Brit. J. Phil. Sci. **68** (2017), 571–595

Theoretical Understanding in Science

Mark P. Newman

ABSTRACT

In this article I develop a model of theoretical understanding in science. This is a philosophical theory that specifies the conditions that are both necessary and sufficient for a scientist to satisfy the construction '*S* understands theory *T*'. I first consider how this construction is preferable to others, then build a model of the requisite conditions on the basis of examples from elementary physics. I then show how this model of theoretical understanding can be made philosophically robust and provide a more sophisticated account than we see from models of a similar kind developed by those working in the psychology of physics and artificial intelligence.

1 Introduction

In this article I develop an inferential model of theoretical understanding in science. This is a philosophical theory specifying the conditions to be met for a

doi:10.1093/bjps/axv041

Advance Access published on August 12, 2015

scientist to understand a theory.[1] Some philosophers are sceptical of this task and think understanding is a topic for psychology, not philosophy.[2] Thus, to motivate what follows I suggest we think about the following question: given at least some of what philosophers of science are interested in studying, why would we need a philosophical account of theoretical understanding?

One reason is that understanding seems intimately tied to explanation, and we philosophers of science have a long-standing interest in explanation. Explanations help us understand phenomena, so it is an interesting question to ask not only what explanations are but what understanding is too. In pursuing this latter question, we might distinguish between the different ways that we understand and the different objects we understand. For instance, there may be a significant difference between understanding brought about by following a deductive-nomological explanation, and understanding induced via watching a computer simulation. There might also be a difference between what it takes to understand an explanation of some phenomena and what it takes to understand the theory used to give that explanation.[3] Furthermore, these two different objects of understanding may themselves be related in interesting ways. Indeed, Henk de Regt ([2009]) and Michael Strevens ([2013]) both consider understanding a theory to be a necessary condition on understanding a scientific phenomenon. If we join them and assume understanding a theory is a component in understanding some phenomenon, it becomes critical to have a clear philosophical account of what it is to understand a scientific theory. For now, my point is just to illustrate why we might want a philosophical account of understanding theory, and one reason is that it would likely help illuminate our theories of scientific explanation.

Another reason for developing a theory of theoretical understanding is that philosophers of science have always been interested in describing how scientists use theories, and a proper account of scientific understanding that promises to illuminate those discussions would be of great benefit. For instance, in the literature on modelling we see philosophers struggling to iron-out the exact relationship between abstract high-level mathematical theory and concrete low-level models of the phenomena.[4] If we had a better grasp on the role theoretical understanding plays in a scientist's approach to connecting theory to world, we might have a better informed account of this issue.

1 The term 'theoretical understanding' was traditionally used for the understanding of phenomena by theories (for example, by Carl Hempel [1965]). Here the term is used differently. Thanks to an anonymous reviewer for this point.

2 Hempel ([1965]) is famous for taking this position, and more recently we see it in (Nounou and Psillos [2012])

3 For currently discussed accounts of what it is to understand a scientific phenomenon, see (de Regt and Dieks [2005]; de Regt [2009]; de Regt *et al*. [2009]; Grimm [2009], [2010]; Khalifa [2012], [2013]; Newman [2012], [2013], [2014]; Strevens [2013]; Wilkenfeld [2013]; Smith [2014]).

4 See, for instance, the papers in (Morgan and Morrision [1999]).

Additionally, and more concretely, an account of theoretical understanding would likely also help with a potentially devastating problem with accounts of inference to the best explanation (IBE). The problem is that studies in psychology have recently revealed that we humans are really quite poor at evaluating how well we understand explanations.[5] It looks like poor explanatory reasoning extends to scientists and philosophers, and since IBE requires reliable explanatory evaluation, this literature seems to undermine IBE as a reliable rule of inference. If correct, this argument against IBE could prove quite devastating to general philosophy of science. For instance, many scientific realists depend on the reliability of IBE, so this looks to undermine their arguments. An account of theoretical understanding seems likely to shed light on, and perhaps aid in rectifying, such difficulties.

Lastly, an account of theoretical understanding could also shed light on how science makes progress. There are several general approaches to defining scientific progress: progress is just accumulation of knowledge[6]; progress is increasing verisimilitude[7]; progress is increased problem-solving ability.[8] None of these approaches incorporates the idea that progress can be achieved by increasing our understanding. Intuitively, this seems peculiar since some scientists hold that understanding is the overarching aim of science itself. Investigating the nature of theoretical understanding promises to illuminate this discussion and so provides further motivation for our task.

In summary, theoretical understanding is a topic just bursting with potential to illuminate at least four issues in philosophy of science. For ease of reference, in what follows I'll refer to these as our 'traditional problems':

(1) What is the nature of scientific explanation?

(2) How do scientists use theories?

(3) What is IBE?

(4) How does science make progress?

Here's a brief outline of what follows in this article. In Section 2 I distinguish between three different linguistic expressions that might be used to capture the notion of 'theoretical understanding'. I try to disambiguate them and select the most appropriate for analysis in the remainder of the article. In Section 3 I provide an analysis of this expression ('subject S understands scientific theory T') with examples to help motivate constitutive components. In Section 4 I provide an account of theoretical understanding I call the 'inferential model of scientific understanding'. Section 5 presents and explains some

5 See, for instance, recent work by Frank Keil ([2006]).
6 This has recently been advocated by Bird ([2007]).
7 See (Niiniluoto [1999]) for a recent version of this view.
8 Laudan ([1977]) is most commonly associated with the problem-solving view of progress.

empirical work that supports the model. Section 6 provides some thoughts on the ramifications this model might have for our four above-listed 'traditional problems'.

2 The Explicandum/Analysandum

We should now be motivated to investigate theoretical understanding. My task in this section is to disambiguate our object of analysis. The concept 'theoretical understanding' can be captured under a long list of different expressions. I suggest that to determine which is most appropriate we should look at common linguistic usage. When we do, I think we find at least the following three possible constructions for a philosophical account of theoretical understanding[9]:

(1) Subject *S* understands theory *T*;

(2) *S* understands why phenomenon *p* obtains;

(3) *S* understands that *T* explains *p*.[10]

Since we are concerned with understanding scientific theories, not specific phenomena, I think the construction with which we ought to be concerned is the first listed above, which I will paraphrase as: '*S* understands *T*'. To reduce confusion, however, this expression should be distinguished from its above-listed neighbours. I will proceed by looking at examples of each expression, reviewing different interpretations each might be given, and showing that although they are not co-extensive, nothing is lost by adopting my preferred option over the others. Taking the time to do this will significantly reduce confusion later on. Importantly, though, I will not be engaged in discussion over whether any of these expressions reduces to either of the other two (a tempting line of investigation, but a topic for another paper).

