Understanding as Compression

Daniel A. Wilkenfeld[1]

Abstract:

What is understanding? My goal in this paper is to lay out a new approach to this question and clarify how that approach deals with certain issues. The claim is that understanding is a matter of compressing information about the understood so that it can be mentally useful. On this account, understanding amounts to having a representational kernel and the ability to use it to generate the information one needs regarding the target phenomenon. I argue that this ambitious new account can accommodate much of the data that has motivated theories of understanding in philosophy of science, and can also be generally applicable in epistemology and daily life as well.

Keywords:

Explanation, Understanding, Philosophy of Science, Compression, Epistemology

§1 Introduction

What is understanding? This question is both fairly traditional, being a concern of Aristotle (Zagzebski 2001), and surprisingly urgent, given a recent explosion in work on understanding (e.g., the work in de Regt, Eigner, & Leonelli 2009). It also takes on some importance from the hope of grounding the nature of explanation (a pressing issue in philosophy of science) in an account of understanding (a path originally suggested in Scriven 1963, and revisited in Wilkenfeld 2014). My goal in this paper is to lay out a new approach to the question and clarify how that approach deals with certain issues. Much of the literature on the nature of understanding is in philosophy of science (e.g., de Regt forthcoming, Khalifa 2017, Waskan 2006), but there are also important strands of thought in epistemology (e.g., Elgin 2007, Grimm 2016, Kvanvig 2003) and especially in daily life. I aim higher than many such accounts in that I would like to say what is common to and important about understanding in as broad an array of domains as possible. At some level the account developed in any one article will need to be at least somewhat schematic. Nevertheless, enough can be said here to make the new approach promising. My primary guide in constructing the account will be (mostly my own) intuitions about cases,[2] but the goal will be to end up with something that can be justified by more empirical means.

The claim, to be cashed out in more detail in the next section, is that understanding is a matter of compressing information about the understood so that it can be mentally useful. On this

[1] Department of History and Philosophy of Science, University of Pittsburgh

[2] I realize that any reliance on intuitions invites criticism (e.g., Machery 2018), but they strike me as a legitimate starting point. A full defense of this approach is beyond the scope of this paper.

account, understanding amounts to having a representational kernel and the ability to use it to generate the information one needs regarding the target phenomenon.

Here is the general plan for the paper. In §2, I lay out "Understanding as Compression" (UC). I present a pair of guiding examples of understanding and show how they can be accommodated in the proposed framework. In §3 I lay out some of the account's general features and benefits. Then in §4 I consider the relation of my account to other accounts of understanding (and explanation) in both philosophy of science and epistemology.

§2 Understanding as Compression

Relative to most contexts where one would assess understanding, I understand first-order logic.[3] Typically, around the time of my students' first mid-term, they do not. Here is a strange fact about them though—right at the moment they finish cramming for their exam, it is at least conceivable that they could rattle off more facts about first-order logic than I could. For example, they may have memorized a derivation of De Morgan's Laws in the Lemmon-Mates system, which has something on the order of 20 or 30 steps—I certainly do not have that proof memorized. Because I understand, I don't *need* to have it memorized—I have more basic facts memorized, and a bunch of rules, heuristics, and hypotheses that let me recreate the proof fairly easily. The claim of this paper is that being in such a state where you have a representational kernel and an ability to piece together the rest is not just a result of understanding, but what it means to understand. In this section the goal is to see how far we can take that intuition towards building a more precise account of understanding.

The first question one must ask is the scope of one's understanding claim—in what sort of understanding are we interested? Frequently, philosophers of science focusing on explanation (following the pioneering Hempel & Oppenheim 1948) restrict themselves to accounts of explaining why events happen, with forays into understanding laws themselves (again featuring the pioneering Hempel 1965). Similarly, major recent accounts in the philosophy of understanding (e.g., de Regt forthcoming, Khalifa 2017) focus on scientific understanding as well. But we understand all sorts of things—people, paintings, movies, laws, theories, languages,[4] etc. It is natural to think "understanding" might be ambiguous or polysemous, but it would be theoretically beneficial to build as unified an account as possible.[5] One guiding maxim for this paper is a prescription by David Lewis (1980) to "begin by asking for all that we want." (The "we" both here and in Lewis refers to philosophers.) So let us (following Wilkenfeld 2013) take the objects of understanding to be "objects"; this is intended to be maximally general, so as

[3] This is probably not true relative to *all* contexts—if one were considering me and a resurrected Tarski for a job, it might be reasonable to say something like "Tarski is the one who really understands logic," but typically the sort of understanding of first-order logic possessed by someone with a PhD in philosophy will deserve the appellation "understanding".

[4] Languages might prove to be a special case, so I set it aside for now. For a discussion of the relation between understanding objects and understanding languages, see Wilkenfeld (2013).

[5] Unity is often thought to be a virtue of theory choice—see Thagard (1978) for an example.

to include anything that can fall within the scope of an objectual quantifier—events, states-of-affairs, people, propositions, etc.

That said, understanding seems to be a cognitive state of persons. AI theorists might strive to create computers that understand, but note that what they are doing is usually trying to mimic the sort of understanding found in people.[6] Moreover, we would do well to consider understanding relativized to particular contexts; whether or not someone understands first-order logic might vary between a context of assessment where we are deciding whether someone passes a class and a context of assessment where we are hiring a chair for our logic program. (See also n. 3.)

Next, one important aspect of understanding is that what one can do with it be relevant in context (or, as we will see, at least be the right kind of information relevant in a context).[7] Being able to recite the fonts that are used in a particular logic book might technically be information about first-order logic, but it is not information we care about in a context such as evaluating a logic student. One concern[8] with the approach of demanding the products of understanding be contextually relevant is that, intuitively, someone who can answer more questions about a given topic would seem to understand it better, even if understanding is being attributed in a context where those *particular* questions are not immediately relevant. What is going on in this case is that there is a certain *kind* of information—for example, causal information or proof-theoretic information[9]—that is fixed by context, and the better understander could generate more information of that kind. This is potentially constitutive of greater understanding even if the information is not *immediately* relevant. As such, the account of understanding should reward an agent for being able to produce more of the kind of information picked out as relevant by the context.

Finally, it must be decided whether understanding needs in some way to be accurate—for example, did phlogiston theorists under combustion? There are conflicting intuitions among philosophers in this area. de Regt & Gijsbers (2016) claim that they did, [10] whereas others (e.g., Khalifa 2017 on "proto-understanding") have been a bit more guarded. At this point in the dialectic, nothing I have said settles the matter of the extent to which understanding can tolerate inaccuracies one way or another. However, as I have elsewhere (Wilkenfeld 2017) argued that accuracy is a good-making feature of understanding, I leave an accuracy clause in until there is good reason to take it out.

Combining these features with the approach suggested by our first-order logic example, we get:

---

[6] Perhaps institutions can understand things as well, but if this proves derivative on persons' understanding it need not complicate the picture here.

[7] Wilkenfeld (2013) argues that the relevant context is the context of attribution.

[8] I thank an anonymous reviewer for this point.

[9] Obviously there will be some vagueness in how kinds of information are individuated. Hopefully some vagueness can be tolerated at this point; if not I leave the challenge of precisification to a later date.

