Idealizations and Understanding
Much Ado About Nothing?

Sullivan, Emily; Khalifa, Kareem

DOI
10.1080/00048402.2018.1564337

Publication date
2019

Document Version
Final published version

Published in
Australasian Journal of Philosophy

Australasian Journal of Philosophy

ISSN: 0004-8402 (Print) 1471-6828 (Online) Journal homepage: https://www.tandfonline.com/loi/rajp20

Idealizations and Understanding: Much Ado About Nothing?

Emily Sullivan & Kareem Khalifa

To cite this article: Emily Sullivan & Kareem Khalifa (2019): Idealizations and Understanding: Much Ado About Nothing?, Australasian Journal of Philosophy, DOI: 10.1080/00048402.2018.1564337

To link to this article: https://doi.org/10.1080/00048402.2018.1564337

Published online: 22 Jan 2019.

AUSTRALASIAN JOURNAL OF PHILOSOPHY
https://doi.org/10.1080/00048402.2018.1564337

ARTICLE

OPEN ACCESS Check for updates

Idealizations and Understanding: Much Ado About Nothing?

Emily Sullivan[a] and Kareem Khalifa[b]

[a]Delft University of Technology; [b]Middlebury College

ABSTRACT

Because idealizations frequently advance scientific understanding, many claim that falsehoods play an epistemic role. In this paper, we argue that these positions greatly overstate idealizations' import for understanding. We introduce work on epistemic value to the debate surrounding idealizations and understanding, arguing that idealizations *qua* falsehoods confer only non-epistemic value to understanding. We argue for this claim by criticizing the leading accounts of how idealizations provide understanding. For each of these approaches, we show that: (a) idealizations' false components promote only convenience instead of understanding and (b) only the true components of idealizations have epistemic value.

ARTICLE HISTORY Received 25 August 2017; Revised 9 November 2018; Accepted 18 November 2018

KEYWORDS understanding; idealization; epistemic value; truth; models

1. Introduction

Scientists' use of idealizations raises a well-known puzzle: how can falsehoods be epistemically beneficial? Several philosophers answer this question by appealing to idealizations' contributions to understanding [Elgin 2004; Bokulich 2008; Rohwer and Rice 2013; Strevens 2017]. Conversely, these same contributions motivate the idea that understanding is distinct from more shop-worn philosophical concepts such as knowledge and explanation, which typically demand greater accuracy.

This paper disrupts this symbiosis. We argue that any epistemic value that idealizations have is not because of the understanding that they provide. Consequently, philosophers seeking either to redeem idealizations' epistemic value or to advertise understanding's importance should look elsewhere. We begin by identifying the dominant view concerning idealizations' epistemic value and proposing an alternative (Section 2). Sections 3 through 5 then show that, despite their authors' intentions, the leading accounts of how idealizations provide understanding can be reinterpreted so that understanding confers no epistemic value upon falsehoods. Instead, we will argue, the falsehoods need only be conveniences that aid in easing calculations and making things salient.

2. Idealizations, Understanding, and Epistemic Value

Understanding has garnered interest both from epistemologists and from philosophers of science [Grimm, Baumberger, and Ammon 2017]. Unsurprisingly, these different philosophical sub-disciplines assign greater currency to different lines of argument. Epistemologists are especially interested in how understanding enjoys a distinctive epistemic value that knowledge lacks [Kvanvig 2003; Pritchard 2010]. Our argument applies concepts concerning epistemic value to discussions in the philosophy of science concerning idealizations. For epistemologists, this extends their concepts to more concrete examples from scientific practice. For philosophers of science, our framework suggests a way of unifying core features of several realist [McMullin 1985; Mäki 2009] and deflationary [Nowak 1992; Alexandrova 2008] approaches to idealization, and opens lines of communication to epistemologists. In this manner, we hope that our arguments engage philosophers of diverse orientations.

For some, idealizations challenge the idea that understanding is factive [Elgin 2004]; others demur [Mizrahi 2012]. However, the distinction between factive and non-factive understanding enjoys little consensus or precision. More importantly, many arguments discussed below come from self-described factivists who nevertheless hold that understanding's relationship to idealizations is epistemically significant [Strevens 2013; Bokulich 2016; Rice 2016].

Importantly, all of our interlocutors appear committed to idealizations' having some kind of epistemic value because of the understanding that they provide. For this reason, epistemic value, rather than understanding's factivity, grounds the relevant distinction. Since idealizations can depart significantly from the truth, imbuing them with epistemic value is surprising and interesting. For instance, 'truth monism'—the view that only true beliefs have fundamental epistemic value—enjoys a somewhat privileged status in the epistemic value literature. Even truth monism's critics frequently impose some kind of accuracy or veridicality requirement on the other epistemic goods that they tout as fundamental, such as understanding and cognitive achievements. Hence, idealizations' role in understanding may have far-reaching implications about which epistemic statuses are of fundamental epistemic value and why.

The most plausible view that accords idealizations epistemic value is this:

> *Epistemic Instrumentalism*. Some falsehoods have instrumental epistemic value because of the understanding that they provide.

On this view, idealizations are a means to obtaining understanding. Most of our interlocutors appear to endorse some version of epistemic instrumentalism. For instance, Bokulich [2016: 261] writes that idealizations are 'fictions [that] can be an effective means by which we can come to understand truths that would otherwise be difficult to grasp'. To our knowledge, only Elgin [2004] endorses a stronger view about idealizations' epistemic value, and even this is debatable. At any rate, our criticisms of epistemic instrumentalism apply to Elgin's position (see section 3), and can readily be extended to any views that take idealizations to have non-instrumental epistemic value.

By contrast, we will argue that all extant examples of idealizations providing understanding can be interpreted according to the following:

> *Non-Epistemic Account*. Any epistemic value that falsehoods have is not because of the understanding that they provide.

