Synthese (2017) 194:1273–1293
DOI 10.1007/s11229-015-0992-x

ORIGINAL PAPER

MUDdy understanding

Daniel A. Wilkenfeld[1]

Received: 7 July 2015 / Accepted: 9 December 2015 / Published online: 17 December 2015

Abstract This paper focuses on two questions: (1) Is understanding intimately bound up with accurately representing the world? (2) Is understanding intimately bound up with downstream abilities? We will argue that the answer to both these questions is "yes", and for the same reason-both accuracy and ability are important elements of orthogonal evaluative criteria along which understanding can be assessed. More precisely, we will argue that representational-accuracy (of which we assume truth is one kind) and intelligibility (which we will define so as to entail abilities) are good-making features of a state of understanding. Interestingly, both evaluative claims have been defended by philosophers in the literature on understanding as the criterion of evaluation. We argue that proponents of both approaches have important insights and that, drawing on both their own observations and a few novel arguments, we can construct a more complete picture of understanding evaluation. We thus posit the theory of there being Multiple Understanding Dimensions. The main thing to note about our dualism regarding the evaluative criteria of understanding is that it accounts for the intuitions about cases underlying both previously held positions.

Keywords Understanding · Explanation · Philosophy of science · Factivity · Epistemology

1 Introduction

This paper focuses on two questions:

(1) Is understanding intimately bound up with accurately representing the world?

✉ Daniel A. Wilkenfeld
daniel.wilkenfeld@berkeley.edu

[1] University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA

(2) Is understanding intimately bound up with downstream abilities?[1]

We will argue that the answer to both these questions is "yes", and for the same reason—both accuracy and ability are important elements of orthogonal evaluative criteria along which understanding can be assessed. More precisely, we will argue that representational-accuracy (of which we assume truth is one kind) and intelligibility (which we will define so as to entail abilities) are good-making features of a state of understanding. Interestingly, both evaluative claims have been defended by philosophers in the literature on understanding as *the* criterion of evaluation. We argue that proponents of both approaches have important insights and that, drawing on both their own observations and a few novel arguments, we can construct a more complete picture of understanding evaluation. We thus posit the theory of there being Multiple Understanding Dimensions (MUD).

Let us start with the link between understanding and accuracy. The traditional view, formulated by Hempel (1965), is that you can only understand why something is the case when all the premises in the explanatory argument are true.[2] Philosophers have since jettisoned much of Hempel's apparatus—such as the requirement that the explanans contain at least one law (e.g., Friedman 1974), the requirement that they form an argument at all (e.g., Salmon 1984), and even the requirement that understanding depends on possessing an explanation (e.g., Lipton 2009; Wilkenfeld 2014)—but most have nevertheless generally maintained that understanding is ultimately grounded in believing true things. We call any view that assesses understanding in terms of its representational accuracy without any reference to subsequent abilities an "ability-free view."

More recently, however, some philosophers have begun to think that the true marker of understanding is not what it says about the world, but rather what you can do with the state of understanding itself. Most notably, de Regt (2009b) argues that one understands a phenomenon when and only when one has an intelligible theory that explains that phenomenon, where theory intelligibility is defined primarily in terms of what scientists can do with the theory. Chang (2012) and Le Bihan (forthcoming) also have views that downplay the importance of representational accuracy. We will argue that even Wilkenfeld (2013)—who characterizes his view in explicitly representational terms—fails to appreciate the significance of representational accuracy. We will call any view in this family an "accuracy-free view."

The structure of the paper is as follows. In Sect. 2 we articulate our central thesis a bit more precisely, specifying what we mean by "representational accuracy" and "intelligibility", and how a tradeoff between them might be effected. In Sect. 3 we address the accuracy-free views. We begin with a critical discussion of de Regt (forthcoming), arguing that his view commits him to some intuitively implausible claims about who understands what better than whom. In Sect. 4 we then turn our attention to worries about ability-free views. We take as our modern exemplar of this family the "Simple View" of Strevens (2013), and will rely heavily on a published reply to that piece by Wilkenfeld and Hellmann (2014). We will also

[1] By "downstream" abilities we mean that which one can do with the understanding, contrasted with that by which one comes to possess it.

[2] Hempel defends this most explicitly in (1965, pp. 248–249).

show that our concerns apply to work by Kelp (forthcoming). In Sect. 5 we discuss accounts (i.e. Wilkenfeld 2013; Khalifa in press) that are more amenable to MUD, and show why incorporating this approach would mark an improvement on those accounts. In Sect. 6 we look at the positive case for MUD, arguing that it gives the best account of cases philosophers have put forward. Finally, in Sect. 7, we draw out an interesting connection between MUD, explanation, and inference to the best explanation.

2 MUD

We begin this section by providing a sense of each of the two major metrics of understanding evaluation. We will then employ these metrics to characterize our positive proposal a bit more precisely. Finally we turn to how this positive proposal can be implemented in practice.

The first proposed metric of understanding evaluation is representational accuracy. We do not have an account of what it means for a representation to be accurate but presumably the general idea is that the actual state of affairs of the world is in some important sense similar to the state of the world as depicted in the representation. Importantly, if we assume a correspondence theory of truth, any true propositions will be representational accurate. However, approximately true propositions[3] and non-propositional representations (e.g., maps) can be more or less accurate as well. Thus, the requirement that understanding be representationally accurate is broader than the requirement that a state of understanding be composed of true beliefs. We take this broader notion because we mean to show that even it cannot account for all the dimensions along which we evaluate understanding, without being supplemented by an account of ability.

The second proposed metric is intelligibility. In its original formulation (de Regt and Dieks 2005; de Regt 2009b), intelligibility was strictly a matter of being able to draw out qualitative consequences of a theory without performing precise calculations. However, as with truth, we take a very broad notion of intelligibility, in order to gain maximum traction from the claim that even this broader notion does not suffice for a theory of understanding evaluation. We will count as intelligible any understanding such that the possessor of it can as a result do certain things or make certain (good) inferences.[4] Importantly, the relevant abilities include *downstream* effects of the understanding, not just the capabilities that went into understanding acquisition.

We are now in a position to state our central thesis a bit more precisely:

[3] For our purposes, we do not need a theory of what it means for something to be close to being true. See Oddie (2014) for a discussion of some interesting possibilities.

[4] This terminology is sub-optimal in at least two respects. First, where de Regt defines intelligibility as a property of theories, we treat it as a property of the understanding those theories lead to; this is motivated entirely by the absence of a good word for the property of being ability-conferring-understanding. Our usage is also not to be confused with that of Waskan et al. (2014), who treat intelligibility as a property of explanations that cause understanding.

Multiple Understanding Dimensions (MUD): the quality of a state of understanding is evaluable along multiple orthogonal dimensions, including (but perhaps not limited to)[5] *both* representational accuracy *and* intelligibility.

