Factive inferentialism and the puzzle of model-based explanation

LSE Research Online URL for this paper: http://eprints.lse.ac.uk/110749/

Version: Published Version

Article:

ORIGINAL RESEARCH

Factive inferentialism and the puzzle of model-based explanation

Philippe Verreault-Julien[1]

Abstract

Highly idealized models may serve various epistemic functions, notably explanation, in virtue of representing the world. Inferentialism provides a prima facie compelling characterization of what constitutes the representation relation. In this paper, I argue that what I call factive inferentialism does not provide a satisfactory solution to the puzzle of model-based—factive—explanation. In particular, I show that making explanatory counterfactual inferences is not a sufficient guide for accurate representation, factivity, or realism. I conclude by calling for a more explicit specification of model-world mismatches and properties imputation.

Keywords Models · Representation · Explanation · Inferentialism · Factivity

1 Introduction

Scientific models often contain assumptions that simplify, abstract, or idealize from reality. One issue that arises from this state of affairs concerns the epistemic status of these models. For if models misrepresent and contain assumptions that falsely describe reality, how can we use them to learn about target systems? In particular, models allow for *surrogative reasoning* (Swoyer 1991). That is, models enable or license their users

---

✉ Philippe Verreault-Julien
  p.verreault-julien@lse.ac.uk
  https://pvjulien.net

1 Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, UK

🖄 Springer

to draw inferences about their targets. These inferences may serve various purposes and can be about, e.g., the sort of entities that exist in the targets, their properties, causal relations, or the future value of variables.[1] For instance, viewing that *X* causes *Y* in the model may enable the inference that *X* causes *Y* in the target. In other words, modellers build and study models in order to draw conclusions about the things they are surrogates of. Accounts of scientific representation face the challenge of spelling out the conditions under which models may allow surrogative reasoning (Frigg and Nguyen 2020, p. 3 ff.). Why is it that studying a representation helps us to draw conclusions about its target?

One purpose models are used for is to explain phenomena. How idealized models can license our explanatory inferences appears to pose a special challenge insofar as our theories of explanation typically require that explanations be factive (e.g. Craver 2006; Hempel 1965; Strevens 2008; Woodward 2003b). How could models that seemingly misrepresent ever warrant us to conclude about true explanations? One influential answer, representationalism, holds that it is because models faithfully represent their target that they can serve their various purposes (e.g. Bailer-Jones 2009; Giere 2010; Mäki 2011; Morgan 1999; Weisberg 2013). If models represent some aspects of the world, then scientists can use them to explain phenomena of interest. But what relationship must there be between a model and its target for the former to represent accurately the latter? Various accounts of the representation relation have been proposed. So-called structuralists have suggested that accurate representation is based on shared structural features like isomorphism between models and their targets (e.g. French and Ladyman 1999; Da Costa and French 2003; Van Fraassen 1980; Suppe 1977; Suppes 2002). Models represent since they preserve the structure of the target. Others have instead proposed that the similarity between models and their targets grounds the representation relation (e.g. Giere 1988, 2004; Poznic 2016; Teller 2001; Weisberg 2012, 2013).

The inferentialist account of representation (Suárez 2004, 2015; Suárez and Solé 2006) grew out of dissatisfaction with the structuralist and similarity conceptions (Suárez 2003). Instead of explaining the capacity of models to allow surrogative reasoning in terms of them representing, inferentialism explains the representation relation in terms of models allowing to draw inferences (Frigg and Nguyen 2017). Thus, the fundamental feature becomes surrogative reasoning itself in place of the representation relation. However, Suárez's inferentialism is silent about whether models may represent and satisfy the factivity of explanation.

I argue that what I call *factive inferentialism* (Kuorikoski and Ylikoski 2015; see also Kuorikoski and Lehtinen 2009; Ylikoski 2009; Ylikoski and Aydinonat 2014; Ylikoski and Kuorikoski 2010) does not provide a satisfactory solution to the puzzle of factive model-based explanation. Section 2 introduces factive inferentialism and the claim that it is compatible with factivity and realism. However, I show in Sects. 3 and 4 that correctly answering 'what-if-things-had-been-different' questions is plausibly not a sufficient guide for accurate representation, factivity, or realism. Then, in Sect. 5 I argue that a more complete answer would need to specify explicitly how we should

---

[1] 'Inference' should be understood broadly. I follow the literature in taking the key feature to be *reasoning* from a model to a target. That reasoning can be deductive, inductive, analogical, abductive, etc. (see, e.g., Suárez 2004).

interpret model-world mismatches and, more importantly, what sort of properties we should impute to the world. Section 6 concludes.

2 Representation, inferentialism, and explanation

Critics of representationalism point out that it has proven hard to specify notions of similarity and structure that would solve what Frigg and Nguyen (2017, 2020) call the ER-Problem.[2] The ER-Problem is one of identifying the conditions that make model *M* an *epistemic* representation of *T*. For instance, one way of characterizing the similarity account's answer to the ER-Problem is that a "scientific model *M* represents a target *T* iff *M* and *T* are similar" (Frigg and Nguyen 2017, p. 58). However, an adequate account of representation should also allow for the possibility of *mis*representation. Whether something is a representation is a separate issue of whether it is an accurate one. Thus, even if a representation fails to be similar to its target, we still want to say that it is a representation.

One line of defence is to rely on pragmatic factors such as the users' intentions to specify that a given model is meant to be a representation. So even if a model fails to be similar, but was intended to be, then it would still count as a representation. However, introducing pragmatic factors, Frigg and Nguyen (2017) argue, shifts the role of similarity or structure. It is not clear what they are doing for us in solving the ER-Problem if intentions are the key factor. Inferentialism aims to solve this problem by reversing the relationship between representation and surrogative reasoning. Instead of viewing the inferential affordances of models as depending on them representing, inferentialism characterizes the representation relation in terms of these inferences (Frigg and Nguyen 2017). Surrogative reasoning itself becomes the fundamental relation instead of representation.