[10] Le Bihan (2016) has a different compelling argument for "enlightening falsehoods".

- Understanding as Compression (UC): A person p_1 understands object o in context C more than another person p_2 in C to the extent that p_1 has a representation/process pair that can generate more information of a *kind* that is useful in C about o (including at least some higher order information about which information is relevant in C)[11] from an accurate, more minimal representation.[12]

One idea that suggests itself from this formulation is that what marks (better) understanders is the fact that information about is in some way (more) compressed. To be a bit clearer about what this amounts to, we'll take a brief foray into the sort of compression envisioned. Nevertheless, on the present proposal UC captures what it really means to say that someone understands, with a more fleshed-out picture just providing details on one way understanding might be implemented. It is certainly conceptually possible that our concept of **understanding** might, unbeknownst to us, latch onto one privileged means of data compression. However, it seems more likely that it would pick out a *kind* of cognitive state that could be realized in multiple ways

§2.1 Compression

When one has a block of information, one can store it more or less efficiently. One could store each individual datum, but to do so would be to overlook regularities in the data that would enable more efficient summarizations. For example, if one has an arbitrarily long sequence that just alternates between 0s and 1s, memorizing the whole sequence would be both unnecessary and odious—one could simply remember the rule "it alternates between 0s and 1s" and the fact that it starts with a 0.[13] One can imagine every piece of information one receives as more or less constrained by the information one has regarding the system (one's hypotheses) and the specific data one has already encountered.

The goal then is to construct an account of understanding in terms of the minimal information required (e.g., general mathematical strategies and a few axioms) to re-create the maximum information possible (e.g., full proofs). There are several ways to describe how much information is compressed—in principle I expect most could work for present purposes. I will adopt a variant of the Minimum Description Length[14] (MDL) described in Grünwald (2005), where what is minimized is the sum of the length of a description of a general hypothesis about how the data is structured combined with a description of the specific data in terms of that hypothesis. One nice feature of the MDL framework is that it turns out not to matter that much

[11] This clause is required to avoid the counterintuitive result that someone who can generate more of the right kind of information but cannot discern which information is relevant understands better than someone with more targeted understanding. I thank Colin Allen for this point.

[12] Strictly speaking, there should also be a provision that the representation does not generate too much false information, lest one think that a contradiction plus classical logic was maximally understanding-producing. I doubt most philosophers (and certainly most non-philosophers) are worried about the implications of logical explosion for an account of understanding, so I leave this condition implied for the remainder of the paper.

[13] There are other ways to simplify encoding this sequence.

[14] I will be using the "Crude" two-part version—complexities introduced in the reduction to the more advanced one-part version do not matter for present purposes.

what language is chosen, so long as it's a sufficiently expressive language and the data being encoded is rich enough.

Grünwald (2005, p. 5) characterizes the central insight of the literature on MDL thus:

> …any regularity in the data can be used to compress the data, i.e. to describe it using fewer symbols than the number of symbols needed to describe the data literally. The more regularities there are, the more data can be compressed. Equating 'learning' with 'finding regularity', we can therefore say that the more we are able to compress the data, the more we have learned about the data.

MDL proponents put forward MDL as an empirical claim about how people actually learn. My move is to claim that, at least at some level of abstraction, what they have (also) identified is a conceptual truth about what it means to learn well—i.e., to understand.

§2.2 Understanding

I can now put the view a bit more precisely:

> Understanding as Compression* (UC*):
>
> A person p_1 understands object o in context C more than another person p_2 in C to the extent that p_1 has a representation/process pair that can generate more information of a *kind* that is useful in C about o (including at least some higher order information about which information is relevant in C) from an accurate, more minimal description length.

Before proceeding any further, it makes sense to respond to some of the more glaring objections to UC*.

First, is *any* compression of information sufficient to improve understanding? Technically the answer should be no—if there happen to be a string of similar symbols in a proof, remembering a quick rule for those symbols and then memorizing the remainder by brute force would not seem to aid understanding. Here it is important to note that what is presented by UC is what understanding consists in for *persons*—what is relevant is what would actually enable a person to reconstruct the relevant information. As such, most sorts of compression that people would actually find useful do seem to qualify as improving understanding. Memorizing a syntactic pattern of symbol-strings is a fine way for a computer to encode a proof, but it would be extremely surprising for any *person* to be able to encode a proof in that way. The result that the sort of compression people are likely to use constitutes understanding whereas other sorts generally do not is perfectly reasonable—understanding is something we care about as it exists within people. Indeed, to the extent that one thinks that understanding might be characteristically *human*, then it is a feature rather than a bug if it ends up being tied to how we actually process information.

One can come up with contexts where a human might compress information like a computer, but in those contexts it often seems like there is some understanding to be had. For

example, for most of us, being able to compress information about specific color patterns would be a horrible way to understand a painting. But that isn't always the case. If one were an artist trying to understand a propensity of the painting to provoke certain visceral reactions, then storing color patterns might be a good way to do so. Even if one is not an artist, knowing "the next splotch is probably some shade of blue" can be critical in understanding Picasso's "The Old Guitarist".

As another objection, one might think that I can have understanding of why an event happens by knowing some facts about its causal history, even if none of the information is relevantly compressed.[15] For example, my six-year-old daughter understands perfectly why I was late from work today when I say that a meeting ran long. This can be true even if she has no broader hypotheses from which to generate additional facts. One might well wonder then whether she has any compressed information on this matter at all—there seems to be nothing to unpack beyond the singular causal claim "daddy was late because his meeting ran long".[16] My daughter in such a scenario certainly *knows* why I am late, and UC can even allow that she has some understanding of why I am late.[17] All UC demands is that someone who can generate either more information from the same description or the same information from a more minimal description understands *better*, and that seems to be the case here. My spouse, who has been stuck in meetings before, can generate information about the sort of things that can make meetings run late, the sorts of feelings I might have had during the meeting, etc.; it seems reasonable to say that she understands my lateness to a greater extent than my daughter. This is all that UC demands.

§2.3 The Logic of the Invisible Hand

To see how the account does, I will present two examples—one of current understanding a student might attain, and one of a more historical example of understanding achievement.

§2.3.1 An Illustration: Soundness of a Natural Deduction System for First Order Logic

Suppose I want to understand a standard proof of the soundness of a natural deduction system for first-order logic. No doubt many readers will have been made extremely familiar with this proof at some point in their education. Here then is a challenge to such a reader: recite the proof, from beginning to end…NOW.

This challenge is patently unreasonable. Having every word of such a proof memorized—as some of us were once forced to memorize the "Tomorrow, tomorrow" speech

[15] I owe this objection to Edouard Machery, with related comments from Kareem Khalifa.
[16] I will return to a discussion of understanding singular events in the discussion in §4.1 of recent work by Skow.
[17] I follow Kelp (2015) in thinking that attributions of outright understanding are derivative on attributions of better or worse understanding, combined with contextualist semantics. Defending this claim is outside the scope of this paper.

from *Macbeth*[18]—would be a ridiculous way to store and recall information. Not only could we not do so orally, but the same mathematician, asked to write out the proof three times, would be highly unlikely to use the exact same words. We understand the phenomenon by means of a proof that is not stored explicitly at any cognitive level—what we have instead is enough higher-order information that we could recreate the proof. (I take it that "the proof" is not individuated by the precise words the recreator happens to use, but by its underlying mathematical function—it would be odd to count every re-articulation as a new proof.)