If there is more than one dimension along which understanding is properly evaluable, it is natural to wonder how we can form an overall assessment of a state of understanding, as we might want to do when hiring someone to do a job or judging to whom we should defer (cf. Wilkenfeld, Plunkett, & Lombrozo, forthcoming). However, we see no special problem here for our theory of understanding; most things in this world are evaluable along multiple dimensions—we are all sadly familiar with trying to choose our food while comparing both flavor and nutritional value. It is an empirical (and interesting) matter to find out how scientists actually value accuracy and intelligibility in practice, but, regarding the more normative question of how it would make sense to value these dimensions, we think the right answer is already suggested by de Regt (e.g., forthcoming): look to the context. de Regt argues that context determines what sort of problems we care about when determining whether a theory is intelligible. We propose going one level higher—a sort of meta-contextualism—where the context also fixes how much we value intelligibility in the first place relative to representational accuracy.[6] In some contexts, such as building a bridge, intelligibility should be the dominant criterion when determining whether someone has understanding, though all else being equal a bridge-builder would according to MUD possess *better* understanding if she also had a more accurate picture of the basic physics. In other contexts, such as the writing of a theoretical physics textbook that purports to describe the world, representational accuracy will be more important, though again intelligibility will still be a relevant axis of evaluation.

We now turn to our defense of the claim that both representational accuracy and intelligibility are independently good-making features of a state of understanding.

3 Representational accuracy: a criterion of understanding evaluation

As representational accuracy was the more traditional metric of understanding evaluation, we begin with a defense of its importance. The structure of the argument is not to reiterate the benefits of accounts of understanding that require representational accuracy, but rather to present a new look at the costs of an account that does not require it. Specifically, we will be looking at de Regt's (forthcoming)[7] intelligibility-centric account [though, as noted in the introduction, we expect that many of the same considerations can be marshalled against Chang (2012) and Le Bihan (forthcoming)].

[5] We suspect that some other putative virtues of understanding—such as mechanism or fecundity—can be captured in terms of the two discussed here, but are certainly open to the possibility that there are others.

[6] In other words, we are positing a one-to-many function between contexts and ordered tuples of values to be placed on each dimension of understanding evaluation, of which we have identified two.

[7] It is less clear how our criticisms apply to de Regt (2009b) than to de Regt (forthcoming). In the former, he does at points seem to suggest that accuracy is at least generally a good thing; if that is right, we take the current section to be a plea to return to an earlier position. That he has jettisoned accuracy entirely is also suggested by de Regt and Gijsbers (forthcoming).

3.1 de Regt's game of truth or dare

As the case for evaluating understanding on the basis of intelligibility might be less obvious and intuitive than the case for evaluating understanding on the basis of representational accuracy, in this sub-section we briefly outline de Regt's argument that it is intelligibility, rather than accuracy, which is relevant for evaluating understanding. He bases his argument primarily on an examination of real scientific cases—both modern and historical—where he argues that there is understanding generated by a theory that is not even approximately true.

The first example of understanding based on falsehoods comes from the realm of economic modeling. Citing Hindriks (2013), de Regt (forthcoming) points out that most assumptions behind economic models are not only false, but *very* false and *obviously* false. To take the most familiar example, much of classical economics is founded on the assumption that human beings are rational utility maximizers, whereas we know well from psychology and behavioral economics (as well as common-sense) that this assumption is rarely if ever satisfied. Nevertheless, it would be extremely revisionary to claim that classical economics provides no understanding of, for example, the functioning of markets. (Hindriks goes even farther, suggesting that it is precisely such models' deviations from reality that accounts for their conduciveness to understanding.) de Regt offers as another example ecological models, which are similar to economic models in being constructed on knowingly implausible assumptions.

de Regt offers two other classes of example. First, he considers the use of model organisms in biology, such as the use of fruit flies to model evolution. Obviously fruit flies themselves cannot be either accurate or inaccurate, but de Regt nevertheless cites their use as further evidence that understanding does not have to be representationally accurate. The reason is that the very factors that make fruit flies convenient to study also make them poor exemplars of the species they are supposed to model. Thus, conclusions made from examination of fruit flies, for example that genes behave thus-and-so, are knowing oversimplifications of the real mechanisms in the world.

The final class of examples de Regt considers are those instances in the history of science where people seemed to have real understanding, even on the basis of theories that later proved false. The example he gives is of Newton's understanding of the tides. The acknowledgement of Newton's understanding as unimpeachable however also commits de Regt to admitting that the phlogiston-theorists might have had understanding of, for example, combustion (forthcoming, p. 12). On his view, this admission is softened by two things. First, de Regt points out all the myriad of things the phlogiston-theorists did right, and all the predictions they could make that would prove empirically adequate. Second, he argues that while there is no *principled* reason to deny the phlogiston-theorists understanding, we can nevertheless deny them understanding on the practical grounds that there are no contexts in which their putative understanding would be more useful than a modern chemist's (or Lavoisier's intermediate theory) (ibid.).

The general contention is consistent: understanding is a matter of being able to do the right things, and theories that enable you to do more are better, regardless of the accuracy of the claims that make up that theory. de Regt (forthcoming, p. 6)

concludes that understanding is a version of "dare" rather than "truth"—its role is to be "productive", not "representational".[8]

Notice that this conclusion is fairly radical. Even philosophers who have agreed that "fictional" elements such as idealizations pose a prima facie problem for the accuracy requirement have generally tried to salvage some version of it. For example, Mizrahi (2012) argues that understanding can be advanced by theories with falsities, but tries to cordon off the damage by saying that better understanding is achieved when we're aware of the theory's boundary conditions and limit our application of it to a domain where it is basically accurate. Elgin (2007), among others, has challenged even this conventional wisdom, arguing that understanding is not a matter of being in possession of the truth. However, even she ultimately holds understanding accountable to the truth: "Pretty plainly, understanding somehow answers to the facts. The question is how it does so" (Elgin 2007, p. 33). de Regt severs this link, at least in principle.

3.2 Enlarging the bullet

In this sub-section, we aim to show that the concession that de Regt makes in regard to the phlogiston-theorist is more costly than he at first suggests. We argue this first by an extension to the real example of Ptolemaic astronomers, and then to hypothetical case of surprisingly accurate practitioners of hepatoscapy (the divination of the future on the basis of examination of the livers of dead animals).