Since we restrict our critique to positions arguing that idealizations have epistemic value in virtue of providing understanding of empirical phenomena, we leave open the possibility that idealizations might have epistemic value because of something other than their contributions to understanding phenomena. Importantly, this falls well short of our interlocutors' aspirations for idealizations. Furthermore, we are arguing only that *extant* accounts of idealizations' role in understanding are compatible with idealizations lacking epistemic value, so we do not argue that falsehoods' role in understanding cannot possibly afford them some kind of epistemic value. Instead, we use non-epistemic accounts primarily to shift the burden of proof onto those with loftier visions of idealizations' contributions to understanding, and to highlight unexplored conceptual territory.

In what follows, we reinterpret the leading proposals concerning idealizations' role in understanding in terms of a particular kind of non-epistemic account. This non-epistemic account holds that the falsehoods within idealizations are a mere convenience that aids in easing calculations and making things salient.[1] We show how nothing is lost in this reinterpretation, even when we use the same notion of understanding as our interlocutors do. Hence, epistemic instrumentalism is unsupported.

3. Flagging Irrelevancies

One mark of understanding is discriminating information that is relevant to an object's behaviour from irrelevant information. According to Strevens [2008; 2017], idealizations frequently aid in this task by highlighting or 'flagging' explanatory irrelevancies. As we shall argue, flaggers' claims can be replicated by a non-epistemic account without loss.

As an illustration, consider a common statistical-mechanical derivation of the ideal gas law.[2] On this approach, the equation of state follows from the partition function given by a sum over all states of the system in terms of the energy E of each state:

$$Z = \sum e^{-E/kT} \tag{3.1}$$

where k is Boltzmann's constant and T is the temperature of the gas.

Assume that the system consists of N non-interacting particles. Then each particle's available phase space is proportional to the system's volume V. The partition function will thereby depend on volume as V^N:

$$Z = V^N f(T)^N \tag{3.2}$$

for some function $f(T)$. This leads to the ideal gas law:

$$PV = VkT\frac{dlnZ}{dV} = NkT \tag{3.3}$$

[1] In earlier work [Doyle et al. forthcoming], one of us—Khalifa—endorsed an epistemic instrumentalism that assumed that salience and easier calculation are of instrumental epistemic value. Khalifa now rejects this assumption.

[2] Strevens discusses Boyle's law; borrowing from Doyle et al. [forthcoming], we discuss the more encompassing ideal gas law.

where P denotes the pressure of the gas. Note that Nk is equal to nR, where n is the number of moles of gas and R is the ideal gas constant.

The assumption of non-interacting particles is false, but provides understanding of the ideal gas law. However, at low density and high temperature, particle interactions in ideal gases make no difference to the relations between pressure, temperature, molarity, and volume. Indeed, idealizations make this irrelevancy especially salient by arbitrarily fixing the irrelevant parameters to zero. Thus, representing these interactions as non-existent is an especially vivid way of communicating or flagging their irrelevance to the behaviour of ideal gases. By contrast, more accurate derivations wrongly suggest that particle interactions are relevant to the behaviour of ideal gases.

We contend that flaggers' insights can be recast in terms of a non-epistemic account without loss. In particular, we will argue that if an idealization satisfies the following structure then the role of the falsehoods in understanding is not of epistemic value, but is rather a mere convenience for creatures like us to make easier calculations and notice what is salient.

1. *Downplaying*. The idealization makes certain calculations easier to perform or certain features of the phenomenon more salient;
2. *De-Idealizing*. By either eliminating the idealization or replacing it with an approximate truth ('de-idealizing'), all of the epistemic goods purported to figure in the 'idealized' understanding could still be inferred, albeit with more complex calculations or through more demanding acts of attending to the features in question;[3] and
3. *Demythologizing*. Only the parts of the idealization that approximate their de-idealized counterparts provide the aforementioned epistemic goods.

Using the ideal gas law as an illustration, we draw out the philosophical implications of this '3D' (downplaying, de-idealizing, and demythologizing) approach to non-epistemic accounts.

Begin by downplaying the idealizations' epistemic contributions—that is, showing how idealizations promote convenience. Above, we saw that the idealization—that particles do not interact in ideal gases—flagged particle interactions as irrelevant. Flagging an irrelevancy is nothing more than making it salient. Of course, if we only showed this much, then idealizations could still be epistemically valuable. By going 2D, we show how a more accurate and de-idealized inference can replicate all of the epistemic goods billed as benefits of idealizing. Consequently, the idealization has no *distinctive* epistemic benefits. Combined, downplaying and de-idealizing show that the only reason to favour idealizations over their more veridical counterparts is their convenience.

In discussing de-idealizing, we assume that approximate truth, rather than strict truth, is a more appropriate constraint on scientific statements. Of course,

[3] On McMullin's [1985: 281] view, de-idealization implies that an idealized model 'serves as the basis for a continuing research program' in which subsequent 'models can be made more specific by eliminating simplifying assumptions' or adding more accurate assumptions. As such, his notion of de-idealization prognosticates about subsequent scientific inquiry. Note that our definition of de-idealization does not prognosticate in this way; scientists may use an idealization for perpetuity owing to its convenience.

something can be approximately true and strictly false. So, 'falsehoods' must be understood as shorthand for 'falsehoods that fall short of approximate truth'. For the purposes of this paper, approximate truths come in two flavours. In this section and the next, approximately true models represent their target systems' underlying dynamics to an arbitrarily high level of precision and accuracy. In section 5, approximately true models are causal models inferred from reliable data by using accepted statistical methods—in short, very pedestrian models devoid of idealizations. We strongly suspect that our interlocutors take their preferred idealized models to outperform these models in the understanding that they provide, and so this provides an adequate foil.

With these clarifications about approximate truth in hand, we note that de-idealization is especially acute in deriving the ideal gas law. If we forgo the assumption of non-interacting particles, then we can derive the virial equation of state directly from statistical mechanics:

$$\frac{PV}{NkT} = 1 + \frac{B}{V} + \frac{C}{V^2} + \frac{D}{V^3} + \dots \tag{3.4}$$

This expansion can be rendered arbitrarily precise by extending the equation indefinitely, with each added term being derivable from increasingly detailed and accurate assumptions about the intermolecular forces. For instance, B corresponds to interactions between pairs of molecules; C, triplets; etc. Hence, it is far more accurate than the idealized model above.[4] Recall that the ideal gas law only obtains at low density (P) and high temperatures (T). Consequently, volume (V) will be large, and so the contribution of the added terms in the virial expansion (B/V, C/V^2, etc.) will be vanishingly small, resulting in the ideal gas law. Hence, a de-idealized explanation of the ideal gas law will identify the explanatorily irrelevant particle interactions without the idealizations, albeit with greater effort.