The central idea is that what we really possess when we understand a proof is a representation of the essential points, a representation of the system of which the proof is a part, and a process that tells our minds how to fill in the rest of the details. Put in this way, the connection to information compression becomes fairly manifest. A picture file does not need to actually encode every pixel—if it has two non-adjacent pixels that are both the same shade of red, it can encode large swaths of the information between them by a simple rule of the form "and it goes on like that." The contention here is that the same could be said for a soundness proof—one remembers certain facts about how the proof is structured, and one has enough general mathematical know-how to simply store "and it goes on like that." Putting the point in line with the MDL framework above, one has a complex hypothesis about the general structure of the proof (e.g., that it's a strong induction on proof length using the satisfiability conditions specified by the semantics) along with a description of the rules of the system.

I am now in a position to show bit more clearly why such compression is important for understanding. I begin by noting that mere memorization is not sufficient for understanding. When I was in graduate school, a student in the year behind me, in order to be able to move maximally quickly on an exam, memorized an entire soundness proof for a first-order system containing the Sheffer-stroke as the only connective.[19] He was, to a certain extent, like a logician in Searle's (1980) Chinese Room. Here is a fact about him—if he had forgotten one step—even potentially a relatively trivial step—in the middle of the proof, he would have been lost. It is reasonable to ask what a seasoned logician has that my friend lacked, and the glaring difference seems to be the latter's ability to reconstruct the proof from simpler representations. Note that *some* representations are potentially still essential—if the logician were to suddenly "blank" on the actual rules of the system, she would be in a fairly similar predicament as my friend—but the total output-to-scaffolding ratio would be significantly higher for the logician. Moreover, some processes are essential as well—if the logician were to suddenly forget how to do proofs by induction, she would get stuck fairly quickly.

§2.3.2 The Invisible Hand

At one point, it appeared somewhat mysterious that effort seemed to be allotted in efficient ways to advance society, and that no amount of state control seemed to improve

[18] Incidentally, note that UC can be applied to understanding the speech as well, if rather than having it memorized one has a general sense of what the speech was aiming for and the sorts of language likely to be used.
[19] This is a true story.

matters. Various theorists produced abstract models to explain these phenomena, but most had very little mark outside of theoretical economics. The turning point was of course the publication of Adam Smith's 1776 *The Wealth of Nations*, which was the most important historical work in advancing understanding of how goods can and should be distributed for optimal economic advancement. Since then, both professionals and laymen take for granted that (under a wide array of circumstances) allowing prices to fluctuate freely can do better for an economy than could be done by a central planner.[20]

The remarkable thing about *The Wealth of Nations* is that most of the important results can actually be encoded very simply in high school economics textbooks, and that people can understand important large-scale facts about the structure of society simply from remembering a few simple principles about supply and demand. What Smith enabled us to do was to produce information about macroeconomic phenomena (I'm relatively confident regarding what will happen to oil prices when a hurricane hits the Gulf Coast) without needing to encode a particularly large amount of information about the actual structure of society. We can also see how *Wealth of Nations* creates minimal understanding in some people but deeper understanding in those who are better able to unpack the implications of that central kernel. (Smith himself foresaw how his principles would lead to the problem of monopolies, which fact more conservative politicians otherwise espousing his view typically overlook.)

(Incidentally, the same points made about supply-and-demand for Smith could be made about the philosopher of science's more stock example of the laws of motion for Newton. I prefer the Smith example because it brings to the fore how compression/understanding can be effected by re-description rather than always requiring wholly new principles (most of the ideas in Smith had been "in the air" for many years), let alone new data.)

§2.4 Whither Compression?

One might be worried that what is doing all the work in the foregoing examples is the requirement that one be able to do useful things, not that one's information is compressed. One commentator proposed that an account that associated understanding with useful knowledge would fare just as well. However, this overlooks the generative nature of understanding, which is at best only incidental to knowledge. Real understanding requires the ability to take what is in one's cognitive possession and apply it to a new array of cases. Once one grants that, it would seem that the ratio of old information to new information should be important. It would be odd if being able to generate one new consequence of a mathematical theory were on a par with being able to generate wholly new proofs. Yet once one has further accepted that the ratio of encoded information to producible information is important, one has essentially accepted the importance of compression (in addition to usefulness). Everything else is just working out the details.

[20] By "better" I mean more efficient coordination of production and consumption and a more efficient allotment of capital to productive areas.

§3 Taking UC for a Spin

In this section, I take the truth of UC as a working hypothesis, and draw out some of its positive features and implications.

§3.1 Complexity and Simplicity

Here are two platitudes that strike me as true of understanding:

1) In some contexts, understanding something apparently simple can require appeal to the exceedingly complex. That objects tend to be attracted to each other in proportion to their mass and inversely proportional to the square of their distance is fairly simple. However, one might think that real understanding of this principle requires being able to relate it to the distortion of space-time by the presence of mass, which is (to some of us) very far from simple.
2) In other contexts, understanding something complicated can be accomplished by relatively simple pictures. There are several classes of examples of scientific understanding that have this simplifying structure:
    a. Lord Kelvin famously said that he was never satisfied until he had a mechanical model of something, and in science classes the world over by (for example) we see the use of orrerries to engender understanding of solar phenomena.
    b. Idealizations such as the ideal gas law simplify understanding of the behavior of innumerable forces by bracketing such complexities as the gravitational pull between molecules.[21]

On the surface, these two platitudes seem to pull in opposite directions—does understanding simplify, or does it complexify? One could of course say that it sometimes does one and sometimes does the other. This might be expected if what were really being tracked were something else—for example, if all understanding involved the grasping of causes, and causes are sometimes simple and sometimes complex. However, that seems to miss out on the role of simplification and complexification in the examples. One understands the complex phenomena in example (2) precisely *because* one has related it to something simpler, not in virtue of something else that happens to simplify. Less obviously, one might think that part of the understanding that comes from examples of type (1) is precisely that one appreciates the complexity of the underpinnings of everyday phenomena. Part of what it is to understand

[21] The role idealizations play in understanding the Ideal Gas Law is a vexed issue, with some (e.g., Elgin 2007) arguing that it demonstrates that false premises can generate understanding, and others (e.g., Khalifa 2017) arguing that it shows no such thing. However, all parties to this debate (so far as I know) agree that there are simplifications and idealizations somewhere in explaining/understanding the Ideal Gas Law, which is all I need for present purposes.

One might worry that the accuracy requirement in UC begs the question against views that argue that understanding need not be factive. Such a result would not be horrible, as the arguments (e.g., Kalifa 2017, Mizrahi 2012) that the inaccuracies play no role in understanding are at least moderately compelling. However, if one is otherwise convinced by non-facitivist arguments (e.g., Elgin 2007, Kvanvig 2003) then the accuracy requirement can simply be jettisoned.

sometimes is to simplify and sometimes is complexify. (Since I will be ultimately arguing for the simplifying nature of understanding, I need not insist on the importance of complexification.)