One influential strand of inferentialism is Suárez's (2004, see also 2015; Suárez and Solé 2006). He argues that we should "adopt from the start a deflationary or minimalist attitude and strategy towards the concept of scientific representation" (2004, p. 770). He denies, among others, that similarity or isomorphism can be necessary or sufficient conditions for representation (Suárez 2003). Suárez also does not identify a substantive property relating a model and its target, not even the capacity to draw true conclusions (e.g., the true value of a variable). According to him, one condition for a representation source system *A* to represent a target system *B* is that "*A* allows competent and informed agents to draw specific inferences regarding *B*" (2004, p. 773).[3] The normative standards of inferential correctness "are inferential merely, and do not depend on the truth or otherwise of premises or conclusions" (Suárez 2004, fn. 8, see also 2010, fn. 25). Making correct inferences is only a matter of inferential validity, as in deductive validity, not of soundness. Inferential validity, in turn, is determined by the inferential rules that a given epistemic community establishes. Suárez's

---

[2] I use the term 'representationalism' to refer to accounts of representation that, as we will see, explain surrogative reasoning in terms of the representation relation.

[3] Suárez (2015, p. 45) says this a general feature of representation and "at best" a necessary condition. The other general feature, which I do not discuss here, is that a there is 'representational force' between a model and its target (see also Suárez 2004).

deflationism is pragmatic since it puts practice at the centre stage of the representation relation. For him, it is impossible for a notion of representation in an area of science to differ from how representation is actually regulated: "representation in that area, if it is anything at all, is *nothing but* that practice" (Suárez 2015, p. 38, emphasis in original).

But even if we assume inferentialism provides a successful answer to the ER-Problem, can pragmatic criteria shed light on all issues pertaining to scientific representation? In particular, there are two questions that it is crucial to separate (Chakravartty 2010; see also Contessa 2007, p. 68; Frigg and Nguyen 2017):

1. What is it to scientifically *represent*?
2. What makes a scientific representation *accurate*?

These questions are separate because, as noted above, we need to allow for the possibility of misrepresentation. A representation may be inaccurate, yet still be a representation.[4] For instance, the Ptolemaic geocentric model is a representation, but an inaccurate one.

Chakravartty (2010) distinguishes between what he calls 'functional' and 'informational' accounts of representation. Functional accounts focus on the inferential and interpretative practices agents perform using representations of target systems. Therefore, they primarily purport to answer the first question. Informational accounts, on the other hand, focus on the objective relations that obtain between representations and their targets and thus provide solutions to the second question. Functional and informational accounts, Chakravartty argues, should not be opposed to each other. Rather, they are complementary as they illuminate the two sides of the representation coin (see also Contessa 2007).

In Chakravartty's terms, Suárez's account is functional since it does not provide any substantive answer regarding representational accuracy. As such, it faces two sorts of challenges. First, it has been objected on the ground that it solves the problem of representation at the cost of obscuring what makes surrogative reasoning possible in the first place (Bolinska 2013; Contessa 2007). If the representation relation is itself made up by the inferential affordances of models, we may have made progress on what constitutes the representation relation, but not on accounting for why it is possible to make these inferences. By contrast, representationalism has a ready-made explanation; it is because models represent the world.

The second challenge inferentialism faces comes from accounts of explanation which typically demand that explanations be factive (e.g. Craver 2006; Hempel 1965; Strevens 2008; Woodward 2003b). Factivity requires that an explanation's explanans and explanandum be (approximately) true (see also Strevens 2013). Representationalism again seemingly offers a straightforward response on how models can offer factive explanations: they simply have to represent faithfully what they explain. For instance, a model may explain minimal product differentiation (Hotelling 1929; see Reiss 2012a) just in case the model faithfully represents the phenomenon and its explanatory factors, here transportation costs. Thus, representationalism serves well

---

[4] See Poznic (2018) for an argument concerning the entanglement of the "descriptive" and "evaluative" aspects of representation.

friends of the factivity of explanation, in particular, and, more generally, of scientific realism.[5] This line of defence, however, is prima facie not available to inferentialism since the normative inferential standards do not require the truth of the premises nor of the conclusions. Without substantive standards of inferential correctness, how do we know if the representation is accurate? It is thus unclear, to say the least, whether such an account is compatible with the factivity of explanation.

Kuorikoski and Ylikoski (2015; see also Kuorikoski and Lehtinen 2009; Ylikoski 2009; Ylikoski and Aydinonat 2014; Ylikoski and Kuorikoski 2010) claim that their inferentialist account of model-based explanation allows to salvage realism and the factivity of explanation. According to them, the "inferentialist analysis of representation" is all we need and they claim that it solves the puzzle of model-based explanation introduced above, namely how highly idealized models may provide (factive) explanations by way of being explanatory representations. In other words, their account purportedly answers the two questions of representation:

> […] a model (as an external inferential apparatus) represents some real world phenomenon by virtue of the fact that (and to the extent that) some cognitive agent (modeler) can use the apparatus to make correct inferences concerning the phenomenon. If the new inferences made possible by the model include counterfactual what-if inferences, then the model is explanatory and represents some crucial dependencies related to the phenomenon by virtue of facilitating these inferences. (Kuorikoski and Ylikoski 2015, p. 3827)

According to Kuorikoski and Ylikoski, there is thus a tight connection between surrogative reasoning, representation, and explanation.[6] To distinguish Kuorikoski and Ylikoski's (see also Kuorikoski and Lehtinen 2009) brand of inferentialism from Suárez's, I will call it *factive* inferentialism (FInf henceforth). FInf provides a three-layered account of representation and model-based explanation. First, FInf's answer to the ER-Problem is that the representation relation is constituted by the inferential affordances of a model. If a model M affords inferences about its target (e.g. about the value of variable Y), then M represents. In that respect, they follow Suárez. Second, FInf also provides a criterion for *accurate* representation, namely the *correctness* of the inferences. The representation relation is constituted by the inferential affordances "and the extent to which these inferences are correct determines how accurate the representation is" (Kuorikoski and Ylikoski 2015, p. 3830). In short, if M affords correct inferences (e.g. it is correct that Y takes value y), then M accurately represents. Third, if some of these correct inferences are counterfactual what-if inferences (w-inferences henceforth) that allow us to answer 'what-if-things-had-been-different' questions (w-questions henceforth), then that representation—or model—is also explanatory. In brief, if a model M allows its user to make correct w-inferences that answer w-questions, then M accurately represents and explains. In effect, FInf thus makes a distinction between explanatory (the w-inferences) and non-explanatory inferences (all the other ones, e.g. about the value of variables).