Importantly, because de-idealized inferences effectively garner an epistemic good (in this case, a truth that figures in understanding), they are plausibly regarded as having instrumental epistemic value. Indeed, virtually every epistemic value theorist recognizes that (roughly) sound reasoning is an effective means for acquiring true beliefs. This is one reason that inferential justification is epistemically valuable. However, if idealizations were only downplayed and de-idealized, they might still be just as effective as their de-idealized counterparts in yielding the same epistemic goods. If that were so, they would be just as epistemically valuable as their more veridical neighbours.

Interestingly, non-epistemic accounts tolerate this result—but only so long as idealizations' *false* components are not epistemically valuable. Here, we turn to our non-epistemic account's third feature—namely, demythologizing—where we show that only idealizations' approximately *true* components provide the epistemic goods that figure in understanding. The only epistemic parts that are valuable for providing understanding are those that approximate de-idealized elements. In our example, had the interactions not dropped out in the de-idealized inference, the idealization would have wrongly flagged them as irrelevant. In short, the idealization delivers true beliefs about what is explanatorily irrelevant, only because it tracks with the de-idealized components of its more accurate counterpart.

[4] See (3.1) and (3.2).

Thus, all told, non-epistemic accounts entail that idealizations' only distinctive contribution is convenience, and that only their accurate components are epistemically valuable. Combined, this suggests that, when it comes to understanding, in so far as idealizations are false, they are of non-epistemic value; and in so far as they have epistemic value, they are (approximately) true. Recall that this is the core of non-epistemic accounts: falsehoods are not epistemically valuable because of their role in understanding. By contrast, epistemic instrumentalism implies that idealizations are epistemically valuable falsehoods. Indeed, this was what made idealizations epistemically interesting. Hence, non-epistemic accounts entail that idealizations' contributions to understanding fall short of the philosophical hype.

While Strevens is naturally interpreted as an epistemic instrumentalist, Elgin [2004] uses this same example to motivate a stronger view that suggests that understanding-providing falsehoods have non-instrumental epistemic value. Elgin focuses on how idealized models can exemplify certain true features of their target systems. However, such a view also amounts to making certain things salient, and thereby courts the same non-epistemic account. Consequently, our non-epistemic interpretation of the ideal gas law explains away both Strevens's and Elgin's accounts—the latter being the only viable version of 'epistemic non-instrumentalism'—without loss.

At this point, flaggers have some possible objections. For instance, they may insist that salience has epistemic value that 3D accounts do not capture. This occasions three replies. First, our goal is simply to stake out idealization arguments' *current* liabilities. It is certainly *possible* to devise epistemically substantive accounts of salience, but clearly the onus is on our interlocutors to do so. Second, it is unclear that salience has epistemic value, as products of ignorance, illusion, and bias can be just as salient as the products of careful and reliable inquiry, but this does not make the former any less epistemically pernicious. Indeed, making these epistemic pollutants salient can frequently increase their harmfulness (as happens in ideologically saturated information bubbles.) Hence, nothing about salience *per se* appears epistemically valuable; it all depends on *what* is made salient.

Third, the reasoning underlying this objection is unclear. We suspect that it assumes this:

> (IEV) Accepting proposition *A* is of *instrumental epistemic value* if, for an agent *S* and epistemic good *G*, had *S* not accepted *A*, *S* would have been less likely to possess epistemic good *G*.

Presumably, the greater cognitive effort of the de-idealized inference would have greatly lowered the chances of scientists identifying the irrelevance of particle interactions to the ideal gas law. As the objection suggests, this would entail that salience is epistemically valuable.

However, IEV is implausible. Imagine benighted people who only recognize the irrelevance of Einstein's IQ to ideal gas behaviour by accepting the silly idealization that ideal gas particles are a community of sentient beings that has banished Einstein. If IEV is correct, then this silly idealization is of instrumental epistemic value. This ties instrumental epistemic value too closely to agents' idiosyncrasies.

By contrast, the 3D account's appeal to de-idealized inferences suggests a more defensible notion of instrumental epistemic value. As noted above, it rests on the

relatively uncontroversial idea that sound reasoning is a paradigmatic kind of instrumental epistemic value. In lieu of IEV, this suggests the following:

> (IEV*) Accepting proposition *A* is of instrumental epistemic value if *A* is approximately true and has epistemically valuable propositions as some of its consequences.

IEV* departs from IEV by tethering epistemic value to objective epistemic features, such as truth. Hence, it readily rules out the silly idealization, as it clearly is not approximately true. IEV*'s superiority over IEV suggests that sound reasoning, rather than convenience, is a more promising locus of instrumental epistemic value.

Perhaps our critics would respond by strengthening IEV by restricting it to *epistemically responsible* agents, as scientists typically are. Intuitively, this would block our 'Einstein IQ example'. However, the proposal under consideration would require epistemically responsible agents to knowingly make unsound inferences, which seems rather *ad hoc*. By contrast, the 3D account plausibly suggests that pragmatic considerations (such as convenience) sometimes override an agent's epistemic responsibilities.

Flaggers may also reply that our objections merely reflect the ideal gas law's idiosyncrasies. Once again, our goal is simply to stake out current liabilities of arguments pegging idealizations' epistemic value to understanding. Hence, it is certainly *possible* to find different examples that vindicate flagging approaches, but clearly the onus is on flaggers to provide an *actual* example. Parallel points apply to other epistemic instrumentalists.