The point about Ptolemaic astronomers is that, when Copernicus first proposed his heliocentric model, its predictions were actually less accurate than the Ptolemaic alternative. Given technological limitations, there was not much a 16th century astronomer could *do* with a theory of what revolved around what, so manipulation and control (the other two elements of intelligibility according to de Regt forthcoming, p. 7) were not really at issue. Thus, the greater predictive powers of the Ptolemaic system seems to imply that it was more intelligible. The problem for de Regt is that this commits him to saying not only that the Ptolemaic astronomers understood why the planets exhibit retrograde motion, but that they actually understand that motion *better* than the Copernicans who correctly put the sun at the center of the solar system. That seems far worse than crediting phlogiston-theorists with some limited understanding. Moreover, it is less obvious how de Regt could make an analogous move to what he says about phlogiston wherein he argues that there is no context in which the Ptolemaic understanding was more useful than the Copernican—right at first, it would have been more useful in any context that put a premium on predictive accuracy.

It is open to de Regt to object that the Copernican system was more constitutive of understanding because it was simpler, and therefore the theory itself was easier to manipulate. However, ease of manipulation theory is something that generally only accrues to a theory over time as people become familiar with it.[9] Astronomers

[8] The terms are actually quoted by de Regt from Knuuttila (2009), but we take them to capture his ultimate position.

[9] The development of Copernican astronomy and its relation to simplicity is a complex issue. Kuhn (1957, p. 169) famously argues that the system's apparent simplicity is an illusion (see also Palter 1970). However,

well-versed in Ptolemy could manipulate that theory. More importantly, the relevant manipulations would be those that made true empirical predictions, and, as already noted, it was easier to coax correct predictions from a Ptolemaic system. The ability to get incorrect predictions from a theory—however effortless it may be—cannot possibly be what makes for high-quality understanding; if it were, a theory that allowed you to conclude anything from anything would be maximally understanding-conferring.

No doubt, intelligibility and representational accuracy quite often are indelibly linked; as a result, in order to construct an example where the two completely come apart it will be necessary to imagine a highly improbable world. One could rightfully object that our intuitions are not sufficiently well-honed to be reliable in such cases, or even that our concepts themselves are indeterminate when applied to cases so beyond their conditions of construction (see, for example, Wilson 2006). If one prefers to keep one's philosophical theorizing to the real world, our critique of taking intelligibility as the sole criterion of understanding evaluation must rest with Ptolemaic astronomy. However, for those willing to measure our concepts and evaluative criteria against more extreme thought experiments, we argue that the problem for de Regt can be made even more poignant. The point is not that the following example shows how intelligibility and ideal understanding can *actually* come apart, but rather that de Regt's view fails to capture the *principled* link between understanding and representational accuracy.

We think the result that Ptolemaic astronomers had more understanding than Copernican astronomers of *why* the planets exhibit retrograde motion is bad enough, but the problem is somewhat worse. At least in the case of Ptolemaic astronomy (and phlogiston-theory), there is some non-accidental link between the posits of the theory and the accuracy of the predictions. However, this is a contingent feature, and need not always be the case. There is nothing in principle impossible about a *wholly* misguided theory, quite by chance, having remarkable predictive power.

There is a possible world where all the predictions in the newspaper hepatoscopy section are both specific and accurate. We can further suppose that the predictions are not the result of mere arbitrariness, but that there is a detailed "theory" which connects the state of a being's liver to likely future events. This theory has central laws, for example that the gods always make animals' livers bigger before a tragedy and smaller before a wonderful event. These laws can serve as both the basis of models of future events and more formal theories from which other predictions can be derived syntactically. We can imagine that there is a full institutional apparatus governing the norms of hepatoscopy, with professional conventions and exchanges of ideas between hepatoscopists.[10] In short, there would be no reason in such a world to

Footnote 9 continued
even assuming it is simpler, that does not entail that it is more intelligible. Typical arguments regarding its simplicity (e.g., Wilson 2012) mainly address its metaphysical or mathematical simplicity, which is distinct from its intelligibility and ease of use, particularly by early adherents.

[10] By this we just mean those who produce or believe in hepatoscopy. Perhaps fortunately, there does not seem to be an existent word for this.

deny hepatoscopy the status of a full-fledged scientific theory.[11,12] As a result, it is also eminently possible that it could be an extremely *intelligible* theory—hepatoscopists could be better or worse at manipulating the theory in order to derive qualitative characteristics.

Suppose that, in such a world, a local hepatoscopy column regularly predicts where Venus will appear in the night sky and is, by hypothesis, always correct. However, there is absolutely no causal or explanatory connection between the structure of their theory and the accuracy of their predictions. There is some possible world where the predictions are correct entirely by chance, and that is the world we are now considering. We contend that, intuitively, the fact that the predictions are supported entirely by chance means that there is a very strong sense in which they are deficient as the basis for understanding. Perhaps there is a sense in which they provide some understanding, as they still do well along one dimension, but their complete failure along the other dimension would predict (what we would take to be the intuitively correct result) that this understanding would on our account be fairly minimal.[13] On de Regt's account, as there is no principled reason to deny Ptolemaic astronomers maximal understanding, there is no principled reason to deny amateur hepatoscopists maximal understanding either. Not only does this seem to us to be the wrong intuitive answer, but it also seems to us that if understanding cannot even distinguish between a theory whose claims have some link to their own truth-makers and one that does not, we would do well to replace it with an epistemic concept that can.

3.3 de Regt's reply

Thankfully, de Regt and Gijsbers (forthcoming) responded to an earlier draft of this paper.[14] Their response to our examples is to deny the validity of the hepatoscopist, and to "bite the bullet" with respect to the Ptolemaic astronomer.

With regard to the hepatoscopist, de Regt and Gijsbers argue that, even if she happens to be accurate, she will not be *useful* unless there is some causal connection between her predictions and their accuracy. If one tried to manipulate the world on the basis of the hepatoscopist's advice, one would constantly fall into error.

We believe this reply underestimates the force of the hepatoscopy objection. The central point is that *whatever* results one might demand of a theory in order for it to count as understanding-conferring, those results might be achievable by coincidence. Just as there is a world where all the hepatoscopist's predictions are correct, there is also a world where all such predictions are correct *and* useful, as well as whatever

[11] We thank an anonymous reviewer for stressing the importance of establishing this point.

[12] For a thorough discussion of the structure of scientific theories, see Winther (2015). Our goal in outlining the features of hepatoscopy theory in our imagined world was to make sure it satisfied the conditions of all of the syntactic, semantic, and pragmatic accounts of what makes something a theory.

[13] We again thank an anonymous reviewer for pointing out that it is consistent with our view that hepatoscopists in such a world contain some minimal understanding.

[14] As that work has not yet been published, much of what we say here might prove off the mark—however, we do think it can be a useful guide to at least the *kind* of maneuvers open to de Regt, and what we would say about them.

other criteria of success one wants. The general strategy is to note that *whatever* other desiderata one places on a theory will not replace accuracy, as an inaccurate theory that satisfied those desiderata by chance would not be ideally conducive to understanding.