---

[5] That does not mean, however, that representationalism entails realism. It is also possible to hold an antirealist position (e.g. Rueger 2005).

[6] It is also closely related to their account of model-based *understanding* as an ability to make counterfactual inferences.

FInf, therefore, goes beyond traditional inferentialism by providing criteria for both accurate and explanatory representation. Prima facie, it seems to dispel the worries friends of the factivity of explanation may have with the deflationary character of inferentialism. FInf not only states what constitutes the representation relation, but also provides criteria to assess its correctness. That is why, according to FInf's proponents, there is nothing mysterious or philosophically puzzling about representation and explanation. A model that provides factive explanations is just a model that affords correct *w*-inferences.

3 *W*-inferences and *explanatory* representation

FInf tells us that if a model affords correct *w*-inferences, then the model accurately represents and explains. But what is it to afford *correct w*-inferences? In the following two sections, I will raise two sets of issues with taking counterfactual inferential affordances as sufficient for explanatory representation. The first concerns the distinction between explanatory and non-explanatory *w*-inferences. The second, that I discuss in Sect. 4, is about what we can justifiably infer from explanatory *w*-inferences.

One key desideratum of an account of model-based explanation is that we should be able "to explicate what makes the difference between truly explanatory and merely phenomenological models" (Kuorikoski and Ylikoski 2015, p. 3818). Phenomenological models describe the covariance of variables, they 'save the phenomena', whereas explanatory models typically provide us with insight about *why* the described reality is the way it is.[7] Models may save the phenomena, yet fail to explain. While it is correct to say that phenomenological models accurately represent some aspects of reality, that representation is not explanatory. Hence, the issue is not that phenomenological models do not represent, it is that they do not provide *explanatory* representations. We thus need a way to demarcate phenomenological models from explanatory ones. Absent a solution to this problem, this poses a threat to FInf insofar as it is supposed to illuminate in virtue of what models can explain. Even if it were sufficient for an account of model-based *representation*, it would not be for an account of *explanatory* representation.

FInf's solution to this is to specify that not all correct inferential affordances are explanatory, only (a subset of) the counterfactual ones. Counterfactual statements typically have the form 'Had *A* been the case, *C* would have occurred'. Relations of counterfactual dependence are change-relating in that they describe how the explanandum variables would change if the value of the explanans variables were to change. It is in virtue of the information of counterfactual dependence that these relations convey that they are explanatory. Evaluating the truth of counterfactuals requires specifying a semantics. A counterfactual may be true according to one semantics, but false under another one. FInf advocates pluralism about which counterfactual relations of

---

[7] For instance, Woodward (2003b, p. 10) says that descriptive knowledge "is knowledge that, although it may provide a basis for prediction, classification, or more or less unified representation or systemization, does not provide information potentially relevant to manipulation. It is in this that the fundamental contrast between causal explanation and description consists". See Leuridan et al. (2008) for the practical value of correlations.

dependence—e.g. causal or otherwise—may support *w*-inferences (Kuorikoski and Ylikoski 2015; Ylikoski 2013; Ylikoski and Kuorikoski 2010). In the following, I take it that FInf broadly adopts Woodward's (2003b) account of explanatory counterfactuals.

The main issue I would like to highlight here is that *w*-inferences simpliciter are not sufficient for explanatory representation. We need additional constraints on what counts as an *explanatory w*-inference. To illustrate, consider Woodward's (2003b, p. 197ff.) discussion of the explanation of a simple's pendulum period by the following relationship, where T is the period, g its gravitational acceleration, and ℓ the pendulum's length.

$$T = 2\pi \sqrt{\ell/g} \qquad (1)$$

Using this relationship, it is possible to make various *w*-inferences about the pendulum's period. Had the length ℓ been different, then the pendulum's period T would also have been different. As Woodward reminds us, it is also possible to derive the length ℓ from the period T and the acceleration g by converting the equation.

$$\ell = \frac{T^2 \cdot g}{4\pi^2} \qquad (2)$$

There is a sense in which that second equation also supports *w*-inferences and, crucially, can answer *w*-questions. For instance, physics students may be asked to derive the length of a pendulum had its period been different. This supposes that there is a sense in which had a pendulum's P period been different, then its length ℓ would also have been different. Woodward acknowledges this:

There is thus *some* interpretation of the counterfactual

(5.3.3) If the period of P were T^*, then the length would be ℓ^*

such that (5.3.3) comes out true. (Woodward 2003b, p. 197, emphasis in original)

The general problem Woodward raises with the pendulum example is that not all relations of counterfactual dependence are explanatory and that we thus need to distinguish explanatory from non-explanatory ones. More generally, for him "any generalization describing a correlation (or pattern of association) will be change-relating, for anything that counts as a correlation must tell us how variations in the value of one variable are associated with variations in the value of another" (2003b, p. 246; see also Khalifa et al. 2020; Pincock 2018). In that sense, correlations may allow to make *w*-inferences in that they relate the value of one variable to another one.

Of course, Woodward's point is precisely that despite the counterfactual dependence between the length and the period, the latter is irrelevant for causally explaining the former. Rather, the length explains the period. This is why he proposes to distinguish general change-relating generalizations from properly causal/explanatory ones with the notions of 'intervention' and 'invariance'. Basically, the idea is that if one were to physically intervene on the period, it would not have any consequence on the pendulum's length; the length is not causally connected to the period. However, if

one were to intervene on the pendulum's length, one could manipulate the period. A generalization is invariant if it correctly describes how the value of a variable would change under a range of interventions.

