Idealization and Many Aims

Angela Potochnik

Abstract

In this paper, I first outline the view developed in my recent book on the role of idealization in scientific understanding. I discuss how this view leads to the recognition of a number of kinds of variability among scientific representations, including variability introduced by the many different aims of scientific projects. I then argue that the role of idealization in securing understanding distances understanding from truth, but that this understanding nonetheless gives rise to scientific knowledge. This discussion will clarify how my view relates to three other recent books on understanding by Henk de Regt, Catherine Elgin, and Kareem Khalifa.

Forthcoming in *Philosophy of Science: Proceedings of PSA2018*

1 Introduction

Scientific understanding is something of a topic du jour in philosophy of science. One good reason for this is that the nature and status of scientific understanding had been pretty well ignored until recently, for the most part treated merely as a tag-along to scientific explanation. Explicitly considering the nature of our scientific understanding, and how that understanding relates to the more thoroughly discussed topics of scientific knowledge and scientific explanation, provides an opportunity to more fully explore how scientific practices relate to the practitioners of science—the epistemic subjects in pursuit of understanding. And so it is that considerations of scientific understanding naturally give rise to discussion of idealization as well.

In this paper, I outline the view of idealization and understanding developed in (Potochnik, 2017) and position this view in relation to Henk de Regt's, Catherine Elgin's, and Kareem Khalifa's recent books. In Section 1, I sketch my view of the role of idealization in securing scientific understanding and a conception of understanding that supports this role. In my view, idealizations regularly have direct epistemic value, even at the cost of accuracy, insofar as they promote understanding. I highlight accordances between this view and Elgin (2017) and de Regt (2017). In Section 2, I discuss what I take to be a point of difference, or at least difference in emphasis, between my view and both of those authors' views. Put abstractly, I want to emphasize the significance of a number of kinds of variability to this kind of position, including variability introduced by the many different aims of scientific projects. Finally, in Section 3, I address what this view suggests about the relationship scientific understanding bears to truth and knowledge. In my view, scientific understanding is regularly furthered by some sacrifice of truth, but this non-factive view of understanding is nonetheless deeply related to the pursuit of scientific knowledge. It's just that most scientific knowledge is not strictly speaking of the phenomena under investigation, but of causal patterns they embody. This discussion sets up a contrast between my view and Khalifa (2017).

2 Idealization and Understanding

Idealizations are assumptions made without regard for whether they are true and often with full knowledge they are false. Familiar examples include the assumption that a population is infinite in size, the assumption that no other objects exert force on a two-body system, and the assumption that some gas is ideal, that is, composed of non-interacting point particles. I want to start this discussion by offering a general diagnosis for the importance of idealization in science. I don't think these ideas are particularly controversial, but they will point the way toward what I see as the distinctive contribution of idealization to securing understanding. This discussion relates closely to chapters 2 and 4 of (Potochnik, 2017).

Our world is complex. In particular, it's causally complex: causal processes leading to any given event can be traced indefinitely far back in time, and at any given point in time, several causes influence any given event. (Here I include the causes often considered to be background conditions.) Moreover, it's not uncommon for causes to interact, that is, to influence each other's action, such as in feedback loops. And yet, faced with the need to grapple with this complex world, scientists face cognitive, and other, limitations. These limitations make it difficult to secure causal knowledge, to make accurate predictions, and to pursue science's other aims in this causally complex world of ours. Or, perhaps better, this point can be phrased positively: simple patterns are cognitive valuable. Simple patterns support human influence on and understanding of our world. There is thus a basic mismatch between the cognitive value of simple patterns and the world's complexity. (See especially Wimsatt, 2007.) So, in the face of this mismatch, we often resort to lying a little bit: we artificially simplify the parts of accounts that we aren't interested in to improve our access in a variety of ways to the parts we are interested in. This is one service that idealizations provide.