Regarding the Ptolemaic astronomer, de Regt and Gijsbers simply bite the bullet, acknowledging that she understood the motion of the planets better than her Copernican contemporary. We do not think this is an outlandish position; however, we take it to be an undesirable result, if there is an alternative theory that can accommodate all of de Regt's evidence (see Sect. 6) while still showing that there was something better about the Copernican's understanding than the Ptolemaic's. We maintain that MUD outlines such a theory.

3.4 An objection and a clarification

One might worry that our concern about de Regt's being forced to countenance accidentally accurate theories will take revenge on MUD, particularly when one accepts the sort of meta-contextualism advocated in Sect. 2. The objection points out that if representational accuracy is measured against intelligibility based on context, there will surely be some contexts where we care only about intelligibility and not at all about representational accuracy. But in that context, in the examples cited above, the Ptolemaic astronomer and the surprisingly accurate hepatoscopist will still count as understanding.

Our response to this objection is to point out that the role of (meta-)context determines which understanding-virtues we should care about, not whether intelligibility and representational accuracy are both virtues at all. Even in a case where we only care about intelligibility, a lack of representational accuracy is still a bad-making feature of understanding.[15] An analogy is perhaps helpful here: in American football, a running back's tripping just short of the goal-line is a bad-making feature of his value as a running back even if, quite by chance, we happen to be in a context where we as his teammates or fans are indifferent to or even against his scoring.[16] What we are claiming is that context determines how much we comparatively *value* the two dimensions of understanding-evaluation, not how *valuable* they are. This admittedly leaves us without a guide for how to tradeoff intelligibility and representational accuracy when determining a state of understanding's objective value, but we do not take our (present) inability to measure a quantity as proof of its nonexistence.

In Sect. 6, we will return to the question of what MUD has to say about de Regt's examples that purport to show that understanding is intelligibility—rather than

[15] Of course, if our current theories also are representationally inaccurate, that would be a mark against our current understanding as well. We think that this is a feature of our view, rather than a bug—it would be mere temporal chauvinism to assume that our understanding is unimpeachable. If one imagines a future where we discover that our present theories were inaccurate, we anticipate that we would take this as a mark against our present understanding to precisely the degree of the inaccuracy. To soften this conclusion, notice that our current theories can still have some representational accuracy even if they are not completely true, so they might still score reasonably well along both dimensions of evaluation, even if they do not achieve maximal scores.

[16] A scenario where it is better if one teammate does not score does occur periodically, for example near the end of Super Bowl XLVI.

accuracy-dependent. We first turn to examine views (especially Strevens 2013) at the opposite end of the spectrum, which treat understanding as very much akin to propositional knowledge, followed by two views (Wilkenfeld 2013; Khalifa in press) that seem to split the difference.

4 Intelligibility: a criterion of understanding evaluation

In this section, we discuss the more traditional view that all there is to understanding is grasping a correct or true explanation, which entails that the downstream abilities a state of putative understanding enables are not relevant to its status. Our main objections to this picture are taken from de Regt (2009a) as well as Wilkenfeld and Hellmann (2014).

4.1 Strevens and the "Simple View": trouble with termites

If abilities are not the sole good-maker of understanding, one might wonder why we should bother with them at all. Perhaps a more traditional view that lauds only representational accuracy is all that should ever have been expected. Specifically, perhaps understanding just is grasping a correct (read: accurate) explanation. This claim has been made most explicitly by Strevens (2013), who dubs it the "Simple View". Similar claims have been put forward by Khalifa (2012) and Grimm (2010), but for present purposes we will focus on Strevens' formulation.

There is a bit of a caveat to calling any of these views "ability-free", in that all of them incorporate an active element in what it means to "grasp" an explanation. Strevens considers grasping a form of understanding-that, which he takes to be "the fundamental relation between mind and world" (2013, p. 511). This is not mere knowledge, but somehow having cognitive command of all the concepts in question. Likewise, Grimm (2010) requires that we be able to manipulate certain counterfactuals pertaining to the explanation. We will return to what Khalifa requires in Sect. 5.2.

However, what views in this family have in common is that while they might allow that some skill is required in *acquiring* an explanation, none can make room for the possibility that, at least in some contexts, one measure of understanding is in one's post-acquisition abilities.[17] de Regt (2009a) makes this point when he draws attention to the fact that understanding often requires being able to do a derivation or manipulate a model. The problem with the sorts of abilities allowed by Strevens is that they are fundamentally skills in adapting our mind to the world, whereas some of the sorts of examples de Regt invokes are those where understanding critically involves adapting aspects of the world itself (i.e., not just being able to think the right thing, but being able to do the right thing). The idea is that someone who is merely a passive recipient of information does not necessarily understand, and certainly that there are some contexts in which it would be right to judge that she understands less than a more active learner.

[17] At least, post-acquisition abilities cannot be a direct measure. They could correlate with abilities involved in understanding acquisition, but they will not on these views directly contribute to understanding's value. We thank Kareem Khalifa for this point.

The point is defended by Wilkenfeld and Hellmann (2014), who argue for the minimal claim that there are at least some contexts—such as actual performance in a laboratory setting—where understanding requires some abilities that go beyond any epistemic relation one can bear to propositions. They cite the example of complex behaviors that emerge from the interaction of relatively simple rules, as when fish swim in a school or termites clear a landscape of wood chips. Their point is that there are no propositions in the area knowledge of which lets one "see" how the wood chips get cleared (the example is from Roach et al. 2013, p. 42).[18] Rather, one only really grasps how the wood chips get cleared by actually seeing how the simple rules interact over extended periods—i.e. by running a simulation or *using* a theoretical model. Simply being aware of the model, while perhaps sufficient in most contexts (such as reading about it in a journal), would be a deficient sort of understanding for (at least some) people actively engaged in scientific investigation of termites. Relatedly, Wilkenfeld and Hellmann argue that the understanding that constitutes mastery of chess might require a certain degree of actual chess-playing ability.

We believe that while Wilkenfeld and Hellmann were right not to claim that abilities are actually prized in all scientific contexts, they could have gone one step farther than they did by pointing out that even in scenarios where abilities are not particularly salient, understanding is still more complete when they are present than when they are absent. Given two journal readers, it still seems intuitively plausible to say that the reader who can actually utilize a model herself has understanding that is in some sense (read: along some axis) more complete than the reader who cannot.