FInf recognizes the need to distinguish between phenomenological and explanatory models and would accept something like the interventionist test for a large class of *w*-inferences. However, since FInf is openly pluralist, it cannot simply restrict the relevant counterfactuals to those that are invariant under interventions.[8] Ylikoski, for instance, suggests that "prediction is a straightforward case of *what-if* inference: it tells how a system behaves under specific circumstances" (2014, p. 327, emphasis in original, 2009, pp. 101–102). A model that allows to predict may therefore allow correct *w*-inferences since a modeller can on that basis infer the past (retrodiction) or future (prediction) value of variables. Although it adheres to substantive explanatory criteria, the worry is thus that FInf might be too liberal.

Indeed, it is a well-known point that a model may be used to predict, yet may not explain. This is because the capacity of predicting the future value of variables does not require an accurate representation of their explanatory features. As Weisberg observes, maximizing the predictive accuracy ensures a model will be useful to generate predictions, "but gives no guarantee that the models will be useful for explaining the behavior of the system" (2007, p. 653). One way of thinking about this general problem is in terms of fidelity criteria and modelling trade-offs (Weisberg 2013; see also Gräbner 2018). Fidelity criteria determine whether a representation is adequate. Depending on the purpose, these criteria may be different; adequacy is thus purpose dependent. Dynamic fidelity concerns the comparison between model output and the target and is especially important for predictive purposes. But there are also representational fidelity criteria. Representational fidelity concerns whether a representation is sufficiently accurate, typically understood in terms of accurately representing a target's causal structure. For the purpose of explanation, dynamic fidelity is not sufficient, we also need representational fidelity.[9] Importantly, dynamic and representational fidelity do not necessarily accompany each other. In fact, there are often trade-offs between criteria. For example, forecasting success in economics is often obtained at the cost of *not* accurately representing the underlying causal structure (Reiss 2012b). Indeed, since causal relations in the economic world are often bound to break, having a non-causal model is often the only way to secure predictive accuracy. Thus, a model may facilitate *w*-inferences (dynamic fidelity) without being an explanatory representation (representational fidelity). Prediction and explanation do not always go hand in hand.

Therefore, making *w*-inferences simpliciter cannot be a sufficient condition for the accurate representation of *explanatory* relations. This would blur the distinction between phenomenological and explanatory models, a distinction FInf wants to uphold. Friends of FInf could retort that not all *w*-inferences can answer *w*-questions.[10] So even if correlations may support *w*-inferences, they may not allow to correctly answer *w*-questions. Put differently, not all *w*-inferences would be explanatory. This

---

[8] While Woodward focuses on causal explanation, he also acknowledges that there may be non-causal relations of dependence (see 2003b, sec. 5.9, 2018).

[9] See Gräbner (2018) for a discussion of model validation in practice.

[10] I would like to thank a reviewer for pushing me on this.

means that the account of explanation itself is doing all the heavy lifting. FInf answers the worries we may have with more deflationary brands of inferentialism (Suárez 2004, 2015; Suárez and Solé 2006) by appending substantive explanatory standards.

My main qualm with this otherwise acceptable reply is that it is not clear whether the explanatory standards FInf follows can by themselves salvage the factivity of explanation and realism. If one can endorse FInf's broad espousal of a counterfactual theory of explanation without upholding neither inferentialism nor that explanatory models accurately represent their targets, then it is not the explanatory standards that secure factivity, but something else.

4 *W*-inferences and *accurate* representation

The second set of issues concerns the very link between explanatory *w*-inferences and accurate representation. We have seen that FInf holds that if a model affords correct inferences, then that model accurately represents. If some of these inferences are *w*-inferences that allow to answer *w*-questions, then the model accurately represents and explains. Crucially, FInf maintains that this version of inferentialism accommodates both, via accurate representation, the factivity of explanation and realism. We can understand FInf as making a number of different claims concerning the relationship between the explanatory *w*-inferences a model *M* affords and accurate representation. If *M* affords explanatory *w*-inferences (and thus explains), then…

1. *M* accurately represents.
2. *M* provides a factive explanation.
3. *M* captures reality.

At first sight, it appears that all these propositions are intimately intertwined. If a model provides a factive explanation, then it must accurately represent. And if a model accurately represents, why should we doubt it captures—parts of—reality?

However, the problem is precisely that it is contentious whether we are justified in inferring anything from explanatory *w*-inferences. There are two main points of contention. The first one concerns the link between explanation and accurate representation. Many have claimed that models may explain without accurate representation (e.g. Batterman and Rice 2014; Bokulich 2008, 2011, 2012, 2016; De Regt 2015, 2017; Graham Kennedy 2012; Potochnik 2017). Bokulich, for one, argues that what she calls 'model explanations' capture the counterfactual dependencies of a target system *despite* the fictional representational content. As she illustrates, Bohr's model of the hydrogen atom does not, according to contemporary quantum mechanics, accurately represent the actual structure of reality. For her, Bohr's model is not a simple idealized representation of electron orbits. It does not distort reality, but instead introduces fictional entities or processes. Despite its fictional character, Bokulich argues that it explains the atomic spectrum of hydrogen. One reason is because "Bohr's model is able to correctly answer a number of 'what-if-things-had-been-different questions' […]" (Bokulich 2011, p. 43). The model thus captures a pattern of counterfactual

dependence between its fictional entities and spectral phenomena.[11] Bokulich argues that the key difference between phenomenological and explanatory fictional models is that the latter successfully undergo a "justificatory step". This step allows to demarcate which models are relevant for explanation. She notes that "although the range of w-questions that a phenomenological model can answer will typically be more limited, scope alone cannot distinguish between explanatory and phenomenological models" (Bokulich 2012, p. 733). Very briefly, the justificatory step involves (1) establishing a contextual relevance relation, (2) setting a domain of applicability, and (3) translating the model descriptions into correct conclusions about the target.