I will also assume that our explicandum is 'deep' theoretical understanding, rather than just modest or shallow understanding. For instance, we should read (1) as '*S* deeply understands theory *T*'. Making this assumption should prevent misunderstandings due to underestimating *S*'s cognitive achievement. For example, one might think that if *S* only understood theory *T* in a shallow

9 Expressions (2) and (3) might not strike the reader as obvious candidates. This is good for me, since my sympathies are with (1). However, at least one reviewer has raised concern not to overlook the others, so it seems advisable to address each.

10 These expressions perhaps make understanding too person-specific, and so we might be tempted by a construction that will reflect something less person-specific like, 'the current scientific understanding of theory *T*'. However, I will assume this latter expression can be captured by my analysis to come, though I won't spend time defending that claim in this article. Thanks to an anonymous referee for pointing this out to me.

way, then this might allow an interpretation where *S* fails to be able to use *T* to solve relevant problems, or evaluate the plausibility of *T*, or even gain much at all in the way of knowledge from *T*. I think it would strain our common use of 'theoretical understanding' to allow such a reading; in what follows, let's assume the theoretical understanding we are trying to capture with these expressions is deep.

Let us start then with '*S* understands *T*', which reflects a relation between a scientist and a scientific theory. What does it take to satisfy this expression and is it the same as what we intend when we speak of theoretical understanding? We might get a clue how to answer these questions by first thinking of who satisfies the construction. I take it as uncontroversial that the person in question could be a famous figure from the history of science, like Newton, Maxwell, or Einstein. But it seems obvious that theoretical understanding is also achieved by our current workaday scientists, for instance, a chemistry professor at a nearby university. In fact, anyone whom we might recognize as a 'scientist' would presumably understand some theory deeply. And it doesn't much matter which theory that is, for a philosophical theory that adequately analyses the '*S* understands *T*' construction should provide conditions that when satisfied will in general reflect a subject's cognitive achievement of understanding a scientific theory.

Furthermore, these conditions should capture the essential properties that contribute to the understanding relation, and ideally no others. Those properties will likely be of the kind one finds in a successful scientist, whether they are famous for high-profile discoveries, or hidden in obscurity. And it also seems plausible that such properties will include abilities such as mathematical acumen; theoretical and practical knowledge regarding a physical domain of study; and perhaps the ability to intuitively see the means by which a theory can be used to solve problems. Maxwell obviously possessed such properties, but so too do most chemistry professors. It seems intuitive then that '*S* understands *T*' captures a relation that holds for the broad class of people we call scientists.

I suggest that in satisfying this construction we also mean these scientists not only know the theory they are dealing with, they also know how to use it. After all, it seems that if '*S* understands *T*' is true then this suggests *S* can actually use *T*. What would it mean to understand a theory yet not be able to use it? Furthermore, this 'understanding-entails-use' relation seems to go both ways, such that using the theory seems to entail one understands it. Though this need not necessarily hold—we can often use devices without really understanding them. Still, when it comes to using theoretical knowledge, as in science, it is hard to see how one could use it without theoretical understanding. For instance, to use a theory like classical mechanics to determine an orbital trajectory, it seems one would have to understand it, at least to a fair extent.[11] Besides, as stipulated above, here we are dealing with cases of deep understanding, so appealing to use shouldn't cause us problems.[12]

Consider our second candidate: '*S* understands why *p*'. Why shouldn't we use this to capture 'theoretical understanding' instead of '*S* understands *T*'? To figure out the answer, we first have to distinguish between understanding *p* in virtue of an explanation, and understanding it without an explanation.[13] In each case we also have to disambiguate a deep from a shallow reading of the phrase.

Assume first that we treat understanding as requiring an explanation. In this case there is a shallow and a deep reading. A scientist might be said to shallowly understand why *p* merely in virtue of following an explanation, but little more. For instance, *S* is told that the explanation for why light splits into a spectrum when passing through a prism is that different wavelengths travel at different velocities in glass as opposed to air. There is a very shallow sense here in which *S* could be said to understand why *p*, which is not in virtue of knowing how to use the theory. This form of understanding has the structure '*p* because *e*', where *e* is the explanation just given. But this is not the sense of understanding we are looking for. We want something much more robust, reflecting the understanding a scientist possesses, whereas this interpretation differs little from merely superficial knowledge of *e*.

There is, on the other hand, a stronger reading of '*S* understands why *p*' (again with explanation) where *S* does have significant knowledge or skills related to theory. This can be captured by the construction '*S* deeply understands why *p*'. Here the achievement is significant. For instance, *S* not only knows '*p* because *e*', but also has enough relevant theoretical information to inform how *e* is providing its explanation of *p*, or perhaps how one might use *e* to solve other types of problems, or even why *e* might be the best of the competing explanations. On this deep reading, we are getting much closer to

[11] Of course, one might say that quantum mechanics proves the above assumption wrong, since lots of scientists use that theory without understanding it. But this cannot be right. A scientist may not understand how some phenomenon can be as a theory says it is, but this is wildly different from the accusation. The scientists themselves surely do know how to use the theory, which entails knowing how it works. Whether the theory adequately explains the phenomena is of course a different issue.

[12] There are interpretations of '*S* understands *T*' that must be distinguished from the one we are using; these are the more historical, sociological, or even philosophical readings, which are less interested in how *S* is using the theory, but more interested in knowing the theory's intrinsic or extrinsic properties, and how it came to be. On these readings, '*S* understands *T*' can be unpacked in terms of *S* knowing why, how, where, and by whom *T* was invented/discovered; the social influences that led to *T*'s development and acceptance; or logical properties such as if it is inconsistent, complete, or true. These three readings treat *S* and *T* as objects to be studied in their own right, whereas our interest is first and foremost with *T* as an object to be used. These alternative readings are not of primary importance for our task.

[13] Strevens ([2013]) thinks all understanding is via explanation; Lipton ([2009]) thinks understanding can be accomplished without explanation.

the sort of thing we want from a theory of theoretical understanding in science.