4.2 Another ability-free account: Kelp

Where we have been using Strevens as a foil, arguments in favor of the importance of intelligibility can also be applied to other accounts of understanding. For example, in a recent piece, Kelp (forthcoming) has argued that ideal understanding is maximal interconnected knowledge, with other cognate terms being defined derivatively. However, at least on the surface, even maximal interconnected knowledge does not add up to possessing any abilities. This issue is of course somewhat nuanced—Wilkenfeld and Hellmann (2014) include a discussion of the different ways that knowledge does entail abilities (e.g., via conceptual competence), and argue that for the most part they do not suffice to constitute the sort of abilities that at least sometimes play a role in understanding. Kelp (forthcoming, p. 8) seems to embrace the consequence that on his view understanding is divorced from ability, as he takes it as a *feature* of his view that an omniscient being that was incapable of making any inferences or manipulating its own representations would count on his view as understanding. We take this result to be something of a bug, as it seems

[18] For those who are curious, the rules (as stated in Roach et al. 2013, p. 42) are:

(1) Walk randomly until you encounter a grain.
(2) If you have a grain, drop it.
(3) If you do not have a grain, pick one up.
(4) Repeat.

to leave Kelp without the resources to explain why scientists in the sort of lab environments discussed by de Regt (2009a) as well as Wilkenfeld and Hellmann (2014) have their understanding *improved* in virtue of their being able to *do things* with the data.

As to the case of the omniscient being that does not make inferences itself, even if we were to grant to the omniscient knower possessed understanding, that does not preclude the possibility that understanding that could be acted upon would be somewhat better. This might be taken to go against the claim that the omniscient being is "omni-understanding, in the relevant sense" (Kelp forthcoming, p. 8). But, if one takes MUD seriously as even a live possibility, we can see that claiming that some understanding is "omni" is ambiguous between saying the understanding gets a maximal score along one dimension of evaluation and saying it gets a maximal score along all dimensions of evaluation. If the former, then the omni-understanding of the ability-free omniscient being is no threat to MUD. The latter reading would pose more of a problem, but, once it is disambiguated from the former, we find it significantly less intuitive. Moreover, even if one were to grant to such a being omni-understanding in every sense, it does not follow that intelligibility would play no role in the evaluation of non-optimal understanding—it is possible that the "distance from fully comprehensive and maximally well-connected knowledge of [that which is understood]"—that which Kelp (forthcoming, p. 18) takes to be constitutive of degree of understanding—can actually be minimized by taking into account criteria that play no direct role in evaluating maximal understanding. Having complete physical possession of all the actual land, money, life forms, and objects in the world could very well suffice for being maximally wealthy; however, absent such a massive hoard, more abstract criteria such as stock options might constitute an important dimension of wealth-evaluation. (As Kelp does not provide a metric of distance from ideal understanding, we take this last observation about the possibly shifting bases of evaluation from ideal to non-ideal cases to mark a friendly amendment rather than a disagreement.)

5 Almost MUDdy—Wilkenfeld and Khalifa

In this section, we discuss two theories of understanding that we think best accord with MUD, and argue how the explicit inclusion of MUD would mark an improvement for both.

5.1 MUDdy Wilkenfeld

Given what we have said, it seems that a theory of understanding evaluation should include reference to both representational accuracy and intelligibility. One promising theory that would seem to fit the bill is that presented in Wilkenfeld (2013). The theory presented in that paper is that understanding consists of representing that which is understood, but that which representations constitute understanding in any particular context is determined by their usefulness in helping with contextually relevant tasks. Here is the official statement of the view there:

> URM (Understanding as Representation Manipulability): A statement, attributed in context C, that thinker T understands object o, is true if and only if T possesses a mental representation R of o that T could (in counterfactuals salient in C) modify in small ways to produce R', where R' is a representation of o and possession of R' enables efficacious (according to standards relevant in C) inferences pertaining to, or manipulations, of o. (Wilkenfeld 2013, pp. 1003–1004, endnotes omitted)

As a first point, notice that, as stated, URM is a theory of understanding simpliciter, not understanding evaluation. While there is certainly a place for such a theory of outright understanding, we argue that an evaluative theory is in some way more useful—if we can determine that a given mental state is particularly *bad* sort of understanding, the question of whether it is any sort of understanding at all becomes somewhat less interesting. (Incidentally, we do not take ourselves to have provided a full *theory* of understanding evaluation that could actually say when understanding is bad—rather, we are merely laying out a programmatic outline of what sorts of factors such a theory would have to take into account.)

The more important issue for comparing MUD to URM is that we actually think explicit inclusion of the former would patch a hole that was found in the latter. Any view of understanding should have as a consequence that someone who just performs well by luck does not understand, and indeed that is a desideratum of URM (Wilkenfeld 2013, p. 1004). URM was thought to rule this out in two ways—first, by requiring that the success of one's practice be a result of one's representations, and that the success obtain in contextually relevant counterfactuals as well. And indeed, these conditions do work to rule out the hepatoscopy example in Sect. 3.2. The extremely lucky hepatoscopist would not be able to function effectively in almost any nearby counterfactuals, and it is not clear that even in this world she would be right on the *basis* of how she represents the world.

The Ptolemaic astronomer—or a fictionalized variant of her—however, poses a more serious challenge. The problem is that her true beliefs are based on how she represents the world, and that representation is non-accidentally related to their truth. Specifically, a geocentric system can be represented by an appropriately deformed set of equations that make it equivalent to a heliocentric theory. And we can even imagine some scenarios in which the appropriate transformed system of equations would function in relatively nearby counterfactuals, which might be all that is required in some particular context.[19] But we think the realist nevertheless has solid grounds to say that a picture wherein the world revolves around the sun is a better picture than one wherein the sun revolves around the earth.

In such a scenario the astronomer would have a useful but fundamentally inaccurate representation; whether or not this constitutes *some* understanding (as URM would predict), we maintain that (at least intuitively) an understanding that places the sun in

[19] In principle, one could perhaps also make the same move for the super lucky hepatoscopist, for whom it is the case that not only are her predictions correct, but she happens to be situated in a neighborhood of possible worlds where predictions made in nearby worlds would be correct as well. We think the present example, where the link between the representation and the correctness of its predictions is not entirely accidental, is more persuasive.

the center is (all else being equal) *better*. Again, as URM is not a theory of understanding evaluation, this is not necessarily a criticism, but does seem to indicate a place for fruitful augmentation of that theory. MUD provides just such an augmentation—what the envisioned Ptolemaic astronomer was missing was as representationally accurate a picture of how the solar system really looks.

5.2 MUDdy Khalifa

The other philosopher whose views we think are most compatible with MUD is Kareem Khalifa. In this section, we will first look at some of his earlier work and argue that, while perhaps in tension with MUD, it is vulnerable to the ability-based criticisms that we argue beset Strevens, as well as the considerations discussed in Sect. 4.2 regarding extrapolating from a theory of ideal understanding to a theory of non-ideal understanding. However, in more recent work Khalifa has shifted to a view that is more compatible with MUD but, will we argue, usefully augmented by it.

The central point Khalifa has consistently advocated about understanding X is that it involves a true belief that some Y correctly explains X. At first blush, this seems to leave very little room for abilities and intelligibility, but, upon closer inspection, one finds his view—particularly in its most recent formulation—to be more nuanced.