If it's right that simple patterns are valuable for understanding and control, and idealizations aid us in our pursuit of simple patterns, then idealizations have value. In particular, because understanding is an epistemic achievement, idealizations thereby have epistemic value. It's worth clarifying the exact claim I intend to make here. Let's distinguish idealizations from idealized representations. By idealized representations, I mean representations, such as scientific models, that incorporate idealizations. By idealizations, recall, I mean false assumptions (or, more carefully, assumptions made without regard for whether they are true). The assumption that some gas is composed of non-interacting point particles is an idealization; the ideal gas law this assumption facilitates is an idealized representation. Many philosophers have argued that idealized representations have epistemic value. Indeed, this seems to be unquestionably true. What I take to be more contested is the idea that idealizations themselves—false assumptions—have direct epistemic value. I say 'direct' because I do not take idealizations to be a first step toward scientific understanding, to be improved upon later, but full participants in the epistemic success of achieving scientific understanding.

So, in my view, idealizations can be used to facilitate representation of simple patterns to generate scientific understanding. In particular, idealizations aid in the representation of causal patterns. Causal patterns are, I think, a common focus in science. Causal patterns are *patterns* insofar as they are regularities that are limited in scope and that may permit exceptions. The ideal gas law characterizes the approximate behavior of most gases, though its predicted relationships break down at low temperatures and at high pressures. It also ignores molecular size and intermolecular forces. Recall the idealization of an ideal gas composed of non-interacting point particles; this idealization achieves that neglect. Accordingly, even within its scope of application, the ideal gas law has exceptions. Causal patterns are *causal* at least in James Woodward's (2003) manipulability sense: to represent a causal pattern is to show how changes to a system would, over some range of circumstances, precipitate changes in other feature(s) of the system. The ideal gas law shows, for example, how temperature increasing in a sealed container of gas with a fixed volume increases the pressure. Mastery of causal patterns is exactly the kind of thing that beings who prize simplicity need in order to operate in and grapple with a causally complex world like ours.

No phenomenon is determined by just one causal pattern. Rather, given causal complexity, causal patterns abound. Any phenomenon embodies quite many causal patterns; any of these patterns may be foregrounded in our investigation of the phenomenon, depending on our interests. Here's a very simple example: any phenomenon that embodies the pattern characterized by the ideal gas law also embodies a version of the van der Waals equation. But, that same phenomenon also embodies a host of other patterns. An illustration will take a bit of additional imaginative work. Suppose the temperature increase in a sealed container of fixed volume was in fact a can of aerosol hairspray left in a car on a hot day. This phenomenon embodies the pattern described in the ideal gas law. It also embodies the pattern of the greenhouse effect: the short wavelengths of visible light can enter through the glass of the closed windows, but the longer wavelengths of infrared light radiated by the objects in the car that absorbed the light cannot exit through the glass as easily. These patterns relate to different aspects of the phenomenon, and which is of interest depends on which aspects we are focused on. As these simple examples show, different patterns embodied by some phenomenon may be closely related to one another or wholly unrelated (or anywhere in between).

Thus, idealizations can aid in the representation of causal patterns embodied by phenomena, even as they introduce falsehoods of phenomena. This is the key to idealizations' contribution to understanding. In my view, scientific understanding of some phenomenon requires (a) grasping a causal pattern (b) that is embodied in the phenomenon and (c) focal to the cognizer(s). I'll say something about each part of this requirement. First, the emphasis on grasping I draw from Stephen Grimm, Michael Strevens, and others who have analyzed scientific understanding (e.g. Grimm, 2012; Strevens, 2013). Grasping a causal pattern in particular requires appreciating the nature of some causal dependence along with the scope of that dependence, that is, the range of conditions under which the dependence obtains. Causal patterns as central to understanding of phenomena fits with dominant themes in the literature on scientific explanation, as the explanatory value of causal information and of laws or patterns have each been emphasized (see Strevens, 2004).

Now, consider (b) and (c) in my above articulation of scientific understanding. Understanding has a dual nature: it is both an epistemic achievement and a cognitive state. Understanding thus must achieve both the proper relationship to the world, to the object of understanding, and the proper relationship to the cognitive agent, to the subject of understanding. (b) and (c) address these respectively. For grasping a causal pattern to yield understanding, the causal pattern must actually be embodied in the phenomenon to be understood—the phenomenon must fit the pattern. And, that pattern must be focal—it must relate in the right way to the subject's cognitive needs.