4. Explanatory Fictionalism

Perhaps another understanding-providing relationship would avoid flaggers' difficulties. While theories of understanding have taken root only in the last decade or so, virtually everyone agrees that explanations provide understanding. Hence, cases in which idealized models explain some phenomenon appear especially promising for epistemic instrumentalists. Bokulich [2008] provides several examples of these 'explanatory fictions'.[5] Most of her examples involve explanations of quantum phenomena that invoke classical assumptions. Physicists prefer these 'semiclassical' explanations to those that advert only to quantum mechanics.

Consider one such explanation. Quantum billiards are the quantum equivalent of classical billiards systems with chaotic geometries. In some of these systems, such as the 'stadium billiard', only a few patterns repeat themselves. These trajectories are known as periodic orbits, and frequently exhibit simple patterns (such as bow-tie, rectangular, double-diamond patterns). Somewhat surprisingly, the probability density of many wavefunctions is strongly localized around a few classical periodic orbits. This phenomenon is known as wavefunction scarring. Orbits will recur over time if and only if Gaussian wavepackets are launched on certain classical orbits. The explanation involves the fiction of classical orbits, for classical trajectories simultaneously have a well-defined position and momentum, and thereby violate Heisenberg's uncertainty principle.

Bokulich claims that semiclassical explanations provide superior understanding because they answer what-if-things-had-been-different (w-) questions that their

[5] See also Cartwright [1983], Batterman and Rice [2014], and Rice [2015].

purely quantum-mechanical counterparts cannot. More precisely, Bokulich [ibid.: 153] states, '[Since] there are different *kinds* of w-questions one can ask about the explanandum system, then, I argue, there can be situations in which *less* fundamental theories can provide deeper explanations than more fundamental theories.' Here, of course, semiclassical explanations are supposed to be the less fundamental but deeper explanations.

Among the authors considered, only Bokulich anticipates the objection that idealizations can be subsumed under non-epistemic accounts. For example, she writes, 'these classical structures (such as periodic orbits), are not simply useful calculational devices' [ibid.: 132]. Moreover, Bokulich's focus on explanation leaves her account less susceptible to non-epistemic accounts than flaggers (although see Schindler [2014] and King [2016] for criticisms on this front.) Nevertheless, using the 3D strategy, we shall now show that non-epistemic accounts can subsume explanatory fictionalism without loss.

First, semiclassical models clearly promote convenience, and hence can be downplayed. Virtually everyone recognizes that they involve less demanding calculations, even if there is disagreement as to whether they provide more than that. Furthermore, semiclassical models are widely recognized as making certain explanatory factors salient. Consider the following quote from some of the leading semiclassical theorists, whom Bokulich [2008: 154] cites approvingly:

> The high dimensionality of the problem combined with the vast density of states can make the [full quantum-mechanical] calculations cumbersome and elaborate … Furthermore, the problem of understanding the structure of the quantum solutions still remains after solving the Schrödinger equation. Again, simple interpretation of classical and semiclassical methods assists in illuminating the structure of solutions [Wintgen, Richter, and Tanner 1992: 19].

If we take 'illuminating' to be synonymous with 'making salient', the idealizations' contribution to understanding is effectively downplayed.

Turn next to de-idealization. At first blush, Bokulich appears to take special measures to show that idealized models yield distinctive epistemic benefits [2008: 151]:

> My claim is not that there is no purely quantum-mechanical explanation for this phenomenon, but rather that such an explanation, that omits reference to classical orbits, is deficient. Although one can 'deduce' the phenomenon of wavefunction scarring by numerically solving the Schrödinger equation, such an explanation fails to provide adequate *understanding* of this phenomenon.

Both semiclassical and purely quantum-mechanical models explain wavefunction scarring. Thus, the understanding referred to here must be something over and above an explanation that involves more manageable calculations or that makes certain features more salient. As mentioned above, Bokulich takes the greater understanding to be from semiclassical models' capacity to answer unique w-questions.

However, this claim faces two objections. First, Bokulich asserts this claim without argument. Specifically, sound arguments for this thesis would present at least one w-question that semiclassical explanations are uniquely positioned to answer, yet Bokulich provides no such w-question. Second, her [2008: 153] arguments for why de-idealized inferences fail to answer w-questions are flawed:

> the complexity of the high-energy eigenstates of the quantum stadium billiard also means that the Schrödinger equation for this system can only be solved numerically. Simply showing, as a purely quantum-mechanical D-N explanation does, that certain eigenstate

> morphologies follow from the dynamical law with the relevant boundary conditions for this system, … gives little physical insight into this phenomenon.

Here, Bokulich assumes that anything that can only be solved numerically cannot answer a w-question. However, it is unclear why that should be so. For instance, if different numerical values of the independent variables yield different numerical values for the dependent variables, then answers to w-questions can be purely numerical.

Perhaps, instead of w-questions, Bokulich's claim that semiclassical explanations provide greater 'physical insight' than their quantum-mechanical counterparts points to a kind of understanding that only explanatory fictions provide. If so, this would redeem her proposal. However, 'physical insight' is never unpacked, and appears to be a mere synonym for 'understanding'.[6] So, at best, this proposal requires elaboration; at worst, it is circular. Consequently, we see no strong reason to grant a unique epistemic good that semiclassical explanations provide. So, the balance of arguments favours the idea that explanatory fictions' distinctive contribution to understanding is that of convenience.

Finally, we demythologize semiclassical explanations. Bokulich describes the relevant de-idealized counterparts as 'purely quantum-mechanical explanations'. Much like last section's de-idealized derivation shows why particle interactions do not make a difference to the ideal gas law, physicists can now provide purely quantum-mechanical derivations of quantum billiards that account for the reliability of semiclassical models by showing how wavepackets not launched on classical orbits do not make a difference to scarring [Georgeot and Prange 1995]. As before, this shows that semiclassical models answer w-questions only because they approximate their de-idealized cousins. Indeed, physicists frequently describe these models as 'semiclassical approximations'.