However, it is because the strong reading of '*S* understands why *p*' (with explanation) captures these properties (explaining the success of *e*, enabling further use of *e*, and facilitating explanatory comparisons) that it actually seems quite close to what we already capture with '*S* understands *T*'. If you think about the amount of theoretical information required for the explanation *e* to do all the work it does, then it seems we are not really talking about just an explanation any more, but rather a theory used to give an explanation. And it is *S*'s understanding of that theory *T* that is providing the insight, not the explanation *e* proper. For instance, it is not the explanation 'different wavelengths of light travel different velocities in glass as opposed to air' that allows a scientist to also use *e* to solve other problems. It is the theory of geometric optics that allows her to do this, by giving her a whole set of propositions with which to tackle new problems. So, on the deep reading of '*S* understands why *p*' (with explanation), I think we will find we can do just as well with our previous expression '*S* understands *T*'. Arguably, '*S* understands *T*' is still better at capturing 'theoretical understanding', though. After all, '*S* understands why *p*' seems myopic in scope, picking out only *p* as its subject matter, whereas theoretical understanding is more commonly construed as an umbrella phrase potentially covering a great number of phenomena. I am not proposing any sort of reduction to, or in principle dominance of, '*S* understands *T*' over '*S* understands why *p*' (with explanation), but it may be preferable because it seems broader and better able to avoid potential confusions resulting from ambiguous shallow/deep readings.

Assume now that we do not take understanding to require explanation. Peter Lipton ([2009]) holds this view, claiming that we can come to better understand a phenomenon in four ways. First, he claims we can come to appreciate causal relations by modelling the phenomenon. Second, we can understand *p* better if we have a false, not an actual, explanation. Third, we can understand a phenomenon's relations to other processes and theories by comparing it to exemplars in the discipline. Last, we can appreciate modal properties of *p*, like its necessity, in virtue of argument patterns used to describe it, such as *reductio ad absurdum*. All four of these means of understanding *p* avoid the use of explanation, at least according to Lipton.

We can immediately see some very fruitful possibilities for this non-explanation reading of '*S* understands why *p*'. But one reason against treating this approach as capturing something different from the previous reading (which was with explanation) comes from Kareem Khalifa's work ([2012]), where he shows that for each of Lipton's examples of non-explanatory understanding of why *p*, there exists a correct explanation that would provide greater understanding of *p*. I find his arguments cogent, and think we can reduce Lipton's non-explanatory approach to more exhaustive but explanatory versions of '*S* understands why *p*'. We have just seen above that explanatory versions of this account are to be distinguished from '*S* understands *T*' because, as with the previous reading, they are overly restricted to concerns with only a single phenomenon *p*; whereas theoretical understanding denotes something much broader. So it seems again we are better off with my preferred expression '*S* understands *T*'.

Now let us move on to the third construction: '*S* understands that *T* explains why *p*'. How does this capture theoretical understanding? Again, we have at least two ways of reading this expression. The shallow reading is that *S* only knows that *T* explains *p*, but not how it does so. On this reading, in order for *S* to theoretically understand *p*, all that is required is the very shallow propositional knowledge '*T* explains *p*'. This is surely not what we mean when speaking of the theoretical understanding had by scientists, being as it is merely a descriptive fact about *S*'s propositional attitudes.

On a deep reading, though, we can take the construction to mean '*S* knows how *T* explains *p*'. This is much more interesting and could potentially be preferable to '*S* understands *T*'. Knowing how *T* explains *p* would seem to reflect how scientists use theories to explain phenomena, or how scientists evaluate theories. As such, it might better capture our intuitions about theoretical understanding. But notice how similar this reading looks to the final take I had above on '*S* understands *T*', which amounted to *S* not only knowing the theory, but also knowing how to use it. They are not identical, but '*S* understands *T*' seems to transform into '*S* knows how *T* explains *p*' in the case where explaining *p* just is what *S* knows how to use *T* to do. This may be a limiting case, but we should not be surprised to find '*S* knows how *T* explains *p*' translates into '*S* understands *T*' across different cases. After all, it is quite common to say one doesn't understand something unless one can explain it, and knowing how *T* explains *p* seems to be to know how to use *T* to explain *p*. Although again I want to emphasize that I resist the temptation to argue for reduction here. Since '*S* understands *T*' seems to be the more general expression of what we mean when using '*S* knows how *T* explains *p*', I think it preferable to stick with the former phrase in what follows.

3 Analysis of '*S* Understands *T*'

Having disambiguated our object of analysis, the challenge now is to construct an account of scientific understanding that provides conditions that, when satisfied, guarantee *S* understands theory *T*. Where might we start? Pre-theoretically, we could look to how scientists themselves justify the belief that someone understands a theory. What does it take to display that one understands, say, introductory quantum mechanics? Surely passing exams is a big part of the answer. After all, we take exams to show we have not only learned a theory by memorizing it—a shallow sort of knowing it—but also that we understand the theory; we know how to use it.

It should be no surprise to discover that de Regt ([2009]), after considering some complex examples such as Bernoulli on flight and Boltzmann on the structure of diatomic molecules, concludes that to understand a theory is to be able to use it to construct useful models. In de Regt's case, the idea is that *S* understands a theory if she can use it to build models that explain something. I suppose if we think of solving a textbook problem as a sort of explaining, then our intuitions here seem to line up quite well with de Regt's.

But, you may object, a student of quantum mechanics may be able to plug-n-chug their way to a mathematical solution to a problem without really understanding it. This is true, but as de Regt observes, if we limit the kind of problem-solving to being 'qualitative' in nature then we don't have this worry. A qualitative problem is inherently conceptual rather than mathematical, and is thought to require greater comprehension of the underlying physical principles of a theory. In fact, studies show that novice physics students, who are almost as competent as their expert teachers with mathematical problems, tend to fail miserably when it comes to qualitative problems.[14] To understand, says de Regt, is to be able to make qualitative predictions using *T*, and this rules-out the plug-n-chug objection.

His idea is an improvement on what we've got so far, so let's start building a model of theoretical understanding on the basis of his insight: *S* understands *T* if *S* can use *T* to produce qualitative solutions to conceptual problems. As a slogan, I agree; and the rest of this essay is devoted to unpacking this slogan.

As de Regt is quick to point out, what we have here is only a sufficiency requirement. That means we have a test that is neither exhaustive of the concept 'theoretical understanding' nor provides a fully constitutive account; it doesn't tell us what must be happening inside *S*'s head if they understand *T*. I suggest that although de Regt is moving in the right direction, his sufficiency requirement needs both deeper philosophical articulation and the addition of a necessity condition. If we had both the necessary and sufficient conditions for '*S* understands *T*', we would have a detailed constitutive account of the concept.

An immediate reaction at this point is to rebuke me for adopting an outmoded method of explication, one that goes back to the bad old days of traditional analytic philosophy with its conceptual analysis and logic-chopping.[15] But we need not take things that far to get where we want to go. I want to provide a naturalistically informed account of understanding that appeals to

[14] For instance, see (McMillan and Swadener [1991]).

[15] And, by the way, it isn't such an outdated method. Peter Achinstein ([2001], [2013]) puts it to good use.

what our best sciences have to say about the relevant issues, while also respecting our intuitions about our concept of understanding. So, I won't be looking to suggest any radical revisionary reading of understanding—although I do think the naturalistic analysis I give may ruffle some internalist feathers.