Bokulich's account is particularly interesting because it relies, like FInf, on the idea that explanation involves correctly answering w-questions with counterfactual relations of dependence (Woodward 2003b). However, whereas FInf concludes that correct w-inferences imply accurate representation (because they are a subset of correct inferences), Bokulich reaches an opposite verdict. According to her, the fact that a model is a fiction does not prevent the model from correctly answering w-questions. Bokulich's account thus suggests that explanatory w-inferences can cohabit with a view that denies accurate representation.[12]

FInf's proponents could reply that, pace Bokulich, Bohr's model *does* accurately represent. Accuracy is a matter of degree and while the model may accurately represent some aspects of spectral phenomena, it may misrepresent others. In fact, Bokulich herself appears to suggest something along those lines when she says that models explain only when a model's counterfactual structure is isomorphic to its target (Bokulich 2011, p. 39). There are two problems with this reply. First, the general issue is that the quantity and quality of explanatory w-inferences we can make with a model does not only depend on its representational accuracy. According to Ylikoski and Kuorikoski (2010; see also Kuorikoski and Ylikoski 2015), explanatory w-inferences can be assessed along five dimensions: non-sensitivity, precision, factual accuracy, degree of integration, and cognitive salience. Factually more accurate explanations are those that include "fewer falsehoods" (Ylikoski and Kuorikoski 2010, p. 212). Factual accuracy, thus, is crucial for representational accuracy, the factivity of explanation, and realism. Without going into the detail, the key point is that there are sometimes trade-offs between these dimensions. For instance, an increase in non-sensitivity (viz., robustness) may require *less* factual accuracy. And, as Kuorikoski and Ylikoski (2015, p. 3833) also recognize, including fewer falsehoods "might not improve explanatory understanding". Therefore, the quantity and quality of explanatory w-inferences a model affords will not only depend on its representational accuracy.

FInf's proponents may respond that even though there is not a perfect mapping between explanatory w-inferences and accurate representation, a representation could

---

[11] Bokulich (2011) indicates that a fictional model may accurately represent the *pattern* of counterfactual dependence. In other words, we could say that while the explanans is fictional and therefore false, the pattern is true.

[12] One reviewer helpfully suggested that we can understand FInf as a package of the three tenets introduced at the end of Sect. 2. Then, we can regard Bokulich as accepting the third tenet about explanation, but rejecting the two first tenets about (accurate) representation. Framing the problem in these terms highlights, I believe, that in fact it is the account of representation and not the explanatory standards that safeguards the factivity of explanation.

not be *completely* inaccurate and explain. This may be so, but then the second problem we run into is that it is not clear where that leaves us with respect to the factivity of explanation and realism. For example, even if we accept that answering *w*-questions with Bohr's model implies that there is *some* accurate representation, it appears to be inconsistent with the factivity of explanation. Using the model, one can infer how the emission spectrum of hydrogen would change if the orbits had been different. But, Bokulich reminds us, there are no stationary state orbits. Thus, in what sense is that explanation of emission spectrum supposed to be 'factive'?

What 'realism' this entails is likewise unclear. Woodward himself holds a minimal form of realism which he calls 'instrumental realism' (2003a, see also 2003b, pp. 223–224; Baird 1988; Saatsi 2020). This view is instrumentalist in that what matters is getting facts about relations of counterfactual dependence right, not the ontology. Models that capture similar patterns of dependence but postulate different entities would, according to Woodward, have the same representational content. Yet, the view is also realist because it downplays the significance of the observables/unobservables distinction and holds that scientific explanation goes beyond saving the phenomena (cf. Van Fraassen 1980). This account is "minimal" (Saatsi 2020) insofar as it entails much weaker ontological commitments than standard readings of realism.

Instead, maybe our commitment to realism should be a sort of 'epistemic realism'. Rice argues that it should concern the *understanding* we obtain with models (2021, 2016). According to Rice (2021, p. 4099), "pervasively inaccurate descriptions of reality" may afford factive understanding, which essentially involves being able to answer *w*-questions on the basis of true modal information. Models may thus misrepresent and give access to that modal information. Bokulich (2016, pp. 270–271) aptly puts the point by saying that "[t]he key move here is the recognition that scientific understanding requires having true modal information and the ability to draw correct inferences, but that one can achieve this 'factive understanding' without having a true or accurate representation".[13] Perhaps FInf's proponents would accept either of instrumental or epistemic realism. There is nothing wrong per se with these positions. However, as I discuss in more detail in the next section, neither form of realism seems to answer the problems FInf was trying to solve in the first place.

Nothing that I have said constitutes a conclusive blow against FInf. What I wanted to highlight is that we can readily accept that explaining—or understanding—phenomena is fundamentally a matter of making *w*-inferences that answer *w*-questions without accepting that these explanations are factive, that the models that ground them accurately represent, or that they capture reality. There is significant variance in the literature concerning what we can justifiably infer from correctly answering *w*-questions. None of the views presented above are obviously wrong. Hence, nothing seems to follow directly from adopting FInf concerning accurate representation, the factivity of explanation, and realism. We could assert with FInf that it is "conceptually impossible" (Kuorikoski and Ylikoski 2015, p. 3830) for a model to allow to answer *w*-questions and not represent accurately, but it is hard to see how this is not just stipulating one's

---

[13] Interestingly, Bokulich (2016, p. 271) reads Kuorikoski and Ylikoski (2015) as being proponents of the view according to which one can have factive understanding of a phenomenon without accurate representation. This interpretation does not appear to be supported by Kuorikoski and Ylikoski's conception of the representation relation and their realist commitments.

way out of the puzzle of model-based explanation. At any rate, even if we accept the positive reasons to adopt an inferentialist perspective on representation, we would still need substantive criteria to demarcate correct from incorrect inferences.

5 Factive inferentialism: the way forward

The discussion in the previous section and this one suggests that the explanatory *w*-inferences a model affords may be a sufficient condition for accurate representation, factivity, or realism only if we disambiguate in virtue of what these *w*-inferences are correct in the first place. FInf constitutes progress over more deflationary brands of inferentialism (Suárez 2004, 2015) because it appends substantive explanatory standards. But if these standards do not imply by themselves accurate representation, then they cannot safeguard the factivity of explanation. This also suggests that the key criteria for factive explanation are to be found in the account of representation, not the explanatory standards. What is the way forward?

