: Philosophia Verlag.
Suárez, M. (2015). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science, 49*, 36-47.
Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis, 55*(3), 393–415.
Templeton, A. R. (2006). *Population Genetics and Microevolutionary Theorey*. Hoboken, NJ: John Wiley & Sons.
Toon, A. (2010). Models as make-believe. In R. Frigg, & M. Hunter (Eds.), *Beyond Mimesis and Convention: Representation in Art and Science* (pp. 71-96). Berlin Springer.
Toon, A. (2011). Playing with molecules. *Studies in History and Philosophy of Science, 42*, 580-589.
Toon, A. (2012). *Models as Make-Believe. Imagination, Fiction and Scientific Representation*. Basingstoke: Palgrave Macmillan.
van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press.
van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.
Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science, 58*, 207-233.
Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
Woods, J. (Ed.). (2010). *Fictions and Models: New Essays*. Munich: Philosophia Verlag.
Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica, 43*(1-2), 99-124.