To help make things more concrete, I recommend that we think about a few simple examples of how scientists use theory to solve a problem, and extract from these cases some relevant properties indicative of understanding. I start with elementary physics, and in what remains I can only justifiably claim to cover this field. It seems likely to me that the model that follows will also apply to modern physics and other fields of science, but that for now is merely a conjecture.

Below are some examples of qualitative problems. To test my claims in what follows, it will be best for you to try and answer these for yourself before looking at the solutions, which are given in Footnote 16:

(1) When you drop a ball, it accelerates downward at $9.8\,\mathrm{m/s^2}$. If you instead throw it downward, then its acceleration immediately after leaving your hand, assuming no air resistance, is
  (a) $9.8\,\mathrm{m/s^2}$;
  (b) More than $9.8\,\mathrm{m/s^2}$;
  (c) Less than $9.8\,\mathrm{m/s^2}$;
  (d) Cannot say, unless speed of throw is given.

(2) An inflated balloon with a heavy rock tied to it submerges in water. As the balloon sinks deeper and deeper, the buoyant force acting on it:
  (a) increases;
  (b) decreases;
  (c) remains largely unchanged;
  (d) need more information.

(3) The electrical force of attraction between an electron and a proton is greater on the:
  (a) proton;
  (b) electron;
  (c) neither; both are the same.

(4) If the sun collapsed to become a black hole, Planet Earth would:
  (a) continue in its present orbit;
  (b) fly off on a tangent path;
  (c) be sucked into the black hole;
  (d) be pulled apart by tidal forces.

(5) Because there is an upper limit on the speed of particles, there is also an upper limit on their:
  (a) momentum;
  (b) kinetic energy;
  (c) temperature;
  (d) all of above;
  (e) none of these.[16]

It seems that in each case we follow a simple pattern of reasoning.[17] First, we imagine the scenario described, perhaps forming a mental model. We then try to figure out which of the principles of elementary physics is relevant to the problem. If there are multiple principles, we might select each with an eye to different parts of the problem. We then strategize and apply the selected principle(s) as best we can to produce a result. And again this is repeated for each sub-part of the problem. While all this is going on, we are simultaneously trying to determine the best overall means of proceeding through the problem.

For instance, in response to question (4) above, we first seem to break down the question phrase into two parts. We consider 'If the sun collapsed to form a black hole' first, perhaps imagining parts of the solar system, with our Sun at the centre. The image of the Sun then disappears, replaced by a black patch. There is no light from the background distant stars or galaxies passing through it. We next think of the expression 'the Earth would . . .' and construct from our image something like a 'problem model' by considering what would happen next. We imagine the earth in its orbit and try to picture the next step in its motion, perhaps visualizing it falling in towards the black hole or flying off in a straight line (we might even see orbital path lines as white streaks in the pitch blackness of space). Each image changes as we work our way down the list of possible solutions. Then we pause to actually try to figure out the solution to the problem. We might have a strong intuition about what will happen, but still need to decide on a physical principle that we deem

[16] These are taken from Paul Hewitt's 'Conceptual Physics' webpage: http://www.arborsci.com/60-questions-physics-students-should-know:

(1) (a) is correct. Choices (b) and (d) indicate confusion between speed and acceleration.
(2) (b) is correct. Other choices indicate failure to see that the balloon is compressed by water pressure—and compression is greater with greater depth—displacing less water.
(3) (c) is correct. Other choices don't consider Newton's third law.
(4) (a) is correct, as no values change in Newton's formula for gravitation.
(5) (e) is correct, for momentum and kinetic energy (and, potentially, temperature) approach infinity as speed approaches the speed of light.

[17] Some of these questions may already be familiar to the reader, and in that case it is most likely one just remembers the solution. This scenario is not my concern here. I am trying to evaluate what we do in cases when we actually have to figure out the answer.

appropriate to determining the correct solution. This forms the principle selection phase, where we search our minds for gravitational principles. If we understand Newton's theory, we select his force law: $F = Gm_1m_2/r^2$ and move into the principle application phase. We apply that law to our case, trying to figure out how to use it to conceptually establish the relation that provides the right answer. Once we establish the correct manipulation, we infer that there will be no change in motion to the Earth's trajectory because there is no change in force (because there is no change in value to any of the relevant variables). We conclude (a) is the correct answer. All the while, we proceed by planning how each step is best going to help us get to the correct (or any) solution—judging our progress along the way.

Now, this is all quite simple-minded so far, and there will no doubt be at least minor variations on this general procedure when solving other kinds of conceptual problems. However, I don't think these steps are likely to diverge too much for other tasks. I think the most important factor is that *S* must be able to draw on the appropriate physical principles, as well as the theoretical strategies for the use of those principles, and apply them to solve relevant problems. We can consider both tasks as generally inferential in nature so I think it reasonable to conclude inference plays an essential role in both retrieving such principles and using them throughout the entire problem-solving process.

4 The Inferential Model

On the basis of these simple examples, I would like to suggest the following characterization of a model for understanding a scientific theory—I call it the 'inferential model of scientific understanding' (IMU):

> IMU: *S* understands scientific theory *T* if and only if *S* can reliably use principles, *Pn*, constitutive of *T* to make goal-conducive inferences for each step in a problem-solving cycle that reliably results in solutions to qualitative problems relevant to that theory.

There is a lot going-on in this condition, most of it not immediately obvious from what I have so far argued, so let me first explain some of the terminology, after which I will defend some of the philosophical components. First of all, what is a principle and when is a principle constitutive of a theory? Here I am thinking of principles in several senses. Obviously this concept should cover scientific laws, such as Newton's laws or Maxwell's equations. It should also cover those principles one finds in the articulation of a theory, such as Einstein's two postulates in special relativity. I also intend for it to cover those definitions of what occurs in a system or how to treat components of a system. For instance, in question (4) above, it was necessary to know that a black hole does not have significantly different mass than the star from which it was created, and in special relativity there is a very specific definition for how to determine proper length.[18] It is these sorts of principles that are part of a theory's description to which I refer with *Pn*.[19]

Second, what is a goal-conducive inference, and what constitutes a step in a problem-solving cycle? Here I take each problem one might address with theory *T* to be solvable in a series of steps, much like those I described for question (4) above. A cycle is just a series of steps required to solve a problem. We first develop a situation model that represents the information in question, and we then transform this into a problem model in light of inserting a task to be achieved. There are thus two separable steps in a series, which hopefully results in solving the problem. A third step is to move into principle-selection, where we search our cognitive space for relevant physical principles to help proceed with the task. And so on, as described above, to complete an entire problem-solving cycle. The notion of a step is not clearly definable then, but we can identify one and distinguish it from others by the different cognitive processes involved. A goal-conducive inference at each step is simply an inference that facilitates movement on through the problem cycle to the next step, and ultimately on to the goal: the solution. Of course some inferences take us on a walk-about, but to the degree that they pragmatically serve the purpose of enabling us to reach a solution, they are still goal-conducive.