One avenue would involve weakening FInf. It submits that correctly answering *w*-questions is sufficient for accurate representation, factivity, and realism. However, a safer reading would be to view it as only sufficient for a sort of 'epistemic realism' without any further commitment with respect to how the model captures or represents reality. As we have seen in Sect. 4, Rice (2021) and Bokulich (2016) both argue we can obtain factive understanding without accurate representation. Here the strategy would be to disconnect cognitive achievements like understanding from other explanatory requirements like factivity. Having modal information that allows us to answer *w*-questions may be sufficient for understanding, the argument goes, but it may not be for factive explanation.[14]

This has some prima facie appeal insofar as an inferential conception of understanding is part and parcel of FInf.[15] Since the epistemic benefit we arguably want to salvage is understanding, severing it from explanation seems to solve partially FInf's indeterminacy regarding explanatory representation. We lose little by conceding that answering *w*-questions via modal information affords understanding, yet sometimes falls short of explanation. Moreover, that understanding can be had without an actual explanation has some support in the literature (Lipton 2009; Gijsbers 2013; Rice 2016; Verreault-Julien 2019).[16] If this potentially provides a suitable solution to the puzzle of model-based *understanding* it does not, however, fully solve the puzzle of model-based *factive explanation*. The initial motivation was to account for how highly idealized models may accurately represent and provide factive explanations, not understanding.

---

[14] Although not explicit, there is a reading of Ylikoski that would support this interpretation (2009, p. 102, 2014, p. 327).

[15] They say that "[e]xplanation and understanding are not different kinds of activities, mental processes, or methods. Successful explanations convey the ability to provide answers to contrastive what-if questions, and understanding is based on a body of knowledge that provides the basis for these answers" (Kuorikoski and Ylikoski 2015, p. 3820).

[16] For opposing views, see De Regt (2009), Khalifa (2012), Strevens (2013), or Trout (2002).

One advantage inferentialism has over standard representationalism in solving the puzzle is that it can more easily avoid the pitfalls of *literalism*. Literalism is "the claim that models have to be interpreted as sharing features with their targets in order to be accurate representations of those features" (Frigg and Nguyen 2021, p. 2435). Shared features accounts are those that hold that models accurately represent their targets if and only if they share features with them.[17] As Frigg and Nguyen remark, similarity and structuralist accounts of representation epitomize this tenet. In principle, inferentialism has no such built-in requirement. If a model affords (correct) inferences, then it (accurately) represents. Inferential correctness becomes the metric of accuracy instead of shared features.

Although FInf avoids the language of shared features, it also invites us to take at face value—i.e., literally—that some parts of the model are *true* and that it is those parts that are responsible for a model's explanatory power and realism.[18] FInf broadly follows what Rice (2019, see also 2018) calls the "decompositional strategy".[19] In the context of modelling, the strategy involves decomposing models in various components, some misrepresenting their targets and others accurately representing them. A model's idealizations—the misrepresenting parts—are not necessarily problematic to the extent that the model also accurately represents the relevant ones, for instance a phenomenon's causes. One benefit of the decompositional strategy is that it allows, in principle, to salvage literalism. Only a model's relevant parts should be interpreted as sharing features with its target, not the idealizations.

FInf's endorsement of the decompositional strategy becomes clear in its discussion of the different types of assumptions and their truth. Models contain substantial, Galilean, and tractability assumptions. Tractability assumptions are those literally false assumptions that are nevertheless necessary to derive results. Galilean assumptions (or idealizations) isolate the difference-making features of interest by assuming away the presence of other features. Substantial assumptions are those that represent the explanatory factors of interest. While the substantial assumptions need to be true or realistic, the idealizations or the tractability assumptions may be false.[20] According to FInf, only the former are epistemically relevant, not the latter: "it is the truthlikeness of the substantial assumptions that ultimately carries the epistemic weight in a model" (Kuorikoski and Lehtinen 2009, p. 127; see also Kuorikoski and Ylikoski 2015, p. 3827ff.).[21]

---

[17] See Khosrowi (2020) for a recent discussion of shared features.

[18] One could also ask to what extent FInf fully bypasses shared features. If making correct inferences depends on the truth of the substantial assumptions, then it seems model and target actually need to share features; a substantial assumption is true insofar as what it states obtains in the target. At any rate, much seems to hang on the notion of 'truth' involved.

[19] See also Sullivan and Khalifa's (2019) '3D' discussion of idealizations, in particular the downplaying and demythologizing aspects.

[20] This does not imply that the same assumption has the same function across contexts (see Kuorikoski and Ylikoski 2015, p. 3831). For instance, assuming that preferences are transitive may be a substantial assumption in a given context, but a mere tractability assumption in another. I would like to thank a reviewer for bringing this to my attention.

[21] This is closely related to the discussion on how robustness analysis can help identify which of a model's assumptions are substantial or merely support tractability (Kuorikoski et al. 2010, 2012). See Cartwright (2009) and Rice (2019) for challenges to this possibility.

One reason why FInf presumably frames the issue in terms of the truth of the substantial assumptions is to salvage the realist thesis that science successfully refers to reality (Chakravartty 2007; Psillos 1999). Models that allow us to answer w-questions have true substantial assumptions that (truly) capture reality. However, the key issue is that we cannot make a straight jump from correctly w-questions to accurate representation and, even less so, to the realisticness or truth of a model's assumptions. This is because correctly answering w-questions may not require the truth of substantial assumptions in that literal sense. Provided FInf is serious about its commitment to inferentialism, what we need is a way of working our way back from the correct inferences to truth and accuracy without having to deny that models have literally false components. We would want to say that models may have these literally false components, yet can successfully explain or accurately represent. This, I submit, requires abandoning literalism.

Frigg and Nguyen's (2016, 2017, 2018, 2020, 2021) account of scientific representation helps to understand how FInf could fully move past literalism and how it may differ from, e.g., Bokulich's (2011, 2012) and Suárez's (2004, 2015) views. According to Frigg and Nguyen (2021, p. 2438), a model M is a scientific representation of target T iff:

1. Denotation: M denotes T.
2. Exemplification: M exemplifies z-properties $F_1, \ldots, F_n$.
3. Keying up: A key K associates the set $\{F_1, \ldots, F_n\}$ with a set of properties $\{G_1, \ldots, G_n\}$.
4. Imputation: M imputes at least one of the properties $G_1, \ldots, G_n$ to T.