We conclude by anticipating two possible replies to our non-epistemic recasting of explanatory fictionalism. First, purely quantum-mechanical derivations of wavefunction scarring admit of no analytic solution and are thereby intractable. One may object that this precludes the possibility of a de-idealized inference, which, in turn, undermines the possibility of demythologizing. However, de-idealized inferences need not be deductive, and, without this requirement, the lack of analytical solutions is moot. Furthermore, Bokulich's own writing suggest some reasons to be wary of denying the possibility of de-idealization. As she argues, all explanatory fictions must have a 'translation key … from statements about the fictions to statements about the underlying structures or causes of the explanandum phenomenon' [2012: 735]. Such translation keys distinguish fictions that are genuinely explanatory from mistaken theories that fortuitously save the phenomena (Ptolemaic astronomy, for example). This suggests something very close to demythologizing: in effect, Bokulich is arguing that if the fictions could not be reinterpreted as approximate truths, then they would not be explanatory. Indeed, if the false parts of the model were to deliver the results without approximating the de-idealized equations, it would be fair to worry that the non-analytic results of the former were mere

[6] On the same page, Bokulich [2008: 153] concludes her discussion of scarring with the claim, 'quantum-mechanical explanations are deficient [because] they fail to provide adequate physical insight into, or understanding of, these phenomena.' This suggests that 'physical insight' and 'understanding' are synonyms.

artefacts of the modelling process. Hence, demythologizing and (by implication) de-idealization appear necessary for falsehoods to provide understanding.

A second objection charges 3D accounts with failing to respect the strong intuitions that semiclassical models provide greater understanding than purely quantum-mechanical models do. In response, we propose that these intuitions result from conflating two kinds of understanding. Convenience is conducive to *understanding what a model says*—hereafter called *legibility*.[7] Models expressed in difficult notations (including foreign languages) are illegible. By contrast, those who tout falsehoods' epistemic value all claim to be discussing scientific *understanding of phenomena*. A model's legibility does not entail that it provides understanding of phenomena, as many crackpot models are legible. Conversely, inconvenient de-idealized models—such as the purely quantum-mechanical model of wavefunction scarring—tend to be illegible because they are expressed in complex idioms, but this does not preclude them from providing understanding of phenomena. Thus, 3D accounts suggest a powerful error theory to dispel our interlocutors' intuitions: they conflate a model's legibility with its capacity to provide understanding of phenomena.

One might retort that illegible models fail to provide understanding that is accessible to inquirers with our kinds of cognitive limitations, and that is the only kind of understanding that matters. However, non-epistemic accounts are consistent with the following set of plausible claims: (a) models only provide understanding of phenomena to scientists if the scientists find these models legible; (b) sometimes idealized models are legible and their de-idealized counterparts are not; and (c) understanding phenomena is epistemically valuable. Indeed, (a) and (b) make our error theory all the more compelling, for if the connection between legibility and understanding phenomena is as tight as they imply, then it is easy to see how these two kinds of understanding could be conflated. However, undermining our non-epistemic account requires further arguments that (d) legibility is epistemically valuable. No such arguments for (d) have been offered. Furthermore, for 3D accounts, legibility need not involve more than ease of calculation and salience, which seem especially implausible as epistemic goods. Hence, even if we grant that illegible models do not provide 'human' understanding, this does not show that idealizations have epistemic value because of their role in understanding.

Thus, all told, Bokulich has not shown that semiclassical explanations provide understanding that quantum-mechanical explanations fail to provide. Holding that quantum-mechanical explanations answer at least as many w-questions as semiclassical explanations, but involve more torturous calculations and make the underlying dynamics less salient, is equally defensible. Moreover, in defending this view, two general strategies for non-epistemic accounts have emerged—arguments that demythologizing is pivotal to distinguishing genuine understanding from mere artefacts of a model, and an error theory involving legibility.

[7] Legibility is a broad notion that includes simple linguistic understanding of what the model *says*. This need not include a loftier facility with the model, such as being able to manipulate the model or draw inferences from it, that some have tied to understanding models [Kuorikoski and Ylikoski 2015].

5. Idealization and Causal Understanding

Explanatory fictionalists claim that idealizations provide understanding by explaining certain phenomena. In contrast, our remaining epistemic instrumentalists claim that idealizations offer a less direct means of identifying causal influences. These proposals—dubbed 'contrastivism', 'isolationism', and 'modalism'—differ only in *how* idealizations provide this causal information. In what follows, we argue that none of these positions redeems idealizations' epistemic value. As an added bonus, the 3D account unifies these approaches into a single framework for interpreting idealizations' role in causal understanding.[8]

5.1. Contrastivism

Many explanations are of the form '*p* rather than *q* because *r* rather than *s*.' According to contrastivists, idealized models provide understanding when they are compared to more realistic models of the same phenomenon [Diéguez 2013; Hindriks 2013]. Roughly, more accurate models provide values for *p* and *r*, and idealized models provide values for *q* and *s*.

Consider the Modigliani-Miller (M-M) theorem from economics. The model states that a firm's value is independent of its debt-equity ratio [Modigliani and Miller 1958, 1963]. The M-M model is highly idealized: it assumes perfect financial markets, in which all parties have equal information, actors are completely rational, and transactions have no costs, such as taxes. Real world financial markets lack all of these properties. Furthermore, real-world firms' value *depends* on their debt-equity ratio. Despite this, economists still use the M-M model to understand firms' value. Understanding is achieved by *comparing* the M-M model with more realistic models, such as models that include the cost of taxes [Kraus and Litzenberger 1973; Jensen and Meckling 1976]. The differences between the models provide answers to the contrastive question, 'Why is a firm's value dependent on (rather than independent of) its financial structure?' Answers cite factors present in real world markets but absent from perfect markets. In this sense, the idealized contrast allows modellers to identify important causal mechanisms present in real-world systems. The M-M model provides a baseline from which economists can assess real-world factors' causal impact.

These idealizations would appear to function as an epistemically valuable means to achieving understanding, since causal information provides understanding. Indeed, one of the M-M model's architects expresses something in the spirit of epistemic instrumentalism: 'showing what *doesn't* matter can also show, by implication, what *does*' [Miller 1988: 100]. Similarly, in discussing the M-M model, Hindriks [2013: 528] asserts that the discovery of differences between idealized and increasingly realistic models is 'the (or at least a) way to achieve understanding of a mechanism by means of models'.