Finally, what is a problem relevant to a theory? By this I simply mean it has to be a problem solvable by using *T*—one that includes, but is not exhausted by the constitutive principles of *T*. It would be silly to criticize *S* and claim she lacks understanding for failing to be able to solve problems when they are not even solvable with *T*. Still, there is the question of how hard the problems might be. After all, if you make the questions extremely difficult there is a good chance that the chemistry professor down the road won't be able to answer them, even though we'd all consider him to be an expert with theoretical understanding. A reasonable approach then is to use our everyday notion of 'expert' to help answer the question, and say the limit on how difficult a problem can be should match the sort of difficulty we would expect an expert to handle, on average.[20] This covers most of the terms in IMU, but there are several important, more philosophical issues I'll now defend that drive my formulation.

[18] This answer assumes the outer 'shell' of the original star has long since departed, so its mass is ignored.

[19] One might worry the 'star mass/black hole mass principle' is not really a principle constitutive of astrophysical theory at all. I would agree. It is a principle constitutive of theories regarding black holes. Thanks to an anonymous reviewer for pointing this out, as well as helping me clarify the notion of a 'problem-solving cycle' and a 'problem relevant to a theory'.

[20] An objection here is that circularity ensues: I am defining the relevant questions by appeal to our notion of theoretical understanding. In defense, a theory of theoretical understanding should after all reflect our intuitions about experts. So while I admit to circularity, I'm not sure it is vicious.

4.1 Which problems are we talking about?

I have assumed that theoretical understanding is something we find in scientists who are teacher-experts, not just famous discoverer-experts. The kinds of problems these experts are capable of solving, but that novices cannot, are qualitative problems. This raises two important questions: what is a qualitative problem, and how is this concept related to that of explanation?

We have already seen several qualitative problems, but another example would be ‘which of two balls dropped from the same height hits the ground first, a cannon ball or a tennis ball?’. This problem does not request an explanation; to answer it requires conceptual competence, and this is something potentially missing from a simple explanation that appeals only to a physical principle. For instance, this question requires that a problem-solver draw on, and implement, Newton’s gravitational force law for two objects, and ‘see’ that their masses cancel. On the other hand, if we were asked for the explanation of why cannon balls and tennis balls dropped from the same height hit the ground at the same time, answering that question only requires pointing to the appropriate principle (Newton’s gravitational force law). The explanation-seeking version of the second question can be answered much more easily than the one asking for qualitative analysis because we need only appeal to knowledge of which principle is applicable. This is a skill that a novice can learn by inductive generalization from past examples, and requires no demanding conceptual competence. In fact, if the question’s problem model is sufficiently similar to one in memory, there is no need for more than recollection ability. This is not the threshold we should set for expertise, so merely being able to answer an explanation-seeking why-question is not sufficient for understanding because it may not require conceptual competence.

We thus have reason to think that requests for explanations put too low a bar on qualifying *S* as an expert who understands some theory. A novice can explain why two balls hit the ground at the same time: they obey Newton’s gravitational force law. Experts, however, reflect the thicker notion of deeply understanding a theory: they can take a complex conceptual problem and work through it to provide a qualitative answer. So ability to ‘build a model’ in the sense of providing an explanation through only recognizing principle applicability is not what we are looking for in a theory of understanding. We want *S* to be able to solve qualitative problems. This is why I include in IMU that *S* be capable of solving qualitative, and not merely quantitative, problems. Occasionally, of course, there may be a problem that causes *S* particular trouble even though *S* understands *T*. This should encourage us to adopt a picture that does not demand *S* be able to solve every possible conceptual problem related to *T*, but only in general be able to solve such problems.

4.2 Does the solution have to be true?

The answer here would seem to be yes. For how can an expert reflect understanding of a theory while providing incorrect answers to qualitative problems? But this answer ignores the possibility that either the input information or the theory itself may be false. This is really a trivial issue. In a logic problem whose premises are false, the correct application of deductive rules may lead to a false conclusion, even though the argument is valid. So also goes physics. A problem situation may not be real, and more importantly the theory may not be true. That doesn't prevent *S* from understanding the theory. So we do not want to require of *S* that she produce true answers when reflecting understanding of *T*. We only demand she produce answers that are correct according to *T*. That is to say, we want the answers to be true if the input and theory *T* itself were true. This means we want a conditional account of problem-solving.

4.3 Does each cycle have to be correct for *S* to understand *T*?

Not even the best of scientists is infallible. They sometimes make mistakes and, when they do, it may not always be appropriate to diagnose them as being any the less capable with *T*. *S* and *S** may both understand *T* equally well, yet *S* may be having a bad day (headache, relationship woes, or whatnot). So although we would typically assume *S** more accomplished in understanding *T*, even if *S* fails to solve problems at quite the high level of success as *S**, this does not entail that *S* fails to understand *T* as well as *S**—at least for the short run. In the long run, if *S* and *S** differ in their reliability at solving problems with *T*, then we should conclude that they do not understand *T* equally well. This intuition reflects the importance of reliability for our notion of understanding. *S* only has to be as reliable as *S** in problem-solving over the long run; the occasional difference in their performance is not significant. So, I include the clause in IMU that *S* be reliable at solving qualitative problems.

4.4 What is problem-solving reliability?

One intuitively appealing way of characterizing this idea is to treat it in a similar fashion to accounts of reliable cognitive processes in the justification literature. Goldman ([1986]) treats reliability as a property of some process with a high truth ratio: a high ratio of truths being outputted given true inputs. How can this be translated into the problem-solving idiom? We can take the process–input to be the problem situation, and process–output a solution to the problem posed. Then *S* is a reliable problem-solver if they have a high problem-solving ratio: a high ratio of correct outputted solutions to input problems.

Adopting this reliabilist account of understanding avoids the unpalatable consequence of judging *S* to lack understanding just because of a poor cycle or two. The reliabilist move here accords well with our intuitions that mistakes are not utterly devastating.