One fundamental feature of the DEKI account—for denotation, exemplification, keying up, and imputation—is that it emphasizes the need to interpret via a key the properties exemplified by a model in terms of properties imputed to a target. The properties $F_1, \ldots, F_n$ a model exemplifies need not all be literally imputed to a target. First, a key translates the F-properties in terms of G-properties. Then, the latter can be imputed to the target.[22] Crucially, that does not require the truth of any 'assumption'. Accurate representation is a matter of correctly imputing (or, inferring) properties to a target and does not depend on the truth of any part of a model. The keying up step allows for a mismatch between the model itself and its target. Therefore, we can—but do not have to—read models literally.

I submit that the reason why we cannot make the jump from correctly answering w-questions to accurate representation, factivity, and realism is because the keying up and imputation steps are underspecified. In fact, we can use these steps to understand the major differences between FInf, Bokulich, and Suárez. Suárez's inferentialism implies that keying up and imputation are purely pragmatic. Practice determines the keys and what properties are imputed to a target. Therefore, we cannot infer from 'correct' inferences that a given model imputes real properties or features to the world.

---

[22] Prima facie, the DEKI acccount is not consistent with inferentialism as they diverge concerning what constitutes the representation relation. My point is simply that we can use use the keying up and imputation conditions to shed light on inferential practices. For example, inferentialists may deny that models denote or exemplify properties, but broadly accept that we need to translate (keying up) what is in the model in terms of features of the world (imputation).

Bokulich's case is interesting. In the previous section, I emphasized that, according to her, models that misrepresent may nonetheless allow us to answer *w*-questions. In particular, her fictional view of models intimates that there can be a considerable model-world mismatch and that properties should not be keyed up in terms of identity. For instance, although Bohr's model of the atom exemplifies the electron moving in a classical stationary state, the key tells us that we should not associate this property as is to the world. Of the keyed up properties, only patterns of counterfactual dependence are imputed to the target, not the electron orbits moving in classical fashion. But Bokulich also claims that her "explanatory fictions lies within a broadly realist approach to science" (2016, p. 261) and that "explanation and understanding are 'success terms,' in that they require getting something right about the way the world is […]" (2018, p. 796). We can thus interpret her as endorsing that science aims at and achieves true imputation of patterns of counterfactual dependence, but not necessarily the entities and processes that underlie them.[23]

FInf's implicit proposal, then, resorts to more demanding realist keying up and imputation steps. For model-based explanation, this minimally requires that the imputed properties that concern the explanandum and the explanans be true. What sort of properties need to be imputed will depend on one's favourite account of explanation.[24] It may only be patterns of counterfactual dependence, but it may also involve other properties like, for instance, being a law of nature. But, crucially, it will also depend on one's broader realist commitments. One may be comfortable with a form of instrumental realism (e.g. Woodward 2003a) and leave ontology alone. Others may give more weight to getting facts about reality right. Contrary to Suárez, FInf at least requires inferences to true conclusions about the imputed properties. And Bokulich's realism is arguably less demanding than FInf's insofar as hers does not involve imputing to the world many of the properties that we usually associate with realism.

Specifying what different brands of factivity and realism imply in terms of keying up and imputation is beyond the scope of this paper.[25] My more modest point is to submit that we need an explicit act of interpretation, keying up, and imputation that provides the accuracy, explanatory, and realist standards. Without information about these steps, we cannot assess in what sense the representation is accurate, the explanation, factive, and the science, realist. And, critically, we cannot judge what to infer from inferential success.

6 Conclusion

The capacity of models to help us learn about the world has traditionally been accounted for in terms of a representation relation. It is because models represent that

---

[23] Note the similarities between the keying up and imputation steps and Bokulich's (2011, 2012) requirement of a "justificatory step" introduced in Sect. 4. The third step states "that *there must be something like a key* that allows scientists to translate statements about the fictional or idealized elements in the model into correct conclusions about the target system" (Bokulich 2012, p. 736, my emphasis).

[24] Of course, adopting a non-factive account of explanation would make different demands to keying up and imputation.

[25] See Frigg and Nguyen (2021) for a brief discussion of this issue.

we can draw inferences about phenomena of interest and explain them. Inferentialism reverses that relationship and claims that surrogative reasoning is the fundamental activity. However, it has trouble showing how models can offer factive explanations. FInf builds on inferentialism and claims it is compatible with the factivity of explanation and realism.

I argued that, by itself, FInf does not solve the puzzle of model-based explanation. Basically, the problem is that the answering of *w*-questions only entails minimal commitments with respect to representational accuracy, the factivity of explanation, and realism. I then suggested ways FInf could solve these issues, more specifically by abandoning literalism and making more explicit the properties FInf proposes to impute from models to targets. This, I believe, would make FInf's solution to the puzzle of model-based explanation stronger and genuinely distinct.

References

Bailer-Jones, D. M. (2009). *Scientific models in philosophy of science*. Pittsburgh: University of Pittsburgh Press.

Baird, D. (1988). Five theses on instrumental realism. In *PSA: Proceedings of the biennial meeting of the Philosophy of Science Association* (pp. 165–173).

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

Bokulich, A. (2008). Can classical structures explain quantum phenomena? *The British Journal for the Philosophy of Science*, *59*(2), 217–235.

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*(1), 33–45.

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*(3), 260–279.

Bokulich, A. (2018). Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science*, *85*(5), 793–805.

Bolinska, A. (2013). Epistemic representation, informativeness and the aim of faithful representation. *Synthese*, *190*(2), 219–234.

Cartwright, N. (2009). If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis*, *70*(1), 45–58.

Chakravartty, A. (2007). *A metaphysics for scientific realism. Knowing the unobservable*. Cambridge: Cambridge University Press.

Chakravartty, A. (2010). Informational versus functional theories of scientific representation. *Synthese*, *172*(2), 197–213.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science*, *75*(1), 48–68.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376.