One description of UC, which is neither wholly charitable nor wholly off base, is that one understands something complicated by getting it simple enough to keep in one's own head. This very neatly captures platitude (2), but seems to fail miserably to capture platitude (1). Indeed, this is generally taken to be the major obstacle to giving simplicity any privilege in understanding understanding—if there's one thing we've learned from the history of science, it's that the universe is in fact fantastically complicated (Friedman 1974 p. 10 makes a point somewhat like this). For example, apparently simple heat transfer is the result of the semi-random motion of huge numbers of molecules exhibiting certain statistical properties—if our goal was to simplify, going from thermodynamics to statistical mechanics seems to take us in the wrong direction.

Here though is the rub: while one can delve almost arbitrarily deeply into the details of statistical mechanics, people asking after how statistical mechanics works do not typically require that someone have unfettered access to all those details. Rather what understanding generally amounts to is a higher order representation of the complexities of the theory, combined with a process for recreating the details not explicitly present in the representation. For example, we do not typically require that anyone who claims to understand the properties of heat by reference to statistical mechanics must be able to rattle off Liouville's Theorem—in many/most contexts, they simply need to know the general principles by which ensembles of randomly moving molecules create heat.[22] From the mere assumption that understanding is a cognitive state (broadly construed), we can conclude that the understanding itself is constituted by the sort of information and information-processes that it could actually be in the mind of the understander. But that means that the understanding itself must have a certain degree of simplicity, imposed by human cognitive limitations.

Perhaps one might worry that I am being too promiscuous in my attributions of understanding, and that understanding of heat really does require one have memorized the requisite equations themselves, rather than merely higher order information about what those equations mean and how to manipulate them. The point from the last paragraph—that the understanding itself must have a certain degree of simplicity—can still be maintained. First, suppose one is considering the case where what is being understood is only finitely complicated, as in the example of a specific proof as discussed above. Even in that case, I argued that having a representation that was itself as complex as the theory was unnecessary and, indeed, probably unwanted. Next consider the case where the aspect of the world one is understanding is vastly complicated—this invites the following dilemma:

1) Label the minimal mental state (presumably a representation/process pair) that suffices for understanding of A as "R".

[22] In more complex contexts we would want the more general laws describing entropy—the general point is that what one needs to be able to access immediately are whatever deeper laws are relevant.

2) For a given person p in context C, either R is too complicated to be represented in p's mind or it is not.
3) If it is not too complicated, then R is a simplification of A.
4) If it is too complicated, then it could not possibly be what we denote when we attribute understanding of A to p in C.

If the world is complicated, then when we attribute understanding, we must be identifying a cognitive state that is *about* a more complicated state of the world. What UC does is specify the way in which it can be simpler yet at the same time constitute understanding—it enables one to reconstruct as much useful information as one can about the target phenomenon from information (broadly construed) that one can actually possess.

My claim is that there is a bit of sleight-of-hand when arguing that understanding cannot be simplifying. Admittedly, the world is made up of complicated things—quarks, randomly moving molecules, racial tension—but this would only imply that understanding could not be simplifying if the understanding itself were constituted by those features of the world. But it isn't—understanding is a product of *how we think about* those things in the world, and the way we think about complicated things could be far simpler than the things themselves.

§3.2 Thought and Action

One point of contention among philosophers of understanding is whether understanding is primarily a matter of what one knows/represents or of what one can do. It has recently been argued (Wilkenfeld 2017) that the two approaches are in fact complementary. Once we start thinking of understanding in terms of compression, this result falls out more-or-less automatically.

The official statement of UC requires of understanding a certain degree of accuracy but, as noted, this is not wholly essential to the view. One advantage of UC* framing compression in terms of the minimum description length paradigm is that it is thought (Grünwald 2005) to be more neutral regarding whether hypotheses need to be strictly accurate.[23] That said, it seems reasonable (though admittedly not entailed by UC) that if understanding comprises representation/process pairs that let you generate relevant and useful information, that it would be a good-making feature of that understanding if the representations were more accurate rather than less.

The usefulness criterion falls fairly directly out of UC, based on the fact that understanding is only characterized in terms of one's ability to decode a particular representation. I also contend that the processes that are involved in unpacking a representation closely mirror those activities that philosophers have argued are constitutive of genuine understanding. For example, on Hills' (2016) account, the requisite abilities have to do with being able to construct, modify, follow, etc. an explanation—yet (for example) constructing

[23] Grünwald argues that it is more neutral than Bayesian learning accounts regarding whether or not a simplifying hypothesis needs to reflect a true probability distribution or merely be the best description of the available data.

explanations is just the sort of thing one would frequently do when unpacking a representation to get information about a target phenomenon. Likewise, de Regt (2009) requires that one be able to recognize characteristic consequences of a theory—yet again, this is precisely what one often would be able to do if one has UC understanding.

§3.3 On the Normative Value of Compression and Understanding

One might well wonder why, if UC were true, we should care about understanding. Not all philosophers will be motivated by this concern—some of us think limning a concept such as **understanding** and seeing whether there is anything in the world that answers to such a concept is an intrinsically interesting project—such philosophers can happily skip ahead to the next section. However, others see things differently. Craig (1990) argues that rather than explore the contours of how people actually utilize the concept of knowledge, we would do well to consider what concept of knowledge we *should* have. In a similar vein, Woodward (ms) argues that philosophy of science at its best is a normative enterprise, as its methodology is not conducive to illuminating actual concept structure. It is thus reasonable to ask about the value that UC assigns to understanding.

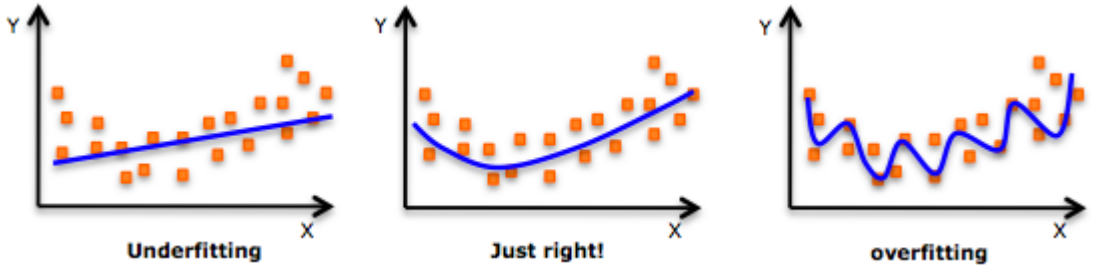
One small benefit of storage compression is, of course, that storage is compressed. While the human brain is a phenomenally complicated device, the human mind is far from unlimited. This is particularly true when it comes to indefinitely extendable sequences--to take a trivial example, almost no one remembers the first 50 Fibonacci numbers, but most people reading this understand their underlying logic (i.e., have a short description of it) and could easily recreate them. The same logic also holds for finite but ridiculously large numbers—if I have an unopened letter from a romantic partner, understanding her thought processes and the sorts of letters she is likely to write me enables me to recreate far more about the content of the letter than I could recreate for a letter from a stranger.