4.5 Does every specific inference of a cycle have to be correct?

Degree of accuracy in using a theory is presumably a measure of understanding. If *S* makes errors along the problem-solving cycle, then we would think *S* less competent than *S** who did not. Still, there are two senses of 'making a mistake' relevant here: First, one could make an error when selecting a principle, or when making an inference. Those both seem to be damaging moves in a cycle—even one that turns out to somehow generate a correct solution. Here it is appropriate to judge *S* less competent than *S**. Second, there are what we might call minor mistakes in process—perhaps solving the problem in a peculiar manner or taking a wayward path in a derivation. *S* does not actually make a substantive error. This may reflect poor planning, but should it downgrade *S*'s understanding of *T*? I think it should. I assume a logic student, *S*, who completes a derivation in a very concise and economical manner is more competent at logic than *S** who takes a perhaps more obvious but less creative route. The reason being that the former route takes more 'insight' on *S*'s part, and this insight would seem to be a reflection of greater knowledge of which principles and strategies to apply, as well as how to make counterfactual inferences. *S* thus has greater understanding than *S**. Both of these kinds of errors are then detrimental to *S*'s understanding of *T*. Still, we should be consistent with the previous commitment to reliability: it is the reliability of a process that is judged, not single instances. Consequently, it is appropriate to apply these negative judgments to *S* only if she regularly makes either kind of mistake.

4.6 Does each inference have to use a principle that is part of the theory?

This seems to have the obvious answer that no, not every inference in solving a problem must use a principle from the understood theory *T*. For instance, in elementary physics we frequently make inferences using background knowledge from algebra, trigonometry, geometry, calculus, and so on when performing calculations. However, we must keep our focus on qualitative problem solutions, and here we don't use these mathematical theories so much. So, I don't think it wise to allow principles not already within the theory to play a role in our evaluation of *S*'s theoretical understanding of *T*. Better to try to recognize when there is incompetence whether it lies with principles instrumental to solving relevant problems, but external to the theory, or those directly included in any comprehensive account of *T*.

There are many more questions we can ask, answers to which will help resolve more nit-picky details of understanding, but I think the current picture will not require too much more adjustment.

5 Theoretical Understanding and Conceptual Expertise: Empirical Considerations

My overarching idea is quite simple: understanding boils down to inferential ability applied at four separate stages of problem-solving. In this section I want to show how the model of understanding we have just developed has a great deal in common with a model of 'conceptual expertise' that has been developed by psychologists of physics. Given their similarities, I think we have good reasons to take the concept of theoretical understanding to be co-extensive with that of conceptual expertise, and perhaps by explaining the one, we will have an explanation of the other. These empirical considerations provide something like use-novel support for IMU. However, it is important we don't jump too quickly to the conclusion that theoretical understanding can be identified with the psychologist's notion of 'conceptual expertise', for as we have seen in the questions from Section 4 above, there is a lot of philosophical analysis that goes into IMU and which thus goes beyond what I am about to report from psychology. Additionally, it is not entirely clear that psychologists even recognize a single concept of 'conceptual expertise', given that it is typically philosophers not psychologists who worry about how to go about defining concepts.

The model of conceptual physics expertise I will use has been developed by Kurt vanLehn and his colleagues, who study expert cognitive processes. They have studied experts and novices in elementary physics for over thirty years. This account is accepted in the cognitive science community researching 'qualitative reasoning', and one can find support for their claims in their references to a large body of studies performed by those working in the field. In particular their work draws on experimental studies from high-profile researchers such as Michelene Chi, K. Anders Ericsson, Paul Feltovich, Ken Forbus, Douglas Medin, and Herbert Simon. In fact, they describe much of what follows as 'old news [. . .] axiomatic' (vanLehn and van de Sande [2009], p. 365) for their research community, meaning many of the distinctions we are about to hear about are very well established in the cognitive psychology of science. Their overarching point is that expertise in elementary physics comprises the following four kinds of knowledge: 'descriptive phase knowledge, and for each principle, mastery of its applicability conditions, its qualitative confluences, and its planning confluences' ([2009], p. 364). In what follows, I will explain this claim.

Just as we did above, vanLehn and van de Sande note two ways of measuring physics competence: quantitative problem-solving (writing down equations for a system and solving them over several minutes) and qualitative problem-solving (this is more conceptual, with little writing or mathematics, and very quick response times). For ease of exposition, I'll stick with their examples. Their example of a quantitative problem is, 'A bomber is flying at 1000 km/hr at an altitude of 500 m. If the bomber releases a 1000 kg bomb, what is its impact speed?' ([2009], p. 358). On the other hand, a qualitative problem might be:

> A dive bomber can release its bomb when diving, climbing, or flying horizontally. If it is flying at the same height and speed in each case, in which case does the bomb have the most speed when it hits the ground? (A) Diving. (B) Climbing. (C) Flying horizontal. (D) It doesn't matter. The bomb's impact speed is the same in all three cases. (E) More information is needed in order to answer. ([2009], pp. 360–1)

The distinction is quite clearly the same as that we drew above.

Surprisingly, as mentioned in Section 4, qualitative kinds of questions are harder to solve. It turns out most novices who approach the expert level of quantitative problem-solving fail miserably with conceptual problems. This is one reason to take conceptual expertise to require, or be co-extensive with, qualitative understanding. It is this kind of conceptual expertise that differentiates the expert from the novice, and reflects the difference between those who have understanding of a scientific theory and those who really do not. Notice that this is a different reason than we had for preferring qualitative to quantitative problem-solving, which was that the former avoids the possibility that a student could solve a problem without really understanding it. The two reasons are probably related, however, since conceptual difficulty and failing to understand seem intuitively to be two sides of the same coin.

VanLehn and van de Sande identify two temporal aspects of conceptual expertise: a description phase and a phase of applying principles. First, the description phase consists of setting-up a mental model of the problem situation. For the dive bomber example, subject *S* imagines the plane in flight. To do this, *S* has to answer small problems such as whether to ignore friction, whether to consider the earth's rotation, whether we should treat the bomb as a point particle or an extended object, and so on. However, this first stage—description phase knowledge—is represented by vanLehn and van de Sande as the activation of 'categorization rules' to generate a problem model: a mental model recognized as posing a particular problem. We do not really have to believe such rules are adequate for representing our mental states, but for illustrative purposes it will be useful to follow these authors here. Their example consists of the activation of rules that categorize the object and its relevant properties and relations. The rules will be complex, but a simplified example is, 'If <*X* is flying and has wings and has engines>, then <*X* is a plane>'. Another might be, 'If <*X* is a plane, delivers bombs, and delivers bombs whilst diving>, then <*X* is a dive bomber>'. The satisfaction of antecedents in these rules generates the activation of the consequents, which themselves may play the role of antecedents in other rules. Spreading activation within the hierarchy of rules amounts to the representation of a problem situation. Again, we don't have to like this rule-based mental model way of putting things, but it will prove to be useful in what follows, and nothing of philosophical importance to our question hangs on it.