If grasping a causal pattern can constitute understanding of a phenomenon, then in the right circumstances, idealizations—false posits—can directly contribute to scientific understanding and, thus, are of direct epistemic value. These ideas about idealizations and understanding bear much in common with both de Regt's and Elgin's accounts of understanding. For de Regt (2017), understanding is explanation based on intelligible theory; it must be both empirically adequate and internally consistent. For Elgin (2017), an understanding of a domain is a reflective endorsement of a network of interconnected commitments in reflective equilibrium. Both also emphasize idealization and general patterns, if not with the same term. The next section will explore an aspect of my view that I believe distinguishes it from these authors'.

3 Variability and Many Aims

As mentioned in the previous section, Elgin also emphasizes the epistemic contribution of idealizations, or a broader class of what she calls "felicitous falsehoods" (2004; 2017). As she notes, if idealizations can directly contribute to scientific understanding, we need a more relaxed requirement than truth for when a posit can be conducive to understanding. She treats truth as a threshold requirement rather than as an absolute: any divergence from the truth must be negligible, that is, safely neglected. And, she points out, whether a posit is epistemically acceptable depends on its role in an argument, explanation, or theory (or, one might say generically, in a representation). The ideal gas law assumes, contrary to fact, that a gas is composed of non-interacting point particles: this is true enough. The ideal gas law represents a range of systems accurately with the aid of this falsehood. But, if a theory of gases posited as a central tenet that gases are composed of non-interacting point particles, this would not be true enough. It diverges from the truth problematically, given that it is intended as an accurate claim.

I think Elgin's insights here are important. Truth might function as a threshold requirement, and where the threshold is for being true enough to contribute to understanding reasonably depends on the role a posit plays, whether it is asserted, conjectured, assumed contrary to fact, or etc. This is a necessary adjustment if idealizations are understanding-conducive. But, in my view, this standard for epistemic acceptability is not sufficiently variable to accommodate all the ways that idealizations facilitate understanding. The epistemic acceptability of posits varies not just based on the role the posit plays in a representation but also based on the specific epistemic purpose of that representation.

In the previous section, I suggested that grasping a causal pattern generates understanding of a phenomenon (embodying that pattern). Recall that phenomena embody many patterns, and which can generate understanding depends on the focus of those seeking explanation. So, that focus—the specific aim of understanding—also shapes when a posit is epistemically acceptable. This gives rise to the following criterion for epistemic acceptability:

> A posit is **epistemically acceptable** when its divergence from truth is insignificant, taking into account: (a) the posit's role in the representation, and (b) the epistemic purpose to which that representation is put.

Consideration (a) is Elgin's insight, while (b) reflects an additional source of variability due to the specific aim of understanding. On this modified definition of epistemic acceptability, different specific aims of understanding can motivate different kinds of idealizations, even when the causal facts are all the same.

Here are two quick examples for how this criterion for epistemic acceptability works and why the variability introduced by (b) matters. Bokulich (2009) discusses the closed orbit account of electron absorption under magnetic force. This account employs the idealization that electrons follow fixed orbits. This posit is not epistemically acceptable as a claim of classical mechanics, taken as a theory. Indeed, scientists have refuted classical mechanics as a theory of subatomic particles, including the claim that electrons follow orbits. Nonetheless, this posit is epistemically acceptable as an idealization, an assumption contrary to fact, in closed orbit theory. One difference is the posit is not a claim but an idealization. A second difference is that the goal is to understand a specific, strange phenomenon of electron behavior, and this idealization is apt given that aim. If the aim were instead to explain electron movement within atoms, this idealization would be unsuitable. These two differences amount to the epistemic acceptability of the posit of fixed electron orbits varying based on both considerations (a) and (b) above.