However, contrastivists succumb to non-epistemic accounting. Causal histories can stretch far back in time, span several levels of analysis, and, in several cases, are influenced by many exogenous factors. By providing a compact representation

[8] Indeed, the framework can be extended to many idealizations used to identify non-causal explanatory relations.

of how the system would behave when certain factors are abstracted away, idealized models alleviate the inconveniences incurred by this blooming, buzzing, confusion of causal details. Hence, contrastive idealizations can be downplayed.

Furthermore, de-idealized models can furnish the same contrasts. For instance, using regression techniques, economists have acquired the same contrastive and counterfactual information as the M-M model by observing that debt-equity ratio makes a statistically significant difference to a company's value when controlling for the relevant confounding variables [McConnell and Servaes 1995]. Unlike the M-M model, this information is reaped empirically rather than theoretically. Importantly for our purposes, such empirical information involves no idealizations.

Furthermore, these de-idealized models underwrite the use of their idealized counterparts, and thereby debunk the myth that they are epistemically valuable falsehoods. Had the empirical results been otherwise, the M-M model would be (as its authors had initially hoped) a realistic model, and hence would serve exactly the *opposite* role in the relevant contrastive explanations. In other words, the following are empirical questions that are answered without appeal to idealization. (a) Do firms behave as the M-M model describes them? (b) Are the differences between observation and the M-M model causally relevant? If the empirical results provide an affirmative answer to (a) or a negative answer to (b), then the M-M model could not function as a contrast or causal baseline.

This last point highlights a general problem that not only plagues contrastivists, but, as we will see, also afflicts isolationists and modalists. Idealizations tend to spring from theoretical motivations. In order for the examples in this section to support epistemic instrumentalism, these theoretical motivations must be sufficient grounds for inferring causes. However, scientific practice and common sense belie this 'causal rationalism:' causal claims need to be inferred from empirical evidence. But steering clear of causal rationalism clears the path for empirical, de-idealized, causal inferences that demythologize the idealizations' success in tracking causal information. This, of course, strongly suggests a 3D account. Hence, the prospects of grounding idealizations' epistemic value in causal understanding appear bleak.

5.2. Isolationism

Isolationists claim that idealizations can identify causal factors by controlling for noise, but, unlike contrastivists, they do not require comparisons between idealized models and more accurate models [Cartwright 1989; Jones 2005; Mäki 2009]. However, like contrastivists, they face the hard choice between causal rationalism and the 3D account. Quite reasonably, they have gravitated toward the latter.

Consider Schelling's [1971] checkerboard model of how types of agents acting on preferences influence spatial proximity. Schelling himself was interested in understanding why many human populations are segregated and how the phenomenon could be explained by the preferences of individuals. The model is a grid with two types of actors, A and B, where both types act on one simple preference—that at least 30% of their neighbours are of the same type. If this preference is not met, the actor moves to the nearest unoccupied space. This simulation is run numerous times until an equilibrium is met. The equilibrium result is a segregated board.

The model is highly idealized. Most drastically, actors have the freedom and means to move. This contrasts starkly with the cities that Schelling analyses, where institutional racism causes much of the segregation. As an illustration, Baltimore's segregation is partly a result of its history with state-sponsored segregation policies that included explicit laws prohibiting integration [Power 1983]. So, while Schelling's model poorly explains the particular history of Baltimore and cities like it, it still has been influential because it seeks to isolate a common cause among diverse cities. Isolationists hold that Schelling's model removes real-world influences, such as integration laws, moving-costs, and red-lining. The resulting model isolates the causal influences of individual preferences on segregration. Importantly, the false assumptions allegedly enable the isolation of an actual causal influence that underlies many diverse segregated systems. Thus, isolationism is a good candidate for redeeming the instrumental epistemic value of falsehoods. However, we will argue that a non-epistemic account can subsume it.

First, Schelling's idealizations can be downplayed as largely simplifying assumptions. In actuality, people incur moving costs, prefer to live in places not only because of neighbourly similarities, and frequently operate with explicitly racial prejudice and within racist institutions. Adding these further considerations would greatly increase the calculational burdens needed to run the aforementioned simulation.

Second, a de-idealized version of Schelling's model readily delivers the same information. Schelling assumes that actors move based on a set preference level shared among all actors. By contrast, other models use survey data to set the preference levels to those of real individuals. Complexities are also introduced where in-group preferences differ depending on age, income, and education. These models confirm Schelling's thesis that in-group preference is a factor in segregation [Clark 1991, 1992]. Furthermore, by accounting for individuals' racial *prejudices* in addition to their positive preferences for their own race, other models include complexities that help to determine the precise causal scope that in-group preferences can have [Bobo and Zubrinsky 1996]. All told, such models more accurately represent the behaviours of people in cities across the US, but still show that segregation can occur because of individual preferences that are independent of institutionalized segregation practices. This means that a less convenient but more accurate model can deliver the same epistemic goods as Schelling's model does.

Finally, we demythologize the Schelling-model's epistemic value. Suppose that the aforementioned empirical models provided decisive empirical evidence that all segregation is the result of racist institutions or individual racial prejudice, so that Schelling's mechanisms for segregation were never realized in any real-world system. Then, quite clearly, we would have no reason to think that Schelling's model uncovers any actual causes of segregation. This shows that Schelling's model provides epistemic goods only in so far as it approximates empirical facts about actual segregated communities. Indeed, Mäki [2009] gestures toward this sort of demythologizing move, while simultaneously distancing himself from causal rationalism. He says, of Schelling's model [ibid.: 40]:

> if we wish to attribute a stronger epistemic form of credibility … it is not sufficient to have some general ontological convictions … more specific information is needed, based on empirical inquires in specific cases. Schelling himself did not provide this information, but probably thought favourably of taking these further steps.

Thus, all told, idealizations tell us how causes act in isolation only in so far as empirical information delivers the same verdict. It is these de-idealized inferences that provide understanding in these cases; the falsehoods are mere conveniences.