Importantly, this descriptive phase fits nicely with the first stage of our analysis that led up to IMU, where we construct a problem model from a given description. Our approach was far less detailed, with no talk of categorization rules or mental models, but nevertheless the two seem very similar in general terms, and it is plausible to count them as co-extensive.

vanLehn and van de Sande point out that although descriptive phase knowledge will be sufficient for solving some novice-level problems, it is not sufficient for solving the dive bomber problem. To solve this problem, the subject has to recognize the relevance of a specific physical principle (the conservation of mechanical energy) and know how to apply it to draw the correct solution. To do this she must possess not only knowledge of the principle, but also mastery of applicability conditions that are rules that help to determine which parts of the problem fit which parts of a general pattern or schema appropriate to the problem. This is the second phase of problem-solving: applying principles. For this example, vanLehn and van de Sande suggest the following applicability condition rule:

> If there is a moving object, and we have or need its velocity at two time points, time 1 and time 2, and there are no non-conservative forces acting on the object between those two time points, then we can apply conservation of mechanical energy to the object from time 1 to time 2. ([2009], p. 362)

This principle can be put in the simple form of the schematic rule: 'If <condition>, then < principle application>'. The condition may consist of a multiple conjunction of possible states of affairs, each of which may only be analogous to the actual system in the problem situation. The system thus will require the additional ability to make appropriate analogies at this stage. The equation for this principle is $KE_1 + PE_1 = KE_2 + PE_2$.

This second stage of vanLehn's model clearly has strong similarities to our second stage, where we determine which theoretical principles apply, even though theirs includes the much more specific 'applicability condition rule', above. Both account for the step when an expert expresses her understanding of a theory with the ability to draw on the appropriate principles. Again, we have strong reason to think that at this stage conceptual expertise is co-extensive with understanding.

Third, aside from description phase knowledge and applicability conditions, vanLehn and van de Sande explain that S also has to know how to derive appropriate information from the principle identified by the applicability conditions. This information they also characterize as a rule, but it is a rule inferred from a qualitative interpretation of the equation above. vanLehn and van de Sande call this inference a ‘principle confluence’. A confluence is a very important concept. They describe it this way:

> A confluence is stated in terms of a qualitative value system, such as {positive, negative, zero, non-zero} or {increasing, decreasing, constant, non-constant}. For instance, if the algebraic form of a principle is $X = Y + Z$, then a confluence based on this value system {increase, decrease, constant} is ‘If X increases and Y is constant, then Z increases’. ([2009], p. 363)

This tells us that scientists make qualitative confluences on principles applied to the problem model. In a confluence, the goal is to reduce the quantitative precision of a potential description while retaining essential distinctions between components. Instead of continuous real-valued variables, each is described qualitatively, taking on only the value +, –, 0, or non-0. The confluence represents competing tendencies in a system. For instance, in a closed container full of air, an increase in volume will influence the pressure and temperature of the gas. If the temperature is kept constant while the volume increases, then the pressure will drop. These simple qualitative relations are maintained in a confluence, where quantitative rigour is sacrificed, but conceptual competence is tested.

As with description phase knowledge and applicability conditions, each qualitative confluence can be given the form of an if–then rule: ‘if <principle application> and <quantity has qualitative value>, <quantity has qualitative value>, ... then <quantity has qualitative value>. Obviously this stage requires further skills for success, such as the ability to call up from long-term memory, and use, specific arithmetic or algebraic relations.

It again seems intuitively plausible to take this step in the conceptual expertise model as co-extensive with its corresponding step in our model of how understanding is expressed through problem-solving. vanLehn and van de Sande’s third step here is remarkably close to ours, where we use relevant principles along with further strategies to make qualitative inference to a solution. Their notion of a confluence appears to be specifically a qualitative counterfactual, and they do not spend time discussing the use of background strategies, instead only noting the use of arithmetic or algebraic relations. Still, these are very close to what we have got, even if they are described in terms of mental rules.

Fourth, on vanLehn and van de Sande's model experts not only possess the ability to extract qualitative confluences, they also mentally plan for solving problems. They note that experts plan how to apply the principles they identify, figuring out how to get to the conclusion strategically. Novices fail to use them almost entirely. They call these rules 'planning confluences'. They give the example: 'If the definition of kinetic energy applies to an object at time 1, and the velocity of the object at time 1 is sought, then its kinetic energy at time 1 should be sought' ([2009], p. 364). Translated into the form of an if–then rule: if <principle application> and <quantity is known/sought>, <quantity is known/sought>, . . . then <quantity is known/sought>. There are hidden cognitive skills here too. For instance, does the system use a 'hill climbing' or 'means–end' method for problem planning?

And again, the parallel between their conceptual expertise model and our model of understanding of theory is obvious, and compelling. So in summary, according to vanLehn and van de Sande, what differentiates physics experts from novices is a complex group of cognitive abilities that stretch from advanced descriptive knowledge and rules for principle selection, to confluences based on qualitative interpretations of those principles and strategic planning on arriving at the correct solution. Let's call these the 'understanding facts' (they seem to track nicely with our intuitions about theoretical understanding, and are derived from empirical studies):

(1) description phase knowledge;

(2) principle applicability knowledge;

(3) principle confluences;

(4) planning confluences.

According to IMU, these same stages are constitutive of expressing theoretical understanding. And if we are willing to accept that the two are co-extensive, we might quite happily endorse their unity. I argued that we should develop our initial model of theoretical understanding into IMU and this model includes a lot of philosophical baggage missing from the conceptual expertise model given by vanLehn and van de Sande, so we have to recognize that these are not quite the same thing. However, we can say that IMU entails their model so the following conditional holds:

> If *S* understands *T* as described by IMU, then *S* has description phase knowledge, knowledge of applicability conditions, principle confluences, and planning confluences.

The entailment here is important for the following reason: IMU is a philosophical account, developed pretty much from the armchair, that receives strong corroboration from recent work by leading scientists in the psychology of physics. It makes some sense of the expertise literature by showing how that literature links-up with our intuitions about understanding scientific theories. This provides valuable support for IMU. In the closing section, I will briefly survey some of the ramifications IMU has for the four 'traditional questions' mentioned at the end of Section 1.

6 Conclusion: IMU and Our Traditional Problems

So far I have argued IMU provides an intuitive and also empirically defensible account of theoretical understanding. I think when we look back on the four traditional problems listed in Section 1, we can now see how IMU might help make progress with them. The following ideas only reflect the early lines of investigation, but I think they support the development of an exciting and profitable research programme based on this novel, and I think plausible, account of theoretical understanding.