The idealization of infinite population size is common in population biology. Unlike electron orbits, this was initially introduced as an idealization; no biologist has ever claimed that some populations actually are infinite in size. So any use of the infinite population size posit satisfies condition (a) in the proposed criterion for epistemic acceptability; acknowledged falsehoods can deviate dramatically from the truth without epistemic harm. But, only some uses of this idealization are actually epistemically valuable. The assumption of an infinite population is useful because it enables biologists to ignore the role of genetic drift in bringing about some evolutionary outcome. Accordingly, this posit is epistemically acceptable as an idealization in an account of natural selection's role in bringing about some trait, since the aim is then to set aside any influence of drift. But it is not epistemically acceptable, even as an idealization, if the aim is instead to represent the role of drift. The posit of an infinite population directly interferes with achieving this aim. If the aim is instead to represent all the important evolutionary influences on some trait, then the epistemic acceptability of assuming an infinite population depends on the causal facts. This idealization is epistemically acceptable if drift is not an important evolutionary influence on the trait but not if drift *is* an important influence. All of this comfortably fits within condition (b) of the criterion for epistemic acceptability I have suggested.

I thus believe that the threshold for what is true enough, for epistemic acceptability, needs to be variable in more ways than Elgin's (2017) account supports. I've suggested that there is variability in the epistemic acceptability of posits depending on the posit's role in the representation and the representation's specific aim. An acceptable idealization (or other posit) for one specific aim may be unacceptable for a different aim. Elgin (2004) says something like this as well:

> There is no saying whether a given contention is true enough independently of answering, or presupposing an answer to the question 'True enough for what?' So purposes contribute constraints as well. Whether a given sentence is true enough depends on what ends its acceptance is supposed to serve.

But in her 2017 book, Elgin instead cashes out epistemic acceptability in terms of a reflective equilibrium of commitments: "whether a representation. . . is acceptable turns on whether it is an element of an account in reflective equilibrium" (89). I don't think this approach to epistemic acceptability succeeds in making sense of the epistemic value of idealizations that Elgin and I both emphasize. Idealizations employed by different representations are regularly inconsistent with one another, in turn giving rise to idealized representations that are, at least on the face of things, also inconsistent with one another. Recognition of the variability introduced by different specific aims—including different varieties of understanding pursued—is in my view a reason to jettison any expectation of coherence among our scientific representations. Our representations, and the idealizations that facilitate them, vary according to their specific epistemic aims. See (Potochnik, 2019) for more on this concern about Elgin's account.

This variability due to different aims is linked to other forms of variability as well. First, in contrast to Strevens (2008) and Elgin (2017), I think even idealizations of causal difference-makers can be epistemically acceptable. Because of our causally complex world, many difference-makers for a phenomenon are incidental to a focal causal pattern. Idealization provides a way to set those causes aside, saying just enough to set the stage for the focal causal pattern to be grasped. Recall my example of the infinite population size idealization just above. When the aim is to understand the role of natural selection in bringing about a trait, it is usually advantageous to assume that the population is infinite, which enables drift to be neglected and facilitates common mathematical models of natural selection. This is so even when drift has influenced the actual outcome, so long as drift hasn't swamped natural selection's causal role. The same is true, I think, for using the ideal gas law to explain a gas's pressure when that pressure deviates some, but not too much, from what the ideal gas law predicts due to intermolecular forces.

Second, because grasping different causal patterns embodied in the same phenomenon can constitute understanding, in my view, there are different explanations, different varieties of understanding, even for a given phenomenon. Indeed, though discussion of this idea is beyond the scope of the present paper, I believe this is so even for a given explanandum (Potochnik, 2016). Different specific aims of understanding—put more simply, different research questions—lead to a focus on different causal patterns and thus occasion different treatments of the same phenomena. A kind of methodological pluralism follows from this view. I don't think science generates a unified understanding or explanatory store but rather different, cross-cutting varieties of understanding, even of a given phenomenon. This contrasts with Elgin's emphasis on a community's endorsement of a network of interconnected commitments in reflective equilibrium. In my view, science does not generate coherent accounts but piecemeal, interest-guided glimpses at some of the causal patterns embodied in the complex phenomena surrounding us.

4 Truth and Knowledge

The view of scientific understanding I have outlined, where different varieties of understanding are occasioned by different specific aims and make liberal use of idealizations, may seem problematically distant from truth and knowledge. Yet aren't truth and knowledge supposed to be at the heart of epistemic achievement and, if you think science is in the business of epistemic achievement, at the heart of science? In this last section, I will discuss how the aim of understanding relates to the aims of truth and knowledge on this account. In the process, I will position this view in relation to Khalifa (2017), who accords understanding a tighter connection to knowledge.