However, the larger benefit of compressing information is that it enables us to avoid overfitting the available data.[24] With any set of data points, there are literally infinitely many different patterns that would generate good (as determined by contextually relevant standards) approximations of the data, of which a very large number are at least in principle memorable. Typically what we learn about the world will come from apprehending some signal (i.e., true features of the world) and some noise (i.e., random fluctuations, often but not always due to how we measure it); thus, if we were to remember a pattern that predicted all of our data exactly we would factor in noise when making future predictions. To take a stock example, if we want to make sense of the following data[25] we are far better off remembering something like "it's a

[24] There are other technical ways of avoiding overfitting—see e.g., Forster & Sober 1994. These accounts are not competitors.

[25] Source: https://medium.com/towards-data-science/train-test-split-and-cross-validation-in-python-80b61beca4b6

quadratic function X with an error term" than either "it's a linear function with an error term" or "it's a decic function".[26]



Avoidance of overfitting is one standard justification for using representations with minimum description length (Grünwald 2005) as an efficient learning tool. If utilizing this tool just is what it means to understand, then understanding will inherit all those virtues. And if we want a way to demarcate which people have the appropriate descriptions to generate accurate future data, then having a concept of **understanding** would thereby be useful.

There is another corollary benefit to being able to talk in terms of who understands and who does not. One might—perhaps having just read Kripke's (1982) exposition of Wittgenstein—wonder whether there is any principled reason to favor the quadratic function or whether we simply treat some people (the quadratic-function-knowers) as going on in the same way we would. Yet note that even if this is all we're doing, it is helpful to have a characterization that demarcates between people who are tracking the same pattern we are and people who are not. Even if there were nothing *epistemically* better about understanding vis-à-vis our future accuracy, there would be something *socially* advantageous about sorting people into those who understand (i.e., use a quadratic function) and those who don't. (Of course we could still worry that people whom we deem to be understanders are using a function that just looks like our quadratic function until some large but finite n, after which it suddenly veers off course—nothing that has been said is intended as a solution to Wittgenstein's more horrifying concerns.)

§3.4 Understanding and Human Processing Limitations

One feature of UC that might strike some as suspicious is that it ties what counts as understanding very tightly to the question of what human beings can actually remember and cognize. This could lead one to several concerns, which I briefly address in this section.

[26] Slight hiccup: according to UC, one would be better off recording the decic function than the quadratic function along with the precise error of each of the 22 data points, whereas intuitively if one could remember the latter it would seem to accord better understanding (as it makes reference to the "real" function). Here I must do a bit of bullet-biting, and acknowledge that if we want to understand these data and no other, the decic function does accord more understanding. However, if we want to understand the place of these data in the broader world, the quadratic function and accompanying error terms gives us a better sense of how the world as a whole really is (because it lets us more accurately generate information about future data).

§3.4.1 Penalized Good Memories

One might worry that UC unfairly penalizes people who have excellent memories, as they will have no need to compress and so no need to understand. However, the fact that one has the cognitive capacity to explicitly represent something without compression does not mean that it's a good idea to do so—there are processing tradeoffs even among savants. Moreover, there is nothing in the account that prohibits having information stored in both compressed and non-compressed forms. I have the entire multiplication table from 1-12 memorized, but this does not entail that I don't understand multiplication—I also have a compressed representation/process pair that takes my explicit representation of the addition table and a process for iterating addition problems and generates information about multiplication.[27] Most importantly, everyone has some things that come naturally and some things that come with difficulty. While my daughter is learning how to multiply, she compresses information about the multiplication table by having representations of addition problems and processes that can convert successive addition problems into multiplication problems. It seems right to say of her that she understands single digit multiplication. By contrast, in addition to my own understanding of multiplication described by the above, I can also use my explicit (non-understanding) representation of the multiplication tables to *reach past* my memorized comfort zone to understand (for example) algebraic manipulations of factors and polynomials.

§3.4.2 Penalized Calculators

One might worry that on this theory, what one understands is a product of what one can actually produce with the representations one has. However, computational aids are not part of a person,[28] and so their products do not constitute part of what one (qua person) can recreate. Therefore, people who need calculators don't understand things.

If this were a result of the theory, that would be fairly dreadful. Thankfully it is not. The question on which to focus is what sorts of things scientists really understand when they require computational aids. They typically understand at least three things, all of which are UC would count as genuine components to understanding:

A) They understand what *kind* of result the theory should predict, even if they cannot compute the precise result,
B) They understand the theory well enough that they can use the computational aid to get a precise result, and
C) Having done calculations pertaining to phenomena, they can understand the properties of that phenomenon by means of remembering the result of those calculations.

Regarding (A), de Regt (forthcoming) has arrayed vital case studies displaying how at least frequently precisely the ability to make good qualitative predictions in the absence of precise calculations is the hallmark of understanding. Regarding (B), the fact that we would in most

[27] Notice that on this point even God can (and presumably does) understand things, as infinite information and processes include both compressed and non-compressed varieties.
[28] I leave to a future project exploring the interactions of UC with theories of extended cognition.

instances attribute understanding of a phenomenon to a scientist whose computer just froze provides some evidence that what we are demarcating by saying that she understands is not anything that depends on the computer producing results. (C) is most to the point—Saul Perlmutter understands Type Ia supernovae *now* largely in virtue of the fact that he remembers having run calculations on them *before*, even if he neither now nor then could recreate every step of the calculation unassisted. But that is all fine on UC—UC does not demand that understanding o requires being able to recreate any information about o without access to any specific fact—it is frequently the case that some specific data (such as the results of calculations) will need to be stored explicitly. UC does predict that he understood such supernovae better after the math had been done then he did before, but that seems like it should be right as well.

Notice that in all three cases, the subject understands in virtue of the cognitive work they can do, rather than somehow getting credit for the calculation itself. That is as should be, as otherwise UC would not count them as possessing the representation/process pair that generates the useful results. For example, if they simply knew which button to push to launch a program that generates extensive predictions without further input, they would not count as understanding (much). This point also defuses another possible objection, which would be that UC makes understanding too transmissible via testimony (cf. Hills 2016). If one simply knows whom to ask for information, that does not constitute understanding on UC, as one does not actually possess the relevant representation/process pair oneself.

§4 UC and Other Accounts

In this section, I explore the connection between UC and other accounts of understanding in the literature, as well as a few accounts of understanding that can be read off of accounts of explanation. Not surprisingly, I will argue that the advantage lies with UC; however, all of these accounts have been valuable, and so part of the goal will be to see what I can coopt from their plausibility. Obviously, I will not be able to survey every account of understanding on offer. Many (e.g., de Regt forthcoming) provide specific accounts of what it is to understand in some particular subset of sciences, and will often fall out as plausible pictures of specific domain-restricted applications of UC. In this section I focus more on those accounts of understanding that could be seen as undermining UC (§4.1-§4.2) or so close to UC as to render the latter superfluous (§4.3-§4.5). I also look (§4.6) at how UC bears on two key questions found in the epistemology literature.