Da Costa, N. C. A., & French, S. (2003). *Science and partial truth: A unitary approach to models and scientific reasoning*. Oxford: Oxford University Press.

De Regt, H. W. (2009). The epistemic value of understanding. *Philosophy of Science*, *76*(5), 585–597.

De Regt, H. W. (2015). Scientific understanding: Truth or dare? *Synthese*, *192*(12), 3781–3797.

De Regt, H. W. (2017). *Understanding scientific understanding*. New York: Oxford University Press.

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science*, *13*(2), 103–121.

Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist*, *99*(3), 225–242.

Frigg, R., & Nguyen, J. (2017). Models and representation. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science, Springer handbooks* (pp. 49–102). Dordrecht: 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.

Frigg, R., & Nguyen, J. (2020). *Modelling nature: An opinionated introduction to scientific representation*. Cham: Springer Nature.

Frigg, R., & Nguyen, J. (2021). Mirrors without warnings. *Synthese*, *198*(3), 2427–2447.

Giere, R. N. (1988). *Explaining science. A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, *71*(5), 742–752.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*(2), 269–281.

Gijsbers, V. (2013). Understanding, explanation, and unification. *Studies in History and Philosophy of Science Part A*, *44*(3), 516–522.

Gräbner, C. (2018). How to relate models to reality? An epistemological framework for the validation and verification of computational models. *Journal of Artificial Societies and Social Simulation*, *21*(3), 8.

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

Hempel, C. G. (1965). *Aspects of scientific explanation: And other essays in the philosophy of science*. New York: Free Press.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*(153), 41–57.

Khalifa, K. (2012). Inaugurating understanding or repackaging explanation? *Philosophy of Science*, *79*(1), 15–37.

Khalifa, K., Doble, G., & Millson, J. (2020). Counterfactuals and explanatory pluralism. *The British Journal for the Philosophy of Science*, *71*(4), 1439–1460.

Khosrowi, D. (2020). Getting serious about shared features. *The British Journal for the Philosophy of Science*, *71*(2), 523–546.

Kuorikoski, J., & Lehtinen, A. (2009). Incredible worlds, credible results. *Erkenntnis*, *70*(1), 119–131.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science*, *61*(3), 541–567.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2012). Robustness analysis disclaimer: Please read the manual before use!. *Biology and Philosophy*, *27*(6), 891–902.

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

Leuridan, B., Weber, E., & Van Dyck, M. (2008). The practical value of spurious correlations: Selective versus manipulative policy. *Analysis*, *68*(4), 298–303.

Lipton, P. (2009). Understanding without explanation. In H. W. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding. Philosophical perspectives* (pp. 43–63). Pittsburgh: University of Pittsburgh Press.

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

Morgan, M. S. (1999). Learning from models. In M. Morrison & M. S. Morgan (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 347–388). Cambridge: Cambridge University Press.

Pincock, C. (2018). Accommodating explanatory pluralism. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 39–56). Oxford: Oxford University Press.

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

Poznic, M. (2016). Representation and similarity: Suárez on necessary and sufficient conditions of scientific representation. *Journal for General Philosophy of Science*, *47*(2), 331–347.

Poznic, M. (2018). Thin versus thick accounts of scientific representation. *Synthese*, *195*(8), 3433–3451.

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

Reiss, J. (2012a). The explanation paradox. *Journal of Economic Methodology, 19*(1), 43–62.

Reiss, J. (2012b). Idealization and the aims of economics: Three cheers for instrumentalism. *Economics and Philosophy, 28*(3), 363–383.

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

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

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

Rice, C. (2021). Understanding realism. *Synthese*, *198*(5), 4097–4121.

Rueger, A. (2005). Perspectival models and theory unification. *The British Journal for the Philosophy of Science*, *56*(3), 579–594.

Saatsi, J. (2020). Realism and explanatory perspectives. In M. Massimi & C. D. McCoy (Eds.), *Understanding perspectivism. Scientific challenges and methodological prospects* (pp. 65–84). New York: Routledge.

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

Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A*, *44*(3), 510–515.

Suárez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, *17*(3), 225–244.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, *71*(5), 767–779.

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

Suárez, M. (2015). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science Part A*, *49*, 36–47.

Suárez, M., & Solé, A. (2006). On the analogy between cognitive representation and truth. *Theoria*, *55*, 39–48.

Sullivan, E., & Khalifa, K. (2019). Idealizations and understanding: Much ado about nothing? *Australasian Journal of Philosophy*, *97*(4), 673–689.

Suppe, F. (1977). *The structure of scientific theories*. Urbana: University of Illinois Press.

Suppes, P. (2002). *Representation and invariance of scientific structures*. Stanford: CSLI Publications.

Swoyer, C. (1991). Structural representation and surrogate reasoning. *Synthese*, *87*(3), 449–508.

Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis*, *5*(3), 393–415.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, *69*(2), 212–233.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

Verreault-Julien, P. (2019). Understanding does not depend on (causal) explanation. *European Journal for Philosophy of Science*, *9*(2), 18.

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

Weisberg, M. (2012). Getting serious about similarity. *Philosophy of Science*, *79*(5), 785–794.

Weisberg, M. (2013). *Simulation and similarity. Using models to understand the world*. Oxford: Oxford University Press.

Woodward, J. (2003a). Experimentation, causal inference, and instrumental realism. In H. Radder (Ed.), *Philosophy of scientific experimentation* (pp. 87–118). Pittsburgh: University of Pittsburgh Press.

Woodward, J. (2003b). *Making things happen. A theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2018). Some varieties of non-causal explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 117–138). Oxford: Oxford University Press.

Ylikoski, P. (2009). The illusion of depth of understanding in science. In H. W. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding. Philosophical perspectives* (pp. 100–119). Pittsburgh: University of Pittsburgh Press.

Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis*, *78*(2), 277–297.

Ylikoski, P. (2014). Agent-based simulation and sociological understanding. *Perspectives on Science*, *22*(3), 318–335.

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19–36.

Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, *148*(2), 201–219.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.