I want to begin by acknowledging that, even on my view, the unvarnished truth sometimes best provides understanding. For example, to understand why some snakes have vestigial leg bones, it can be enough for someone to point out the true fact that snakes are descended from lizards. (In my view, this is a sufficient explanation for someone who has the proper background information, including regarding evolutionary theory, but lacks this phylogenetic information, presuming this is the kind of pattern they meant to inquire about.) But in many other instances, the path to scientific understanding is paved with falsehoods, that is, with idealizations. This is so whenever representing an illuminating causal pattern is benefited by setting aside complicating details, details that may be causally relevant in their own right, but that are incidental to the pattern focal to immediate research. It is in this way that idealizations, falsehoods, can directly facilitate understanding.

I thus suggest that understanding, not truth, is science's ultimate epistemic aim. When I say that understanding is science's *ultimate* epistemic aim, what I mean is that when the pursuit of truth and of understanding qua cognitive achievement part ways, the aim of understanding trumps mere truth. This is so even when we restrict our attention in the customary ways to relevant truths. Truth is in some cases the best way to achieve understanding, and certain kinds of truth may facilitate the achievement of other scientific aims, such as prediction. But science regularly achieves epistemic success not in spite of but in part because of its deviations from the truth. This is the way that science navigates the tension I identified at the outset of this paper between the cognitive value of simplicity and the world's complexity. Simple patterns are enlightening. If, because of our world's complexity, we need to get certain things wrong in order to grasp a simple pattern, so much the worse for accuracy.

This is ultimately why idealizations play a positive epistemic role. On the view outlined in this paper, an account that is less accurate of a phenomenon (i.e. more idealized) can generate better understanding of that phenomenon when it depicts the causal pattern focal to those who seek understanding. It seems to follow that understanding is non-factive, as Elgin (2004, 2017) claims. Explanations do not need to be strictly true of the phenomena they explain. Someone can understand that A because B even if B is not entirely true, and even if B is less accurate than some other accounts of A. (Though, recall, there is still an epistemic requirement explanations must satisfy: they must depict a causal pattern embodied by the phenomenon.) In contrast, knowledge is factive. To pick up our simple example from above, one can understand that the pressure in a balloon doubled because the volume was halved and $PV = nRT$, but one cannot know that $PV = nRT$ in this system (as it's not strictly true). One can, however, know something closely related, namely, that the pattern described by the ideal gas law is embodied by this system.

Khalifa (2017) may consider this view of understanding to be quasi-factive rather than non-factive. This hinges on whether he requires non-factivists to reject (a) the approximate truth of posits used to explain (i.e. the explanans) or (b) the approximate truth of the explanatory relation (Khalifa's *q explains why p*). Some passages are more naturally interpreted in the former way (see p.156) and others in the latter way (see p.157). Notably, for Khalifa, (b) does not require (a), but simply the truth of the explanandum p and that the explanans q include some difference-makers for p (p.157). So, my insistence that some posits used to explain are not true may not interfere with Khalifa's quasi-factivism.

I am indifferent to where the requirements we adopt for factivity come down on this matter. What I do think is important is to recognize how idealizations—falsehoods—facilitate scientific understanding. I believe there is clear evidence that idealizations are not merely tolerated in the vehicles of our scientific understanding but play central, positive roles. If so, scientific understanding is at least sometimes achieved not in spite of but (in part) *because of* a sacrifice of truth or accuracy. I don't think Khalifa's quasi-factive approach to understanding permits this, let alone emphasizing it. For Khalifa, explanations more closely approximate the truth the more difference-makers they cite. This makes it possible to order varieties of understanding relative to an ideal, and it entails that more veridical representation of any difference-makers always improves understanding. This precludes the possibility that a more idealized explanation is epistemically better than a more veridical explanation (when the veridicality regards difference-makers).