What are the ramifications for the first of our four traditional problems? The problem is to provide an adequate account of explanation. How might IMU help in this regard? I think it can help in at least three respects. First, it can help with the epistemology of explanation. For instance, take the rather long-winded question, 'can one come to understand an explanation for some phenomenon, *p*, if that explanation is somehow "lucky", such as having been provided (correctly) by *S* even though *S* fails to actually understand the theory *T* that is used in that explanation?'. Some philosophers working in epistemology of science might argue that it does not matter if the given explanation 'is' correct, so long as it is lucky, it cannot satisfy the strict demands put on knowledge, and hence one cannot really understand it. Certainly IMU can help here by providing conditions for understanding a theory, and hence aid in determining when *S* does or does not get lucky, without which this problem is intractable. Second, IMU can assist with a related metaphysical question: 'can an explanation even be a genuine explanation if it is lucky, such as having been provided by someone who fails to understand the theory *T* used in it?' With an account of theoretical understanding we have a resource for constructing examples with which to test potential solutions. We can work through examples to establish whether luck really does undermine the ontological status of an explanation. Third, I think IMU could even help with more methodological sorts of questions related to explanation. For instance, although science is primarily concerned with explaining the natural world, it can sometimes have cause to explain its own theories, such as when making a selection between competing theories. And just as an explanation of some phenomenon requires understanding of a theory, perhaps an explanation of a theory itself requires some sort of theory. We are used to thinking of selection criteria for theories in terms of empirical adequacy, simplicity, fruitfulness, and so on. But here IMU can help since by providing a condition on understanding a theory, it displays the success conditions for our theory-selection, super-empirical virtues. It thus tells us when a theory has successfully been understood, which might make it more amenable than another in theory selection.

It is also quite likely that IMU can help us to answer questions about IBE and the progress of science. In particular, IMU is a model of understanding that can help us avoid the subjective sense of understanding and hence help us to better defend IBE from subjective psychological biases. Specifically, as I mentioned in the introduction, IBE is subject to criticism on the grounds that our sense of understanding is highly susceptible to bias, and hence error. IMU is an objective account of scientific understanding for theories that should help us rule-out over-confidence and hindsight biases in selection of theories. We can imagine using it to help figure out if we really do understand a theory as well as we think we do.

Finally, I think IMU might help illuminate the nature of progress in science, for not only does progress seem to come from knowing more about the world, it also comes from understanding it better. If our understanding of theories can help us to understand the world better, then an account that tells us how these might be connected seems like it would be illuminating. IMU tells us when we have theoretical understanding rather than mere knowledge, and so seems like a very useful tool with which to assess our success in science.

These are just a few of the ways in which I see an account of theoretical understanding contributing to philosophical progress in general philosophy of science. As we have seen, IMU is based on work in elementary physics, and although it stands a good chance of applying nicely to modern physics, it surely also just begs to be tested in other domains of science in future work.

Acknowledgements

I would like to thank two very helpful anonymous reviewers for this journal.

Philosophy Department
Rhodes College
Memphis, TN 38112, USA
newmanm@rhodes.edu

References

Achinstein, P. [2001]: *The Book of Evidence*, Oxford: Oxford University Press.

Achinstein, P. [2013]: *Evidence and Method*, Oxford: Oxford University Press.

Bird, A. [2007]: ‘What Is Scientific Progress?’, *Noûs*, **41**, pp. 92–117.

de Regt, H. W. [2009]: 'Understanding and Scientific Explanation', in H. W. de Regt, S. Leonelli and K. Eiger (*eds*), *Scientific Understanding*, Pittsburgh: University of Pittsburgh Press, pp. 21–42.

de Regt, H. W. and Dieks, D. [2005]: 'A Contextual Approach to Scientific Understanding', *Synthese*, **144**, pp. 137–70.

de Regt, H. W., Leonelli, S. and Eigner, K. [2009]: *Scientific Understanding*, Pittsburgh: University of Pittsburgh Press.

Goldman, A. [1986]: *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.

Grimm, S. [2009]: 'Reliability and the Sense of Understanding', in H. W. de Regt, S. Leonelli and K. Eiger (*eds*), *Scientific Understanding*, Pittsburgh: University of Pittsburgh Press, pp. 83–99.

Grimm, S. [2010]: 'The Goal of Explanation', *Studies in the History and Philosophy of Science Part A*, **41**, pp. 337–44.

Hempel, C. [1965]: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

Khalifa, K. [2012]: 'Inaugurating Understanding or Repackaging Explanation?', *Philosophy of Science*, **79**, pp. 15–37.

Khalifa, K. [2013]: 'The Role of Explanation in Understanding', *British Journal for the Philosophy of Science*, **64**, pp. 161–87.

Keil, F. [2006]: 'Explanation and Understanding', *Annual Review of Psychology*, **57**, pp. 227–54.

Laudan, L. [1977]: *Progress and Its Problems*, Berkeley: University of California Press.

Lipton, P. [2009]: 'Understanding without Explanation', in H. W. de Regt, S. Leonelli and K. Eiger (*eds*), *Scientific Understanding*, Pittsburgh: University of Pittsburgh Press, pp. 43–63.

McMillan, C. and Swadener, M. [1991]: 'Novice Use of Qualitative versus Quantitative Problem Solving in Electrostatics', *Journal of Research in Science Teaching*, **28**, pp. 661–70.

Morgan, M. and Morrison, M. [1999]: *Models as Mediators: Essays on the Philosophy of the Natural and Social Sciences*, Cambridge: Cambridge University Press.

Newman, M. [2012]: 'The Inferential Model of Scientific Understanding', *International Studies in the Philosophy of Science*, **26**, pp. 1–28.

Newman, M. [2013]: 'Refining the Inferential Model of Scientific Understanding', *International Studies in Philosophy of Science*, **27**, pp. 173–97.

Newman, M. [2014]: 'EMU and Inference: What the Explanatory Model of Scientific Understanding Ignores', *European Journal for Philosophy of Science*, **4**, pp. 55–74.

Niiniluoto, I. [1999]: *Critical Scientific Realism*, Oxford: Oxford University Press.

Nounou, A. and Psillos, S. [2012]: 'Book Review of *Scientific Understanding: Philosophical Perspectives*', *Studies in History and Philosophy of Modern Physics*, **43**, pp. 72–4.

Smith, R. [2014]: 'Explanation, Understanding, and Control', *Synthese*, **191**, pp. 4169–200.

Strevens, M. [2013] 'No Understanding without Explanation', *Studies in the History and Philosophy of Science Part A*, **44**, pp. 510–15.

vanLehn, K. and van de Sande, B. [2009]: 'Expertise in Elementary Physics and How to Acquire It', in K. A. Ericsson (*ed.*), *The Development of Professional Performance*, Cambridge: Cambridge University Press, pp. 356–78.

Wilkenfeld, D. [2013] 'Understanding as Representation Manipulability', *Synthese*, **190**, pp. 997–1016.