§4.1 Causes

Though the view is rarely defended in the literature on understanding (the closest are the "grasping" accounts discussed in the next section), one very natural view that one gets from the literature on explanation is that one understands an event if and only if one can identify its causes (something like this might be in the background of parts of Woodward 2003). UC grants that, quite typically, causal claims will be conducive to understanding. This is so because—on any

plausible account of causation—causal claims must have some degree of generality (see Woodward 2006 for a nice discussion of some axes of generality). Thus one can encode that certain sorts of causes have certain sorts of effects. Combining that with one's representations of specific facts, one can reconstruct all sorts of other information. Nevertheless, the lack of any explicit causal vocabulary in UC suggests that causal information is neither sufficient nor necessary for UC understanding.

UC is not actually committed to causal information being insufficient for understanding—it is possible that causal information always involves such a degree of generality that one would be able to reconstruct information. However, it is highly unlikely that such information would always be relevant in every case. This is plausibly the correct result—causal theorists have known all along that some claims (such as the effect of the gravitational pull of Mars on my typing this paper) are true while being totally irrelevant to anything.

One point that might be less immediately obvious is that being able to identify causes is not necessary for understanding either. There must be non-causal understanding (and explanation), because there is understanding (and explanation) in mathematics and philosophy, where causes are few and far between (as necessary truths cannot be caused). There are also worries about mereological understanding, logical understanding, etc. It is thus fair to wonder about the genus of which causal understanding is a species—it is this genus UC aims to provide.

Still, one might hold out hope that understanding in math and philosophy is one thing, but that understanding events always involves knowing something about their causes. This does not seem quite right either. Van Fraassen (1980) famously gives an example wherein he claims an effect (that a shadow appear at a certain point) explains its cause (that a tower was built to a certain height). This example might or might not be persuasive with respect to explanation,[29] but we can make the case more compellingly for understanding. There are some contexts where at least the main component of understanding the assassination of Archduke Ferdinand is grasping its role in starting WWI.[30] An account that says I understand an event exactly when I know its relevant causes cannot make sense of this intuition.[31] By contrast, knowing that it caused WWI is a good way to be able to figure out all sorts of other information about the event, such as that it was done by foreigners and that he was a relevant enough person to make the event an international incident.[32] These conclusions could of course prove incorrect, but that is as it should be—we might think we have understanding when we don't.

---

[29] It probably is not a telling counterexample against causal explanation, as one can readily re-describe the case as one wherein an intent to have a certain effect explains a potential cause. This reestablishes the normal causal order.

[30] I first used this example in my dissertation *Explaining and Understanding*.

[31] Lange's (2016) description of causal explanation might count this explanation as causal, since it still tells us something about the causal structure of the world. However, as his broader point is the existence of non-causal explanations, his view is no threat to the current proposal.

[32] Here's a mini-example of my view in action: before writing that sentence about Archduke Ferdinand, I couldn't actually remember who he was or by whom he was killed. However, knowing the assassination's role, I was pretty

As one final fallback, the causal theorist might grant that while understanding an event (*simpliciter*) can be had with non-causal information, understanding *why* something happened requires information about its causal history. There are still ways to push back against this claim, as research has recently proliferated in explorations of non-causal explanations in science. For example, Lange (2016) argues that statistics really explain in a non-causal way why genetic drift occurs. (One advantage of UC is that it might account for much of the variance in kinds of non-causal explanation, as when Lange (2016) admits that he "will not even try to portray all non-causal scientific explanations as working in the same way as one another." (p. xii))[33]

Still, many argue that putative non-causal explanations of why events occur are really causal—see for example Reisman & Forber (2005) regarding genetic drift and Skow (2014) regarding a host of other examples. However, at this point I can also grant that all explanations of why events occur require reference to that event's causes. If that is true, then the thing to say is that if we build enough into a question, then we consequently place constraints on what constitutes an acceptable answer. If I ask "why" something happened, and by that I mean to ask what caused it, then of course only causal information will do. If I ask "schwhy" something happened, and by that I mean to ask how it relates to apple carts, then only apple-cart information will do. In neither case do we learn anything particularly interesting about the nature of understanding, but only that in some contexts we might be more interested in having a particular instantiation of understanding than others.

To take one example, Bradford Skow (2014) argues that all explanations of singular events are, appearances notwithstanding, really causal explanations. If the view is meant only to apply to explaining *why* events occur,[34] then I am happy to grant that such restricted explanations (and any understanding based on them) require reference to causes. If we tried to force his view to be a complete theory of understanding events *simpliciter*, then there would be two defenses available to me. First, if we were to take him as saying that all events were to be understood wholly in terms of their causes (as opposed to explained wholly or partially, which is I believe the intended reading), the view would have trouble making sense of understanding an assassination in terms of a subsequent war. Second (and more importantly), it would still not challenge UC as a more general theory of understanding objects other than events. As such, Skow's arguments do not require that I add the presence of causal information as a necessary condition of possessing understanding. Even if possession of some causal information were generally sufficient for some minimal degree of understanding, that is compatible with a more compressed representation/process pair always providing more understanding.

---

confident that the assassins were foreigners and that he was important. Sure enough, he was the heir apparent to the Austro-Hungarian Empire and was killed by a Bosnian Serb.

[33] Whether UC can successfully make sense of all of Lange's example I leave for future work.

[34] Skow's later work suggests that this is the correct reading. Here is Skow (2017, p. 905): "My theory starts with the idea, which has been held by many others, that explanations are answers to why questions."

§4.2 Understanding as Grasping Dependency Relations

Grimm (2006, 2010) argues that understanding is a species of knowledge. Specifically, understanding consists in grasping a good explanation, where grasping involves being able to see how changes in one variable manifest as changes in another variable.[35] Typically, this will be a good way to encode some information, as one will be able to remember general connections between variables along with lists of deviations rather than having to remember an indefinite number of pairwise (or more) relations.

However, it is not clear that Grimm is correct that knowledge is either necessary or sufficient (even when combined with grasping) for understanding.

Regarding whether knowledge is necessary, we seem to be able to understand things where our epistemic grasp of the relevant propositions does not rise to the level of knowledge. Hills (2016) argues that one can understand even when one's justification (and so, presumably, one's knowledge) has been undermined by the presence of defeaters. In her example, she envisions a student who learns from a book everything she needs to know about Napoleon's greatness as a general, and comes to understand why Napoleon was a great general. However, she is then informed by her usually-reliable history teacher that Napoleon is not in fact a great general. If she goes on to ignore her teacher, then she still understands why Napoleon was a great general, even though she no longer really knows why.[36] UC remedies this by requiring not knowledge, but merely a representation (which might or might not comprise knowledge) and the ability to recreate information (that might or might not comprise knowledge). Some (following Kvanvig 2003) have also questioned whether understanding requires the same anti-luck conditions as knowledge—I will return to this question in §4.6.

Whether Grimm's criteria are sufficient depends on how one cashes out what it really means to "grasp" an explanation, but it has been argued (de Regt 2009) that the sorts of abilities Grimm has in mind are not sufficient to account for the actual scientific practice of understanders or understanding-attributors. De Regt requires one actually be able to use a theory to make predictions about phenomena which, as noted, falls directly out of UC as well.

§4.3 Unificationism

On unificationist views of understanding (e.g., Friedman 1974, Kitcher 1989), one understands when and only when one unifies larger stores of information while utilizing fewer cognitive resources—in other words, one might be concerned that UC is just unificationism warmed over.