5.3. Modalists

Thus far, being subsumed by 3D accounts has seemed more compelling than endorsing causal rationalism. This is because both contrastivists and isolationists claim that idealizations provide understanding of the actual causes. However, perhaps epistemic instrumentalists should seek refuge in causal rationalism. Our last target, modalism, does just this, by arguing that idealizations provide understanding of possible causes [Rohwer and Rice 2013; Rice 2015, 2016]. Presumably, if we are seeking only to identify *possible* causal influences, this can be successful even in the absence of empirical information. However, while modalists avoid causal rationalism's problems, the cost is an abandonment of epistemic instrumentalism.

As we saw above, Schelling's model explains the possibility of segregation occurring without any imbedded economic and political institutional racism. If we take this possibility as the *explanandum*, it expresses a true claim: it is *possible* that a population could become segregated based on individual preferences. Furthermore, the *explanans* could also contain a similar 'modal hedge': segregation without explicit racism is possible partly because it is *also possible* that agents incur no costs for moving, etc. Furthermore, this modal hedging applies whether the model *explains* a possibility or merely *justifies* a possibility-claim as Rohwer and Rice [2013] sometimes suggest.

Regardless, in such scenarios, no falsehoods furnish understanding. The key point is that claims of the form 'it is possible that p' can be true even when 'it is actually p' is false. Thus, true claims (about possibilities) are used as a means of understanding other true claims (about other possibilities.) This is precisely what non-epistemic accounts require. By contrast, epistemic instrumentalism requires idealizations to be epistemically valuable *falsehoods*. Hence, if modalists claim only that idealizations promote understanding by providing truths about unactualized possibilities, then only non-epistemic accounts provide a plausible interpretation of their epistemic value.[9]

Modalists could reply that the *explanandum* is not about a mere possibility, but an actual real-world phenomenon. However, this either yields an indefensible form of causal rationalism, or recapitulates the previous section's non-epistemic account. In the former case, a theoretical possibility is mistaken as suitable grounds for inferring an actual cause. In the latter case, the empirical de-idealized inferences provide the same causal information, and also demythologize the idealization's success by circumscribing the conditions wherein the idealization has correctly latched on to actual causes.

[9] Rohwer and Rice [2013] also talk about models teaching scientists about 'hypothetical' scenarios. Similarly, others take models to provide understanding via *potential* explanations [Rice 2016; Reutlinger, Hangleiter, and Hartmann forthcoming]. These discussions do not have any consequential differences between hypothetical scenarios, possible scenarios, and potential explanations. Hence, our objections also apply to these proposals.

6. Conclusion

In summary, we have argued that the most promising accounts of how idealizations provide understanding—by flagging irrelevancies, explaining, structuring contrastive explanations, isolating causes, and imparting modal information—can be reinterpreted as non-epistemic accounts. On this non-epistemic view, idealizations merely alleviate the inconveniences required to use more accurate representations, either by simplifying calculations or by highlighting features that would otherwise be more difficult to spot. Furthermore, in so far as these easier tasks result in understanding, it is because a relatively accurate or de-idealized inference underwrites the use of the idealization. Hence, nothing *per se* about idealizations' role in understanding is of epistemic value.

Furthermore, we have granted our interlocutors three sizable affordances. First, our 3D approach is only one of several possible non-epistemic accounts that could be pitched against the claim that idealizations have epistemic value. Second, even among 3D positions, we have restricted ourselves to the claim that idealizations promote convenience only in two ways (calculation and salience.) Third, we have granted our interlocutors their own conceptions of understanding. Should those conceptions wither under further scrutiny, their arguments would face further obstacles. Hence, our interlocutors can be thwarted in more ways than we have considered here.

However, we also see many constructive lessons and tasks falling out of our discussion. For instance, future work can cement the points made here by offering a positive argument for non-epistemic accounts. Here it is worth noting that some philosophers take easy calculation to be a mark of understanding [De Regt and Dieks 2005], and that others take convenience to track closely with humans' sense of understanding [Trout 2002]. While these two views have been in tension, non-epistemic accounts suggest a possible unifying framework.

Furthermore, in making our argument, we have forged connections between two unlikely bedfellows—the literature on epistemic value and the literature on idealizations. These connections were sufficient for the tasks at hand, but a good deal more can be said here, and it would provide an exciting forum for epistemologists and philosophers of science to exchange ideas. For instance, in an influential essay, Weisberg [2007] provides a typology of idealizations motivated partly by the 'representational ideals of theorizing', which include considerations such as precision, accuracy, completeness, simplicity, causal invariance, empirical adequacy, and generality. It is natural to read Weisberg as promoting a kind of epistemic value pluralism, populated by considerations that are informed by scientific practice. It would be interesting to see how truth monists would engage his and related work. To that end, non-epistemic accounts might well bear on some of Weisberg's claims that idealizations promote representational ideals.

Yet another challenge is to see if idealizations have epistemic value for some reason *other than* the ones considered here. This might take two forms. The more conservative approach would continue the kinds of projects of our interlocutors, by offering new variations on the idea that idealizations figure in understanding in some epistemically valuable way. A bolder approach—and one that is compatible with non-epistemic accounts—is that any epistemic value that idealizations have is because of something *altogether different* than how they figure in understanding.

We think that the time has come to give this bolder alternative more serious consideration.[10]

Funding

Emily Sullivan wrote part of this paper while being supported by a subaward agreement from the University of Connecticut with funds provided by Grant No. 58942 from John Templeton Foundation. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of UConn or John Templeton Foundation.