The most direct discussion of the importance of abilities in Khalifa's early work occurs in his (2012) "Inaugurating Understanding or Repackaging Explanation?" There, Khalifa addresses the requirement, found in de Regt (2009a), that understanding involves ability, and denies that this captures anything important. More precisely—he claims that anything it does tell us about understanding "can be captured by philosophical ideas about the epistemology of explanation without loss" (2012, p. 17). His claim is that what he calls a skill-condition[20] is either never necessary (if substantive) or trivial (if merely the requirement that one be able to follow an argument). We will tackle the former horn of this dilemma, arguing that substantive skills might well be necessary for some instances of understanding.

His argument that skills are never necessary is that the sort of model manipulation required by de Regt can for Khalifa be construed as merely providing more (model-based) explanation. We do not believe this is quite right—we would argue with Wilkenfeld and Hellmann (2014) that, in the termite case, the only explanatory information to be had are the low level rules and the high level description of the wood chips being cleared. However, the scientist who actually observes the ants (or a simulation thereof) in action gains understanding that *cannot* be reduced to any explanatory information (since she already had all the *information* before she saw the process).

The more important point, however, is that even if we were to grant the claim that there can be understanding without skills, that is wholly consistent with the claim that understanding generally be *better* along some axis of evaluation if it does involve skills. The arguments in "Inaugurating Understanding or Repackaging Explanation?"

[20] Khalifa treats de Regt's "skill condition" and "intelligibility condition" separately; however, as intelligibility is defined at least in large part by what one can do with a theory, we take the latter (at least in our expanded formulation) to be an instance of the former.

are thus actually compatible with MUD; to the extent that there is independent reason to accept MUD (see Sect. 6), we would recommend that melding the two will only result in a stronger theory.

Where we get into a bit more trouble with Khalifa's view is in his (2013) "The Role of Explanation in Understanding", wherein he defends the following claim:

> SET: For any non-explanatory way, w, to understand p, there exists a correct and reasonably good explanation, e, such that the understanding of p provided by w is a proper subset of the understanding of p provided by knowing that e. (Khalifa 2013, p. 169)

This at least appears largely in tension with MUD, as it states that one can achieve the very best understanding without any abilities, merely as a result of possessing the right explanation. This view is susceptible to the same critiques that plagued Strevens and Kelp—we would argue that there are at least some contexts where the person who can actually manipulate the termite-model understands *better* than one who cannot. However, even if we grant Khalifa's account of ideal understanding, notice that, as with Kelp, it is fully possible that while ideal understanding does not involve intelligibility, non-ideal understanding might still be made better by it (intelligibility) (see the argument in Sect. 4.2).

While we argue that Khalifa's earlier work is compatible with MUD,[21] his most recent formulation of his views (personal correspondence, in press) outright invites MUDdiness. In his most recent work, Khalifa argues that one of the main ways one can have better scientific understanding is by more closely resembling a scientist, which might well involve active exploration and experimentation. MUD suggests one way to increase resemblance—one can more closely resemble a scientist by being able to *use* one's understanding. Moreover, we think MUD adds to Khalifa's account in several ways. First, it applies not only to scientific understanding, but to all understanding (recall that we are taking a broader notion than de Regt of the abilities involved in intelligibility). Second, and relatedly, it can be applied directly to everyday understanding without assuming that such understanding is derivative of scientific understanding. Third, it specifies two crucial metrics by which one can be close to a scientist. Fourth, it potentially accounts for the persistent appeal of inference to the best explanation (see Sect. 7).

6 In defense of MUD

While we take ourselves to have provided some reason to think that MUD avoids many of the pitfalls of ability-free and accuracy-free views, we have as yet put forward little by way of a positive case for it. In this section, we show that MUD makes good sense of many examples of evaluative judgments of understanding.

[21] At least, the earlier work we have reviewed here. Space requires us to gloss over several other pieces e.g. Khalifa (2011), and the various subtleties they entail. Since our end point will be a general sympathy for his latest formulation, our difficulties with all earlier formulations are less critical to our project.

Our main defense of MUD has, in essence, already been given by the very philosophers who proposed ability- and accuracy views. All of their examples, which purport to show the advantage of one virtue of understanding or the other, can *collectively* be explained by the truth of MUD.

The main thing to note about our dualism regarding the evaluative criteria of understanding is that it accounts for the intuitions about cases underlying both de Regt and his interlocutors' positions. Returning to de Regt's examples, of course economic and ecological models are better when they are more intelligible—that just demonstrates that intelligibility is *a* criterion of evaluation. What it does not imply is the stronger claim that intelligibility is the *only* criterion of evaluation. A theory with equal intelligibility would be better if it were also more accurate. Even a wholly inaccurate theory can be along some axis understanding-conferring. And a less intelligible theory might even be more *valuable*, if it is much more representationally accurate, and more *valued* if we are in a context where such representational accuracy is particularly important. de Regt's other examples can be dealt with similarly. Understanding achieved on the basis of theories of fruit fly genetics will be more valuable than some more accurate theories because it is more intelligible, but it remains possible that it will be less valuable than other theories that are either much more accurate (having been derived from humans) or slightly more accurate and roughly equally intelligible.

Regarding the tides, we can see why Newton's understanding was valuable in virtue of being intelligible, yet one who used only Newtonian mechanics to understand the universe would still have some sort of deficiency in understanding.[22] And we can even grant that there was something to be said for the phlogiston-theorists' understanding. However, what we retain—and this we take to be the major advantage of our theory over de Regt's—is the intuition that phlogiston-theorists' understanding was in some way inferior to modern chemists' in principle, regardless of how useful it happened to be. Our main difference with de Regt is that on our view the phlogiston-theorists' failure was a failure *in principle*, given how they represented the world, and not a consequence of any downstream effects. In addition to getting the correct result that phlogiston-theorists did not possess understanding to the same degree as chemists, we contend that it is important that our view gets the right results for the right reasons.

This same versatility that MUD displays when coopting de Regt's examples can also apply to examples meant to show the importance of representational accuracy (which are, in fact the standard examples—such as a theory's explaining an organism's trait by describing its *actual* evolutionary trajectory).[23] Since representational accuracy

[22] Moreover, we would argue that Newton's explanation was actually better than phlogiston-theorists' along both axes of evaluation since, given how relatively flat nearby space is and how relatively slow nearby objects move, it in many respects still paints a reasonable representational picture of the universe (note that representational accuracy can obtain even in the absence of strict truth). Deciding this contentious point however is well beyond the scope of this paper. For one discussion of what is at issue, see Couvalis (1997, Chap. 4), or de Regt and Gijsbers (forthcoming).