Yet, in my view, idealization can play an epistemically crucial role. Idealized representation enables epistemic agents to grasp a causal pattern that a more accurate representation of the phenomenon would obliterate. The ideal gas law, in part due to its idealizations, applies across a range of systems that have different parameters in the van der Waals equation of state. In doing so, it displays a pattern embodied across that broad range of phenomena. When grasping this pattern meets an audience's epistemic needs, given their interests and background knowledge, this generates understanding—and the van der Waals equation does not, or not so well. The ideal gas law also may be a step on the way to the development of the more accurate van der Waals equation, but that is incidental to much of its epistemic value, which consists in its value to understanding. The ideal gas law is a standalone epistemic achievement, which recognition of other causal patterns won't replace and doesn't directly supplement.

So where does this leave scientific knowledge? Knowledge is factive; that is, something cannot be known without being true. This is in contrast to what I have suggested for understanding, which I have argued can tolerate and even benefit from some departures from the truth. It seems to follow that science may generate understanding without giving rise to knowledge. But no. Rather, my account can quite naturally accommodate the widely held view that science's epistemic success consists in knowledge. I have said that scientific understanding of a phenomenon consists in grasping the causal pattern embodied in that phenomenon that is focal to those seeking understanding. This is what justifies idealized explanations sacrificing truth of phenomena. But these principled inaccuracies of phenomena enable accurate representations of something else: *causal patterns*. Science does generate knowledge, but it is knowledge of causal patterns.

Thus, in my view, the object of our scientific knowledge is not technically the phenomena scientists investigate, but the causal patterns those phenomena embody. Science generates understanding of phenomena, and it does so via knowledge of the causal patterns they embody. Knowledge and understanding go hand in hand, but there is a gap between their objects. The phenomena investigated in science spur our drive for understanding, but because of the cognitive value of simple patterns, scientists regularly choose to sacrifice some accuracy of those phenomena to the end of grasping the patterns the phenomena embody.

The difference between truth of phenomena and of causal patterns is, I think, an important one. I have suggested that patterns bear a many-one relationship to phenomena. Or, the relationship is actually many-many, since different patterns embodied by any given phenomenon group that phenomenon with different ranges of other phenomena. This results in the variability I emphasized in the previous section. Different highly specific aims of understanding lead to focus on different patterns, which in turn motivates different idealizations, amounting to different sacrifices to accuracy of phenomena. All of this variability, the rich ways in which sacrificing literal truth of phenomena improves our understanding, and does so in part due to our specific interests, gets lost if we focus simply on the limited respects in which we can still call our idealized representations true.

At least some of science's successes are epistemic successes and, in particular, generate knowledge. But this knowledge is not strictly speaking of the phenomena under investigation, but of some of the causal patterns they embody. Which causal patterns we have scientific knowledge of depends on the specific interests of the practitioners, audiences, and sponsors of science. And securing this knowledge involves idealizing features of phenomena incidental to those specific interests—not as a first step to a later epistemic achievement, but as one important aspect of a full epistemic achievement in its own right.

Acknowledgments

I owe thanks to my co-symposiasts Catherine Elgin, Kareem Khalifa, and Henk de Regt for a lively and genial exchange. Kareem Khalifa also provided helpful comments on a draft of this paper.

References

Bokulich, A. (2009). Explanatory fictions. In M. Suárez (Ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization*, Routledge Studies in the Philosophy of Science, pp. 91–109. New York: Routledge.

de Regt, H. W. (2017). *Understanding Scientific Understanding.* Oxford University Press.

Elgin, C. Z. (2004). True enough. *Philosophical Issues 14*(1), 113–131.

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

Grimm, S. R. (2012). The value of understanding. *Philosophy Compass 7*, 103–117.

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

Potochnik, A. (2016). Scientific explanation: Putting communication first. *Philosophy of Science 83*, 721–732.

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

Potochnik, A. (2019). Review of catherine elgin's *True Enough*. *Philosophical Review 128*, 363–366.

Strevens, M. (2004). The causal and unification accounts of explanation unified—causally. *Nous 38*, 154–179.

Strevens, M. (2008). Comments on Woodward, *Making Things Happen*. *Philosophy and Phenomenological Research 77*(1), 171–192.

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

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge: Harvard University Press.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.