I grant that the central insight of UC is of a piece with unificationism—one understands by getting more cognitive bang for less cognitive buck. However, this insight has never been what has been at issue with unificationism—where unificationism tends to get bogged down is in

[35] Grimm's current view is a bit more nuanced, but the earlier work represents one fairly typical line of thought.
[36] Khalifa (2017) has a nice account of why the student in such a case has minimal but somewhat misguided understanding.

the details, and those are substantially different from UC. First, unificationism has it that all understanding is made up of arguments or argument patterns, which runs into trouble when one thinks about (among other things) singular causal relations (Scriven 1963). By taking representations/processes as the basic components of understanding, this problem is avoided—one can represent parts of a causal story in ways that let one conclude more information about it without those representations being arguments.

Moreover, one of the more damning objections to unificationism[37] is precisely that it fails to take into account the extent to which people can actually *make use* of representations. Humphreys (1993) argues that systematizations of logic involving the Sheffer-stroke are more unified than systems with more connectives and axioms, but are much less conducive to understanding. UC not only gets this result, but actually explains it—the Sheffer-stroke system is, for most of us, merely a matter of brute memorization rather than compression.[38]

§4.4 Strevens' Kairetic Account

In this and §4.5 I examine the accounts that bear the most similarity to UC, and put forward those considerations that favor the latter.

I begin with Strevens' kairetic account of explanation. Though Strevens' official account of understanding (2013) is close to Grimm's, elsewhere (2004) his work on explanation provides an important precursor to any insight found in UC. Strevens argues (very roughly) that a good explanation is one which lays out the causal history of the explanandum but "strips away" details of that story that were not "difference-makers" with respect to that explanandum's occurrence. Strevens has a specific account of how one should go about determining what the difference makers are,[39] but that does not matter for present purposes—the key is that he sees that explanation is fundamentally about getting rid of extraneous information. UC captures this intuition as well—if a given causal factor did not make a difference to what we're trying to understand, then inclusion of that causal factor would lead to an increase in the amount of space our understanding would take up, without a corresponding increase in the amount of useful information we could recreate; since, *ex hypothesi*, the factor did not make a difference, anything we could conclude on the basis of awareness of that factor isn't useful. Thus, a Strevens-style explanation would typically engender UC-style understanding. Conversely, if one has enough information to determine whether an event occurred (the restriction of understanding events in

[37] Kareem Khalifa made the point that there are a surprisingly large number of problems with unificationism, and that I should ideally go through all of them to show how UC does better. For now I make the admittedly unproven assertion that none of the other problems ramify to UC, and that the burden is on the objector to show that they do.

[38] It has been suggested that the problem with the Sheffer-stroke system is that it does not piggy-back on standard usage in the way other connectives might. Instead of the Scheffer-stroke we can compare natural deduction to Frege's axiomatic system. This system used only common terms (negation and if-then), but the introduction of natural deduction systems still seems like a step forward for understanding.

[39] You abstract away to the maximum level of generality that still entails the explanandum and tells a relatively cohesive causal story.

terms of whether or not they occurred is from Strevens), then one must not have omitted any difference makers.[40] Furthermore, since it is impossible to encode an entire causal history in a finite mind, some information will have to be omitted somewhere. Therefore, having UC understanding will typically involve having all the resources necessary for a Strevens-style explanation. Thus, for the most part UC understanding reduces (in scientists' sense of "reduces", not philosophers') to Strevens-style explanation when dealing with understanding of why events occurred.

UC goes beyond Strevens' account in several important ways.[41] First, causal stories are just one among many packages of information one might want to encode. They might well be the sort most common when understanding scientific phenomena, but they are not omnipresent even there. For example, even Skow, who thinks that all explanations of singular events convey causal information, is not committed to causal information being the *only* route to understanding events. Second, there are some things we might want to understand that do not fit into the causal nexus at all. This is important when it comes time to understand a mathematical proof or a philosophical argument.[42]

UC also gives a somewhat more streamlined answer to the question of how we can avoid massively disjoint explanations. Strevens mandates that an account be maximally cohesive, where an account fails to be cohesive when there are different causal paths to a more disjoint conclusion. For technical reasons, without this constraint his model might favor an explanation of Rasputin's death (by drowning) in terms of his either having been thrown into a river *or* given poison teacakes (over the more intuitively satisfactory explanation solely in terms of his having been thrown into a river). This worry never really gets off the ground for UC though, because additional representations that would tell us how Rasputin might have been poisoned do not give us information about the event of his actual death by drowning. (They might give us information about the broader object "how Rasputin might have died", but as Strevens' account does not even apply to such counterfactually constructed objects, no direct comparison can be made.)

It is worth acknowledging that UC does make the pragmatics of what sort of information is relevant bear somewhat more weight than Strevens' more formal(/objective?) apparatus, which might be viewed as a cost. There are two things that mitigate this concern. First, some of the work that for UC is done by the pragmatics is for Strevens done by his specification of the explanatory question—what makes a difference when asking the question of "why Rasputin died" is different from what makes a difference when asking "why Rasputin died in the way that he did." There is thus still a parameter in a karaetic account for what constitutes an appropriate

[40] One might worry that one could have simply gotten lucky by omitting difference makers that cancel out. On Strevens account these wouldn't count as real difference-makers anyway: the facts that Rasputin was poisoned but survived the poison are not two facts that cancel out but which made a difference to his ultimate death (by drowning), but rather on Strevens' account not difference-makers at all.

[41] This is not meant as a criticism—UC might well reduce to his account in the specific case of explaining why particular events occurred, which was quite explicitly (p. 155) his target.

[42] One might think that there are certain non-causal laws that could be explanatory as well (such as the maximum velocity of objects moving in our universe), but Skow (2014) persuasively argues that these are really causal as well.

bit of information in a given context. Second, it is not totally clear that having a broader role for pragmatics is a bad thing. Intuitively, what counts as understanding the very same phenomenon can vary from context to context—to slightly adapt an example from Wilkenfeld (2013), whether one counts as "understanding why they became friends" (and, perhaps, what counts as "explaining why they became friends") might depend a great deal on whether we are looking for someone to give a talk at a psychology conference or someone to help us navigate a party. If the phenomenon of interest is very pragmatic-laden, then the model for it should be as well.

§4.5 URM

There are also resonances between this account and the Understanding as Representation Manipulability (URM) account presented in Wilkenfeld (2013). That account requires that understanding comprise representations that one can manipulate in useful ways. The accounts share the ideal that understanding is a fundamentally representational capacity, and that it is best cashed out in terms of how one thinks about the understood and what one can do with those thoughts.

Nevertheless, there are several advantages to the present theory. First, URM claims that one's understanding of (for example) the soundness proof consists in the fact that one could correct small mistakes and that one could potentially prove soundness for other logical systems. For reasons discussed in §2, these abilities will quite frequently go along with UC understanding as well, but UC predicts that they are neither necessary nor sufficient. They are not sufficient, because one could have an algorithm for correcting a proof that is even more complicated (in an information-theoretic sense) than the original proof. For example, I could have memorized that if I made a mistake on the proof, I could reproduce large tracts of it by rote—it's not clear what URM would say about such a case, but I do not see how it could rule out the counterintuitive result that it counts as understanding. More to the point, URM-style understanding is not necessary for UC-style understanding. I could have compressed information about this proof, even if as a matter of fact I would (for whatever reason) never make a mistake with it (I'm remarkably meticulous) and would never want nor be able to adapt it to other logics (I believe that classical first-order logic is the one true logic). URM would have it that I would not understand in such a case, whereas UC would have it that I do—again the advantage seems to be with the latter.