[23] Note that even the constructive empiricism of Van Fraassen (1980)—which says that scientific theories aim at empirical adequacy rather than truth—still places importance on representational accuracy. While posits of theoretical unobservables might not need be accurate in the traditional sense, the goodness of claims about observable phenomena will still require representational accuracy.

is one dimension of evaluation, any example in defense of its importance is also an example that supports MUD.

What would be required to form a challenge to MUD is a clear case where a gain in intelligibility marked no *improvement* in understanding at all (or even marked a decrement of understanding) or vice-versa, and we are not aware of any example that has been claimed to have this result.

7 MUDdy inference to the best explanation

We conclude this paper with a discussion of the connection between MUD and inference to the best explanation (IBE). According to IBE, when we determine that an explanation is best, we should/do also go on to conclude that it is true. Moreover, as Lipton (2004) points out, by "best" we cannot just mean "likeliest", as that would render IBE largely vacuous (or at least redundant). Rather, in Lipton's terminology, we must evaluate an explanation for "loveliness", and then take loveliness as a guide to likeliness. The use of IBE is controversial—by scientific standards, there is an odd sort of optimism to assuming that we live in a universe where loveliness tracks truth. Van Fraassen (1980) has argued that assuming IBE is a good guide to truth is simply a mistake.

Bracketing the question of its validity for a moment, we turn to the question of how best to describe IBE. We argue that the best way to construe "loveliness" is as something akin to the capacity to produce understanding. We argue for this along two lines. First, we defend the claim that Lipton's criteria for loveliness tie it very closely to intelligibility which, as argued above, is one of the dimensions of good understanding. Second, we point to an independent account in the philosophy of explanation that suggests that explanation quality is directly tied to that explanation's capacity to produce understanding.

We begin with a look at Lipton's criteria for loveliness. Lipton is careful not to give a precise characterization, but makes many suggestive remarks throughout the book. First, he points out that scientists often take as a guide to inference "aesthetic considerations of theoretical elegance, simplicity and unification" (Lipton 2004, p. 66). Thinking of explanation loveliness in terms of understanding yields an account of what makes these considerations efficacious without appealing merely to their aesthetic qualities. Since intelligibility is one criterion of understanding evaluation, if loveliness is proportional to the capacity to produce understanding, we would expect more intelligible theories to be lovelier. Crucially, in addition to their beauty, the conditions Lipton mentions all serve to locate the phenomenon in a more intelligible theory. If a theory is simpler or more elegant, we can *generally* recognize consequences more easily (note that this need not *always* be the case in order to account for the *prima facie* appeal of the aesthetic characteristics). If it is more unified, we can generally reach more far-reaching consequences. Thus, one way to explain the appeal of the factors Lipton cites is that they lead to better understanding by way of better intelligibility. Similarly, Lipton elsewhere (2004, p. 122) mentions the virtues of "mechanism, precision, scope, simplicity, fertility or fruitfulness, and fit with background belief." We note that—with the exception of the last virtue, which we tend to think is more a

marker of likeliness than loveliness—all of these virtues are also generally conducive to intelligibility. Mechanism and simplicity help us reach conclusions more easily, scope helps us reach conclusions about more distant matters, precision helps us reach more accurate conclusions, and a more fertile theory will lead (more or less by definition) to recognition of more consequences. In short, explanation quality—in the form of loveliness—being tied to understanding-quality by way of intelligibility is a good explanation of why we find lovely the characteristics we do.

One can reach the result that explanation quality is tied to subsequent understanding quality more directly on at least one independently motivated view of explanation. On the view put forward in Wilkenfeld (2014) good explanations are *characterized* by being conducive to good understanding. For present purposes it does not matter whether the good understanding is what *makes* an explanation best, so long as it always for whatever reason tracks explanation excellence. Thus good understanding and good explanation might go hand-in-hand because they are both consequences of a good deductive nomological explanation (Hempel 1965), or whatever other theory of explanation one likes.

Combining the link between explanation and understanding with the criteria laid out in MUD, what can we say about IBE? We will argue that while MUD plus a link between explanation and understanding (from now on, for ease of exposition, we will simply assume this link) on the surface appear to support IBE, they actually end up showing why employing IBE might be a particularly tempting mistake.

On one hand, MUD appears to support IBE, and, strictly speaking, it does. If the best explanation is the one that is conducive to the best understanding, and the best understanding must—in addition to being intelligible—also be representationally accurate, then the best explanation must describe the world in a way that is basically accurate as well. Thus concluding that the statements of such an explanation that are supposed to do the explaining (the explanans) are at least approximately true is not so bold a leap.

There is, however, an obvious concern about this defense of IBE—IBE is supposed to be an inferential procedure, used by real thinkers to acquire knowledge they did not otherwise possess. But, if we are not already in a situation where we know whether that the state of affairs represented by the explanans obtain, then we cannot know whether a particular explanation is *actually* conducive to maximal understanding. If we had some sort of sense that tracked whether understanding were good, we could use that to pick out the best explanations, but, as philosophers have been at pains in recent years to point out (see in particular Trout 2002), the *sense* of understanding is notoriously unreliable.[24] So as a guide to truth, the representational accuracy of optimal understanding does not actually help us.

As already alluded to though, MUD might actually suggest a way in which IBE could be temptingly misleading. The problem comes out most clearly when we recall that all of the criteria Lipton cited as conducive to loveliness foster understanding by making a theory more intelligible. MUD posits that there are (at least) two *independent* evaluative dimensions to understanding—intelligibility and representational accuracy.

[24] Even philosophers who think that understanding is actually a crucial component of explanation (Wilkenfeld 2014; Waskan et al. 2014) grant that we cannot just *feel* when such understanding obtains.

There is a very natural and familiar human tendency, when something is evaluable along multiple dimensions, to simply assume that the value along both dimensions will correlate. People who think it is good for food to be organic also tend to think that organic food is healthier (Magnusson et al. 2003) despite scientific evidence that there is no such link (Smith-Spangler et al. 2012).[25] This can even affect our direct sensory experience—a salad arranged to artistically resemble a Kandinsky painting is considered both tastier and more valuable than a less artistic alternative (Michel et al. 2014). It is then reasonable to wonder whether explanations that fare better on the intelligibility dimension are *thereby* assumed to fare better on the accuracy dimension. But that would suggest that the apparent value of IBE is an artifact of our tendency to *conflate* different ways explanations can be valued. This does not establish that IBE is not a valid means of inference, but it does suggest a somewhat debunking explanation of its popularity.

8 Conclusion: truth and dare

In this paper, we have argued that understanding should be evaluated both in terms of its representational accuracy and in terms of its intelligibility, as those concepts are defined in the literature. We take this to be a natural synthesis of existing views (e.g., de Regt forthcoming; Strevens 2013), while also providing a healthy addendum to other views (e.g., Wilkenfeld 2013; Khalifa in press). We do not take ourselves to have provided a theory of understanding evaluation, but rather have argued for the claim that there are at least two dimensions such an eventual theory must adequately reflect.