References

Alexandrova, Anna 2008. Making Models Count, *Philosophy of Science* 75/3: 383–404.
Batterman, Robert W. and Collin C. Rice 2014. Minimal Model Explanations, *Philosophy of Science* 81/3: 349–76.
Bobo, Lawrence and Camille L. Zubrinsky 1996. Attitudes on Residential Integration: Perceived Status Differences, Mere In-Group Preference, or Racial Prejudice? *Social Forces* 74/3: 883–909.
Bokulich, Alisa 2008. *Reexamining the Quantum-Classical Relation: Beyond Reductionism and Pluralism*, Cambridge: Cambridge University Press.
Bokulich, Alisa 2012. Distinguishing Explanatory from Nonexplanatory Fictions, *Philosophy of Science* 79/5: 725–37.
Bokulich, Alisa 2016. Fiction as a Vehicle for Truth: Moving Beyond the Ontic Conception, *The Monist* 99/3: 260–79.
Cartwright, Nancy 1983. *How the Laws of Physics Lie*, Oxford: Clarendon Press.
Cartwright, Nancy 1989. *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.
Clark, William A.V. 1991. Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model, *Demography* 28/1: 1–19.
Clark, William A.V. 1992. Residential Preferences and Residential Choices in a Multiethnic Context, *Demography* 29/3: 451–66.
De Regt, Henk W. and Dennis Dieks 2005. A Contextual Approach to Scientific Understanding, *Synthese* 144/1: 137–70.
Diéguez, Antonio 2013. When Do Models Provide Genuine Understanding, and Why Does It Matter? *History and Philosophy of the Life Sciences 35*/4: 599–620.
Doyle, Yannick, Spencer Egan, Noah Graham, and Kareem Khalifa, forthcoming. Non-Factive Understanding: A Statement and Defense, *Journal for General Philosophy of Science*.
Elgin, Catherine Z. 2004. True Enough, *Philosophical Issues* 14/1: 113–31.
Georgeot, Bertrand and Richard E. Prange 1995. Exact and Quasiclassical Fredholm Solutions of Quantum Billiards, *Physical Review Letters* 74/15: 2851–4.
Grimm, Stephen R., Christoph Baumberger, and Sabine Ammon, (eds) 2017. *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, New York: Routledge.
Hindriks, Frank 2013. Explanation, Understanding, and Unrealistic Models, *Studies in History and Philosophy of Science Part A* 44/3: 523–31.
Jensen, Michael C. and William H. Meckling 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics* 3/4: 305–60.
Jones, Martin R. 2005. Idealization and Abstraction: A Framework, *Poznań Studies in the Philosophy of the Sciences and the Humanities* 86/1: 173–218.
King, Martin 2016. On Structural Accounts of Model-Explanations, *Synthese* 193/9: 2761–78.

[10] We would like to thank and the audiences at the University of Albany, Models and Simulations 8 Conference at the University of South Carolina, and the Institute for Philosophy and History of Science and Technology in Paris for their feedback on earlier versions of this paper. We owe special thanks to Nahuel Sznajderhaus for insightful comments concerning semiclassical explanations.

Kraus, Alan and Robert H. Litzenberger 1973. A State-Preference Model of Optimal Financial Leverage, *The Journal of Finance* 28/4: 911–22.
Kuorikoski, Jaakko and Petri Ylikoski 2015. External Representations and Scientific Understanding, *Synthese* 192/12: 3817–37.
Kvanvig, Jonathan L. 2003. *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.
Mäki, Uskali 2009. MISSing the World. Models as Isolations and Credible Surrogate Systems, *Erkenntnis* 70/1: 29–43.
McConnell, John J. and Henri Servaes 1995. Equity Ownership and the Two Faces of Debt, *Journal of Financial Economics* 39/1: 131–57.
McMullin, Ernan 1985. Galilean Idealization, *Studies in History and Philosophy of Science Part A* 16/3: 247–73.
Miller, Merton H. 1988. The Modigliani-Miller Propositions after Thirty Years, *Journal of Economic Perspectives* 2/4: 99–120.
Mizrahi, Moti 2012. Idealizations and Scientific Understanding, *Philosophical Studies* 160/2: 237–52.
Modigliani, Franco and Merton H. Miller 1958. The Cost of Capital, Corporation Finance and the Theory of Investment, *The American Economic Review* 48/3: 261–97.
Modigliani, Franco and Merton H. Miller 1963. Corporate Income Taxes and the Cost of Capital: A Correction, *The American Economic Review* 53/3: 433–43.
Nowak, Leszek 1992. The Idealizational Approach to Science: A Survey, *Poznań Studies in the Philosophy of the Sciences and the Humanities* 25/1: 9–63.
Power, Garrett 1983. Apartheid Baltimore Style: The Residential Segregation Ordinances of 1910–1913, *Maryland Law Review* 42/2: 289–328.
Pritchard, Duncan 2010. Knowledge and Understanding, in *The Nature and Value of Knowledge: Three Investigations*, ed. D. Pritchard, A. Millar, and A. Haddock, Oxford: Oxford University Press: 3–90.
Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann, forthcoming. Understanding (with) Toy Models, *The British Journal for the Philosophy of Science*.
Rice, Collin C. 2015. Moving Beyond Causes: Optimality Models and Scientific Explanation, *Noûs* 49/3: 589–615.
Rice, Collin C. 2016. Factive Scientific Understanding without Accurate Representation, *Biology & Philosophy* 31/1: 81–102.
Rohwer, Yasha and Collin C. Rice 2013. Hypothetical Pattern Idealization and Explanatory Models, *Philosophy of Science* 80/3: 334–55.
Schelling, Thomas C. 1971. Dynamic Models of Segregation, *The Journal of Mathematical Sociology* 1/2: 143–86.
Schindler, Samuel 2014. Explanatory Fictions—for Real? *Synthese* 191/8: 1741–55.
Strevens, Michael 2008. *Depth: An Account of Scientific Explanation*, Cambridge, MA: Harvard University Press.
Strevens, Michael 2013. No Understanding without Explanation, *Studies in History and Philosophy of Science Part A* 44/3: 510–15.
Strevens, Michael 2017. How Idealizations Provide Understanding, in *Explaining Understanding: New Essays in Epistemology and Philosophy of Science*, ed. Stephen R. Grimm, Christoph Baumberger, and Sabine Ammon, New York: Routledge: 37–49.
Trout, J.D. 2002. Scientific Explanation and the Sense of Understanding, *Philosophy of Science* 69/2: 212–33.
Weisberg, Michael 2007. Three Kinds of Idealization, *The Journal of Philosophy* 104/12: 639–59.
Wintgen, Dieter, Klaus Richter, and Gregor Tanner 1992. The Semiclassical Helium Atom, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 2/1: 19–33.