A second, speculative advantage of UC is that, if the project shows promise, it has far more potential to be quantified. In principle, for a given context we could compare different claims of understanding based on the degree of compression—URM does not yield such tidy results.

§4.6 Understanding in Epistemology

Most of the accounts heretofore discussed (with the exceptions of Grimm 2006 and Wilkenfeld 2013) have been based in philosophy of science. However, as noted in the introduction, my aim is a bit more ambitious. Obviously it is impossible to connect UC to every

question in the epistemology of understanding, but it might be helpful to look at two central debates.

Frequently, epistemologists (e.g., Elgin 2007, Kvanvig 2003) focus on the requirement of coherence among the propositions constituting an explanation. While this is not required on UC, in most cases grasping coherence relations improves UC understanding. What coherence amounts to is the existence of a governing principle for an entire body of data, which can then typically be represented explicitly along with processes for reconstructing other relevant information. Conversely, UC allows for understanding in the absence of an appreciation of coherence relations, as the ability to produce logical proofs from some core representation seems independent of being able to step back and take a holistic view of how the propositions that constitute proofs relate. (Or, if one thinks that this ability does amount to appreciation of coherence relations, then nothing is lost by saying that philosophers such as Elgin and Kvanvig have identified a crucial property of understanding generally.)

Another topic of interest among epistemologists, as already noted, is whether a state counts as one of understanding if its accuracy came about as a result of luck. Kvanvig (2003) argues that understanding is more compatible with luck than knowledge is, Grimm (2006) argues that understanding and knowledge go together, and Pritchard (2009) argues that, with respect to luck, understanding and knowledge sometimes go together and sometimes not. Since it lacks any etiological requirement, UC has it that understanding is compatible with any kind of epistemic luck.[43] If one wanted, one could bring UC in line with Grimm's claims by explicitly adding an etiology requirement, but I think that would be a mistake.[44] One of the motivations for the anti-luck requirement is that it is thought that the value of understanding hinges on its being a particularly impressive epistemic state. By contrast, UC can account for understanding's value on much more mundane grounds, which are its instrumental role in data compression and modeling.

§5 Conclusion

In this paper I have endeavored to set out a new outline for an account of understanding. I have argued that this account captures some useful platitudes about understanding, is extensionally plausible, and fares well when compared to several other candidate views. Obviously much more work needs to be done—cases need to be examined and details need to be worked out. This paper is only intent on setting an agenda, and showing that it is an agenda that is worth pursuing in its infancy.

[43] The account is silent on knowledge, and so does not say whether or not they go together.

[44] Wilkenfeld, Plunkett, and Lombrozo (2017) admittedly provide some evidence for Grimm's view that you cannot pry apart the folk conceptions of **knowledge** and **understanding** with an etiology requirement. I would be inclined to read that as a sign that philosophers have been too demanding with respect to knowledge rather than too loose with understanding, which interpretation is consistent with much of the data.

Acknowledgements:

I would like to thank the University of Pittsburgh, the Center for Philosophy of Science (and all its Autumn 2017 fellows), and in particular Edouard Machery, Colin Allen, Stewart Shapiro, Pablo Acuña, and Kareem Khalifa for invaluable written feedback. I also thank Tania Lombrozo for helpful conversation.

References

Craig, E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford University Press.

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

De Regt, H. W., & Gijsbers, V. (2016). How false theories can yield genuine understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science. London: Routledge*.

De Regt, H. W., Leonelli, S., & Eigner, K. (2009). *Scientific understanding: Philosophical perspectives*. University of Pittsburgh Pre. Retrieved from

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

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, *45*(1), 1–35.

Friedman, M. (1974). Explanation and Scientific Understanding. *Journal of Philosophy*, *71*(1), 5–19.

Grimm, S. R. (2006). Is understanding a species of knowledge? *The British Journal for the Philosophy of Science*, *57*(3), 515–535.

Grimm, S. R. (2010). The Goal of Explanation. *Studies in the History and Philosophy of Science*, *41*(4), 337–344.

Grünwald, P. (2005). A tutorial introduction to the minimum description length principle. Retrieved from http://sites.google.com/site/mkrishnarhul/IntroMDL.pdf

Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, *15*(2), 135–175.

Hills, A. (2016). Understanding why. *Noûs*, *50*(4), 661–688.

Humphreys, P. (1993). Greater Unification Equals Greater Understanding? *Analysis*, *53*(3), 183–188.

Kelp, C. (2015). Understanding Phenomena. *Synthese*, *192*(12), 3799–3816.

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

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. Salmon (Eds.), *Scientific Explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.

Kripke, S. A. (1982). *Wittgenstein on rules and private language: An elementary exposition*. Harvard University Press.

Kvanvig, J. L. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge University Press. Retrieved from

Lange, M. (2016). *Because without cause: Non-causal explanations in science and mathematics*. Oxford University Press.

Le Bihan, S. (2016). Enlightening falsehoods: a modal view of scientific understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New essays in epistemology and the philosophy of science.* Routledge.

Lewis, D. (1980). Mad pain and Martian pain. *Readings in the Philosophy of Psychology*, *1*, 216–222.

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.

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

Pritchard, D. (2009). Knowledge, understanding and epistemic value. *Royal Institute of Philosophy Supplement*, *64*, 19–43.

Reisman, K., & Forber, P. (2005). Manipulation and the Causes of Evolution. *Philosophy of Science*, *72*(5), 1113–1123.

Scriven, M. (1962). Explanations, predictions, and laws. In *Minnesota Studies in the Philosophy of Science* (Vol. 3, pp. 170–229).

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Skow, Bradford. (2014). Are There Non-Causal Explanations? *British Journal for the Philosophy of Science*, *65*, 445–467.

Skow, Bradford. (2017). Levels of Reasons and Causal Explanation. *Philosophy of Science*, *84*(5), 905–915.

Strevens, M. (2004). The causal and unification approaches to explanation unified—causally. *Noûs*, *38*(1), 154–176.

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

Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, *75*(2), 76–92.

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

Waskan, J. A. (2006). *Models and Cognition*. A Bradford Book.

Wilkenfeld, D. A. (2013). Understanding as representation manipulability. *Synthese*, *190*(6), 997–1016.

Wilkenfeld, D. A. (2014). Functional Explaining: A New Approach to the Philosophy of Explanation. *Synthese*, *191*(14), 3367–3391.

Wilkenfeld, D. A. (2017). MUDdy Understanding. *Synthese*, *194*(4), 1273–1293.

Wilkenfeld, D. A., Plunkett, D., & Lombrozo, T. (2018). Folk Attributions of Understanding: Is There a Role for Epistemic Luck? *Episteme*, *15*(1), 24–49.

Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review*, *115*(1), 1–50.

Zagzebski, L. (2001). Recovering understanding. *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, 235–252.