Clearly, many questions remain. Are there other dimensions of understanding evaluation? How exactly does the context fix how the criteria are differentially weighted? We would also hope future work would empirically explore to what extent MUD governs actual scientific evaluations of understanding. We would like to further develop the connection between MUD and explanation, more rigorously utilizing MUD to see if it has any bearing on evaluating cutting-edge theories of explanation. However, we take the general result to be independently interesting.

Returning to de Regt's game of "Truth or Dare", our response, like that of many middle-school students throughout the years, is to cheat. However, while the traditional mode of cheating is to try to *avoid* both revealing the truth and engaging in some dare/performance, our cheat is rather to *embrace* both, insisting that understanding be answerable to both its accuracy and its practical role. We argue that this captures all of the intuitions behind de Regt's position and Strevens' position without even in principle allowing the possibility that phlogiston-theorists' (and hepatoscopists') understanding could be equal to its more scientific equivalent, or demanding that downstream abilities never play a proper role in understanding-evaluation.

Acknowledgements We would like to thank Henk de Regt for comments on this draft, Tania Lombrozo for her continued support, Kareem Khalifa for clarification regarding his views and comments on a draft,

[25] Actually there might be such a link, per Barański et al. (2014), but it seems safe to say that most people who assume organic food is healthier are not staying up-to-date on the cutting edge of the debate.

Don Hubin for a particularly poignant question regarding Copernican astronomers, Soazig Le Bihan for invaluable conversation, helpful referee comments, and the John Templeton Foundation *Varieties of Understanding* grant and its administrators for funding the present research.

References

Barański, M., Średnicka-Tober, D., Volakakis, N., Seal, C., Sanderson, R., Stewart, G. B., et al. (2014). Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: A systematic literature review and meta-analyses. *British Journal of Nutrition*, *112*, 794–811.

Chang, H. (2012). *Is water H2O?: Evidence, realism and pluralism*. Dordrecht: Springer.

Couvalis, G. (1997). *The philosophy of science: Science and objectivity*. London: Sage.

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

de Regt, H. W. (2009b). Understanding and scientific explanation. In H. W. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding: Philosophical perspectives* (pp. 21–42). Pittsburgh: University of Pittsburgh Press.

de Regt, H. W. (forthcoming). Scientific understanding: Truth or dare? *Synthese*.

de Regt, H. W., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, *144*(1), 137–170.

de Regt, H. W., & Gijsbers, V. (forthcoming). How false theories can yield genuine understanding. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), Explaining understanding: New perspectives from epistemology and philosophy of science. London: Routledge.

Elgin, C. (2007). Understanding and the facts. *Philosophical Studies*, *132*(1), 33–42.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, *71*, 5–19.

Grimm, S. R. (2010). The goal of explanation. *Studies in History and Philosophy of Science Part A*, *41*(4), 337–344.

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

Hindriks, F. (2013). Explanation, understanding, and unrealistic models. *Studies in History and Philosophy of Science Part A*, *44*(3), 523–531.

Kelp, C. (forthcoming). Understanding phenomena. *Synthese*.

Khalifa, K. (2011). Understanding, knowledge, and scientific antirealism. *Grazer Philosophische Studien*, *83*(1), 93–112.

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

Khalifa, K. (2013). The role of explanation in understanding. *The British Journal for the Philosophy of Science*, *64*(1), 161–187.

Khalifa, K. (in press). *Understanding: An epistemology of scientific explanation*.

Knuuttila, T. (2009). Isolating representations versus credible constructions? Economic modelling in theory and practice. *Erkenntnis*, *70*(1), 59–80.

Kuhn, T. S. (1957). *The Copernican revolution: Planetary astronomy in the development of western thought*. Cambridge, MA: Harvard University Press.

Le Bihan, S. (forthcoming). Enlightening falsehoods: A modal view of scientific understanding. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), Explaining understanding: New perspectives from epistemology and philosophy of science. London: Routledge.

Lipton, P. (2004). *Inference to the best explanation*. London: Routledge.

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

Magnusson, M. K., Arvola, A., Hursti, U. K., Åberg, L., & Sjödén, P. (2003). Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour. *Appetite*, *40*(2), 109–117.

Michel, C., Velasco, C., Gatti, E., & Spence, C. (2014). A taste of Kandinsky: Assessing the influence of the artistic visual presentation of food on the dining experience. *Flavour*, *3*(1), 1–11.

Mizrahi, M. (2012). Idealizations and scientific understanding. *Philosophical Studies*, *160*(2), 237–252.

Oddie, G. (2014). Truthlikeness. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy (summer 2014 edition). http://plato.stanford.edu/archives/sum2014/entries/truthlikeness/.

Palter, R. (1970). An approach to the history of early astronomy. *Studies in History and Philosophy of Science Part A*, *1*(2), 93–133.

Roach, J., Ewert, W., Marks, R., & Thompson, B. B. (2013). Unexpected emergent behaviors from elementary swarms. In 2013 45th Southeastern symposium on system theory (SSST) (pp. 41–50).

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Smith-Spangler, C., Brandeau, M. L., Hunter, G. E., Bavinger, J. C., Pearson, M., Eschbach, P. J., et al. (2012). Are organic foods safer or healthier than conventional alternatives? A systematic review. *Annals of Internal Medicine*, *157*(5), 348–366.

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

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/New York: Clarendon Press/Oxford University Press.

Waskan, J., Harmon, I., Horne, Z., Spino, J., & Clevenger, J. (2014). Explanatory anti-psychologism overturned by lay and scientific case classifications. *Synthese*, *191*(5), 1013–1035.

Wilkenfeld, D. A. (2013). Understanding as representation manipulability. *Synthese*, *190*, 997–1016.

Wilkenfeld, D. A. (2014). Functional explaining: A new approach to the philosophy of explanation. *Synthese*, *191*, 3367–3391.

Wilkenfeld, D. A., & Hellmann, J. K. (2014). Understanding beyond grasping propositions: A discussion of chess and fish. *Studies in History and Philosophy of Science Part A*, *48*, 46–51.

Wilkenfeld, D. A., Plunkett, D., & Lombrozo, T. (forthcoming). Depth and deference: When and why we attribute understanding. Philosophical Studies.

Wilson, D. B. (2012). Simplicity in the Copernican revolution: Galileo, Descartes, Newton. In J. Z. Buchwald (Ed.), *A master of science history: Essays in honor of Charles Coulston Gillispie* (pp. 123–133). New York: Springer.

Wilson, M. (2006). *Wandering significance: An essay on conceptual behavior*. Oxford: Clarendon Press.

Winther, R. G. (2015). The structure of scientific theories. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy.