2016

Understanding and Its Role in Inquiry

Benjamin T. Rancourt
University of Massachusetts Amherst

UNDERSTANDING AND ITS ROLE IN INQUIRY

A Dissertation Presented

by

BENJAMIN T. RANCOURT

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2016

Department of Philosophy

UNDERSTANDING AND ITS ROLE IN INQUIRY

A Dissertation Presented

by

BENJAMIN T. RANCOURT

Approved as to style and content by:

_____

Hilary Kornblith, Chair

_____

Alejandro Pérez Carballo, Member

_____

Christopher Meacham, Member

_____

Andrew Cohen, Outside Member

_____

Joseph Levine, Department Chair

Department of Philosophy

ACKNOWLEDGMENTS

This dissertation would not exist were it not for the guidance of Hilary Kornblith and Christopher Meacham. Their advice, discussion and comments helped shape this project from its beginning through to its completion. I would also like to express my gratitude to members of my dissertation committee, Alejandro Pérez Carballo and Andrew Cohen, for discussion and comments that made this dissertation better.

I would like to thank the members of the University of Massachusetts Amherst Philosophy dissertation seminars from Fall 2012 through Fall 2014. The comments I received from the participants in these seminars were constructive and valuable. Special thanks are due to the professors who ran these seminars: Louise Antony, Maya Eddon, and Joseph Levine.

I could not have done this without the support of my wife, Julia Rancourt. I am grateful for the support of my parents, Stephanie and Kenneth Rancourt, and all of my family. I am grateful for all of my friends who helped me stay the course to reach this point.

ABSTRACT

UNDERSTANDING AND ITS ROLE IN INQUIRY

MAY 2016

BENJAMIN T. RANCOURT, B.A., UNIVERSITY OF MINNESOTA

M.A., UNIVERSITY OF WISCONSIN - MILWAUKEE

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hilary Kornblith

In this dissertation, I argue that understanding possesses unique epistemic value. I propose and defend a novel account of understanding that I call the *management account of understanding*, which is the view that an agent A understands a subject matter S just in case A has the ability to extract the relevant information and exploit it with the relevant cognitive capacities to answer questions in S. Since inquiry is the process of raising and answering questions, I argue that without understanding, it would be impossible to engage in successful inquiry. I argue that understanding is indispensable for effective cognition and that it is irreducible to other epistemic categories, such as knowledge, justification, or rationality. Understanding is an irreducible component of epistemic excellence.

In arguing for this account, I focus on the nature and requirements of cognition and pay less attention to intuitions about hypothetical cases, in contrast to previous approaches to understanding. I draw upon psychology and cognitive science, especially work on the frame problem, to demonstrate the importance (and difficulty) of finding relevant information and exploiting it in relevant ways to answer questions. One's cognitive system must be organized in such a way that one has the ability to think the relevant thoughts in the relevant ways to answer the relevant questions. This reveals well-defined directions in which future

research on understanding should move.

In addition to proposing and defending the management account of understanding, I explore several features of understanding, including some interesting consequences of the management account. I argue that the management account and commonly used methods of measuring degrees of understanding imply that understanding can be based on false beliefs. I argue that the management account (and any account of understanding that implies understanding is an ability) implies that understanding can be acquired by luck. I also argue that understanding is ontologically prior to explanation: a good explanation is essentially an act that produces understanding in the relevant cognitive background.

TABLE OF CONTENTS

Page

CHAPTER 1

BACKGROUND

1.1 Introduction

Understanding is generally considered to be a significant intellectual achievement. Providing understanding is considered to be the goal of education. Gaining understanding is considered to be the goal of science. Yet until recently, understanding was largely considered uninteresting as a topic of philosophical study. It received only minimal attention in epistemology. In the past decade or so, however, this has changed and there is now a growing philosophical literature about understanding.

One effect of the neglect of understanding is that the epistemological issues relating to understanding are not as clearly delineated as the issues relating to other topics such as knowledge. We are still working out the logical space of possible views regarding understanding, and the different ways of categorizing those views. We are still coming to terms with the different states that are referred to as 'understanding,' and we are still working out which are relevant to epistemology. This makes the use of intuitions methodologically problematic, for without clearly delineated distinctions among uses of 'understanding,' it will not be obvious whether a wayward intuition is a problem for a theory or whether it merely relates to a different usage of the term. In this dissertation, I will clarify these issues, and argue for solutions to many of them.

In this chapter, I survey previous philosophical work on the topic of understanding. I highlight major themes, developments, and challenges of this work. After highlighting

these features of previous work, I explain how my approach differs.

My approach aims first and foremost to identify the role that understanding plays in our cognitive lives. I aim to explain what role it plays in allowing a cognitive system to engage in successful inquiry. I identify an independent problem that cognitive systems must solve in order to function as cognitive systems at all, and argue that understanding is the ability to solve this problem. This opens new avenues in the study of understanding by tying understanding into several important areas of research and interest.

The chapter is organized as follows. In section 1, I briefly survey some issues with the term 'understanding.' In section 2, I survey previous approaches to understanding and highlight some of the challenges facing them. In section 3 I explain my approach to the study of understanding.

1.2 The term 'understanding'

Natural language allows for several different uses of the term 'understanding,' and not all of these are relevant to the project of this dissertation.

First of all, there is understanding a language and understanding units of language, such as sentences or words. While I believe that the account of understanding I present and defend in later chapters can be extended to give an account of understanding language, I will not address this topic in the present work. "I understand French," and "She did not understand what I meant" will find no account in the present work.

There are some weak uses of 'understand' as well. For example, it can mean something akin to *believe based on rumor*, as in 'I understand you'll be headed to Palo Alto' upon hearing that someone got a job offer at Stanford. This weakened use of 'understand' is not of interest here. Whatever state it refers to must be left unanalyzed.

The forms of understanding that I will talk about include: understanding a subject matter, understanding a theory, understanding why something occurs or is true, understanding an object or work of art or person, and understanding how something works. These forms

of understanding are epistemically significant, as I will show in later chapters, and I will provide accounts of all of them.

1.3 Previous approaches

My approach departs to some extent from approaches to understanding that have been taken in the past. To clarify the difference, I will briefly survey previous work done on the philosophy of understanding. We will start in the distant past, when understanding was considered more central, through a time when understanding was neglected in philosophy, and up to its more recent resurgence.

1.3.1 Ancient philosophy

It is generally thought that the ancient Greek philosophers were more concerned with understanding than with knowledge. Modern concerns about skepticism and the very possibility of knowledge were not as central to ancient philosophy. Less attention was paid to showing that knowledge was possible, and more attention was paid to depth of understanding. There is good reason to think that some ancient works that had been considered to be about knowledge are better thought of as about understanding. (Moravcsik 1979) argues that the use of *epistêmê* in many of Plato's dialogues is better translated as 'understanding' rather than 'knowledge.' (Burnyeat 1984) argued that Aristotle's use of the term *epistêmê* is better translated as 'understanding' than the more common translation 'knowledge,' since Aristotle often implied that *epistêmê* involved greater depth than mere knowledge implies, such as the ability to produce explanations.[1]

To the ancients, it seems, the aim of epistemology was to determine how one achieves understanding of the world. Concerns about knowledge and justification, while still present (prominently in Plato's *Theaetetus*, for example), were considered less pressing than concerns about understanding. However, since that time, the focus has shifted from under-

[1] (Lesher 2001) argues that Burnyeat is correct about many instances of *epistêmê* in Aristotle's work, but not all of them. Some can still be translated as 'knowledge.'

standing to knowledge. The knowledge-centric approach to epistemology has reigned from about the time of Descartes to the present.

1.3.2 Subsidiary to explanation

Moving forward to the twentieth century, philosophical research into understanding was mostly carried out as a subsidiary topic in the explanation literature. There was little sustained attention focused on understanding in its own right. The approach to understanding, such as it was, operated under the assumption that the real work was finding an adequate theory of scientific explanation and that understanding was just a matter of having a good explanation. Philosophers did not pay much attention to what it meant to *have an explanation* or how having an explanation would provide understanding, it was largely just assumed that explanations provide understanding, as in (Hempel 1965), (Friedman 1974), (Kitcher 1981), (Kitcher 1989), (Strevens 1999).

This approach is inadequate. First of all, the exact connection between understanding and explanation tended to be left vague. No one provided a convincing argument that explanations provided understanding, nor did anyone offer a clear description of exactly how explanations provided understanding.[2] If understanding is just a by-product of being exposed to a good explanation, no convincing argument had been given.

In general, the earliest analytic philosophy covering understanding was dismissive, neglecting to give detailed accounts of what understanding was. Generally, the view was that whatever interest understanding had, it derived from its connection to more important topics such as knowledge and explanation. It was therefore not worthy of discussion in its own right; one only needed to give an account of explanation, for example, and then drop one sentence saying that explanations meeting the proposed account provided understanding. The attitude has been defended more recently by (Trout 2002) who claims that proponents of the importance of understanding face a dilemma: either understanding is (i) knowing

[2]This complaint is raised in Section 2 of (Barnes 1992) and in (de Regt and Dieks 2005). I criticize previous explanation-based approaches to understanding in more detail in Chapter 6.

an explanation, in which case understanding is trivially reducible to knowledge and explanation and not philosophically interesting for in its own right, or (ii) understanding is a subjective psychological phenomenon, in which case it lacks epistemic value.

Lipton further describes horn (i) of the dilemma by posing the question: if gaining knowledge helps us to operate successfully in the world, and if knowledge or truth is the primary epistemic goal, then what do we gain by also possessing understanding (Lipton 2004, p. 129)?

Hempel discusses the horn (ii) of the dilemma. He supposed that understanding was a feeling of finding something intelligible.[3] Finding something intelligible might be of interest to a study of subjective experience, but it it hard to see how it plays any significant epistemological role (Hempel 1965, p. 426). (Trout 2002) takes this further, claiming that the feeling of understanding systematically misleads us into thinking that we have found a good explanation when we have not. In addition, the desire to gain understanding leads us to adopt idealizations without verifying that they will work. If an idealization makes things intelligible, it becomes a temptation to adopt bad idealizations (Cartwright 1995).

Among those who accepted that understanding faced the above dilemma, two views regarding understanding developed, depending on which horn was taken. Some argued that understanding is a species of knowledge, and others argued that it was a subjective feeling.

1.3.3 Understanding is a species of knowledge

Consider the view that understanding is a species of knowledge. The most common versions of this proposal being that understanding why p is identical to knowing a true explanation of why p, or knowing the causal history of why p (these views are espoused by (Trout 2002), (Barnes 1992), (Salmon 1998)). This was assumed without much argument

[3] More specifically, he claims that understanding means seeing that an event was to be expected by deriving its occurrence from laws and initial conditions. We can generalize his point to other feelings of intelligibility, however.

for much of the twentieth century.

This proposal has met with resistance. For example, (Kvanvig 2003, ch. 8) presents a line of argument against the proposal that understanding is any species of knowledge. Kvanvig claims that the only external component of understanding is truth, and the other components of understanding derive from the internal state of grasping coherent connections among a body of beliefs. Since knowledge includes an external component to avoid epistemic luck and Gettier problems, understanding does not imply knowledge and so is not a species of it (p. 197). The example he uses to show this is the Comanche Case: someone reads a completely accurate book about the Comanches, forming true beliefs about them and grasping connections among those beliefs. However, it turns out that it is only by luck that the information in the book is true; the author simply guessed. Kvanvig claims that the person understands why the Comanches were dominant at their height, but does not know the relevant facts, demonstrating that understanding is not knowledge.

In response to Kvanvig (and other arguments I will not recite in detail), (Grimm 2006) argues that it is not so easy to prove that understanding is not a species of knowledge. First of all, he argues against Kvanvig that the etiology of beliefs affects understanding. A hallucination that happens to produce true beliefs does not give understanding (p. 520). Understanding rules out luck. Another example he uses to show this is as follows: you see an event where it looks like a hammer smashes nut. In fact, you did see a hammer smash a nut, but every other day the same person does something that looks exactly the same but is only a simulation of nut-smashing. Even though you have true beliefs, the luck means you do not understand why the object was smashed. Thus, Kvanvig's claim that understanding lacks externalist anti-luck conditions fails.

Another objection against the claim that understanding is a species of knowledge is that knowledge is a kind of belief, and belief is far too "thin" to encompass understanding. Beliefs are mere assent, and understanding requires much more. Since knowledge is a species of belief, the knowledge thesis would make understanding a species of belief. Grimm's

response is that belief is not always as thin as mere assent. As an example: to know a logical truth one must grasp enough of the logic behind the proposition to grasp why it is true. Thus, some beliefs may be robust enough to be understanding in some cases.

Though Grimm responds to some problems for the view that understanding is a species of knowledge, several other objections have emerged. (Zagzebski 2001) argues that understanding is the grasping of non-propositional aspects of reality, so it could never be propositional knowledge (p. 244). Both Grimm himself (in (Grimm 2010)) and (de Regt and Dieks 2005), among others, argue that understanding requires the skill to apply knowledge in context. This skill is required in addition to knowledge, and is not reducible to it.

The best defense of the view that understanding is a species of knowledge is now found in (Khalifa 2012). Khalifa claims that understanding why p is knowing an explanation of why p. One issue with this account is that it it explicitly only covers understanding-why; it is not clear how to extend it to cover understanding a person or a subject matter. Beyond this initial difficulty, Chapter 2 and 3 provide detailed arguments that understanding is not reducible to any species of knowledge.

1.3.4 Subjective feeling

The second early theory of understanding holds that understanding is a subjective feeling of satisfaction or intelligibility associated with an explanation or belief. (Trout 2002) uses Peirce's description of understanding as occurring when an explanation is "turned back and forth [in the mind] like a key in a lock." (Hempel 1965, pp. 430–2) identifies understanding with the feeling that an event was "to be expected", or in other words, that it is not surprising. Other descriptions call it the "Eureka!" moment or the "Aha!" feeling.

If understanding is in fact an "Aha!" feeling, then it is likely that understanding plays no epistemological role. As (Hempel 1965) points out, it might be of interest to some areas of research to know what explanations feel satisfying to us, but epistemology is concerned

with objective features that lead to scientific progress. Subjective feelings of satisfaction are not of interest to epistemology.

(Trout 2002) argues that in fact the situation is worse than this; understanding is a feeling that leads us astray. Our feeling of understanding is primarily due to biases, such as hindsight bias and overconfidence bias. These are two well-known ways in which our feeling that something makes sense can lead us astray. They lead us to feel like we have found a good explanation when in fact we have not.

However, even if Trout is correct that understanding is just a subjective feeling, it does not necessarily follow that understanding is of no epistemic merit. (Lipton 2009), based in part on arguments from (Gopnik 1998), argues that our sense of understanding is a reliable, though fallible, guide to inference. Lipton draws upon his claim in (Lipton 2004) that inference to the best explanation is one of our most important forms of reasoning, and in (Lipton 2009) he adds that our sense of understanding is what alerts us to the goodness of an explanation. Thus, assuming understanding is just the sense that an explanation makes sense, understanding is epistemically valuable. He appeals to Gopnik's arguments that the sense of explanation is an adaptation that gives a sense of satisfaction when we find true explanations, and is thus attuned to actually finding true explanations. (Grimm 2010) also says that the *Aha!* feeling comes from genuine understanding, so it is not just an epistemically meaningless feeling.

However, there is good reason to reject the claim that understanding is just a subjective feeling. Intuitively, one can apply one's understanding in order to solve problems. This is nonsensical if understanding is a feeling. When one applies one's understanding, one is exercising a skill, not a feeling. In the rest of my dissertation, I present and defend a view of understanding that implies it is not a feeling and that it is epistemically significant.

1.3.5 Resurgence of understanding

Recently, philosophers have taken understanding more seriously as a worthy topic of study in its own right. Most of this new work on the topic follows the traditional analytic philosophical method of consulting intuitions about cases of understanding and lack of understanding. Judgments about cases are catalogued and compared to proposed theories to determine whether any theory is adequate. Much of this work tends toward general claims, rather than delving into details and specifics.

Most philosophers currently working on understanding tend to reject the claim that understanding is reducible to knowledge. They tend to reject the claim that understanding is subsidiary to explanation. They tend to reject the claim that understanding is a subjective feeling.

Understanding has been hailed by some as the natural object of epistemological concern for epistemologists who are not primarily trying to defeat skepticism. As the discussion shifts away from worries about certainty and justification in the face of skepticism, understanding comes to the fore (see (Kvanvig 2003, pp. 186–7) and (Zagzebski 2001, p. 236)). The thought is that once we move past obsessive concern with skepticism, our concern is not whether and how we have knowledge, but how to apply knowledge and integrate it into a coherent whole. Understanding, the suggestion goes, is a robust state that includes the skills to use knowledge and information. It implies possession of the skills to interact with the world in more comprehensive ways than mere knowledge allows. It is more significant to post-skepticism epistemology than knowledge. (Even if the conclusion of the supremacy of understanding is overstated, the point about understanding's importance is significant.)

Recent accounts of understanding identify understanding with some kind of "grasping". Others identify understanding with some kind of skill, in particular the skill to apply knowledge in the right way. Still others employ both notions, grasping and skill. Taken all together, it is clear that "grasping" amounts to some kind of skill. The general outlook is expressed by Elgin, "Understanding is a grasp of a comprehensive body of information that

is grounded in fact, is duly responsive to evidence, and enables non-trivial inference, argument, and perhaps action regarding that subject the information pertains to." (Elgin 2007, p. 36) So grasping, in addition to its connection to evidence, is associated with cognitive and practical skills of inference and argument formation.

Defenses of "grasping" theories of understanding have been developed in a few ways. First, there is the value-driven approach. Proponents of this approach have grown skeptical of the claim that knowledge is uniquely epistemically valuable and worthy of being the defining concept of epistemology. In part this is in frustration at the proliferation of epicycles required to avoid Gettier problems. The concern is that anti-Gettier conditions appear to have little epistemic value and are only needed to match intuitions, as argued in (Kvanvig 2003) and (Zagzebski 2003). This leads to concern that knowledge does not have a distinctive value above the value of true belief. As a result, some have turned to understanding in the hopes that it has the distinctive value knowledge lacks, for example in (Kvanvig 2003) and (Boylu 2010). They claim that the value one might have hoped to find in knowledge is only available through understanding.

One proposal along these lines is that understanding, and the "grasping" involved, is the skill to apply knowledge, as argued in (Grimm 2010). One does not understand why a plane flies, for example, simply by knowing Bernoulli's principle and the boundary conditions of the plane. One only understands if one can apply the principle to the particular case of the plane. Likewise, understanding geometry is not a matter of knowing a list of geometric theorems, it is a matter of being able to apply the knowledge of geometry to new problems. One has understanding when one knows that the principles are true and is able to apply them. The sense of application that is being used is more than the ability to perform brute force calculations, which is the ability to use a formula without understanding anything about how it works. What is the right kind of application?

Understanding may be the ability to apply knowledge by allowing one to determine what would occur in certain counterfactual situations. Specifically, one understands when

one has the skill to answer questions about what would happen if causally relevant parameters were manipulated. An argument in favor of this view is that understanding requires the ability to produce explanations, and an explanation of an event cites the causal dependencies that lead to the event. There are asymmetries in explanation, for example one can explain the length of a flagpole's shadow by the height of the flagpole, but one cannot explain the height of the flagpole by the length of the shadow, despite structurally identical arguments in either direction. The proposal here addresses this problem by noting that one can manipulate the length of a shadow by manipulating the height of a flagpole; there are well-defined answers to questions about what would happen to the shadow given different heights. However, one cannot manipulate the height of a flagpole by manipulating its shadow. Thus, since understanding is the ability to produce explanations, and explanations appeal to the causal dependencies about what would happen under manipulations, understanding is the general ability to identify causal dependencies and answer questions about what would happen given counterfactual manipulations of parameters.

The skill of being able to answer questions about what would happen given this or that manipulation sheds some light on why brute force calculation is not enough for understanding. We do not get understanding from an oracle that gives us correct answers to counterfactual questions, we can only get understanding if we see how to derive those answers. We understand when we can "look inside black boxes" and see how they work. We can do this better if we have a qualitative grasp rather than just the ability to calculate quantitatively. We understand better when we can detail and identify inner mechanisms of events. One of the reasons for this requirement is that on the one hand, if we can work out what would happen under different scenarios, we are not reliant on a memorized list of scenarios and results. On the other hand, if we can work out the answers without doing calculations, we are able to see wider trends and specific qualitative dependencies, rather than requiring a new calculation for every situation.

There is a proposal that identifies the ability to draw qualitative consequences from the-

ories as the defining characteristic of understanding. This proposal notes that we do not fully understand why an event occurred if all we can do is predict/retrodict its occurrence from some generalizations without a theory that governs its occurrence. We still do not understand if we can merely predict/retrodict the phenomenon by going through an exact calculation from a theory. We must have a more intuitive grasp of the theory to have understanding. (de Regt and Dieks 2005) encapsulates this in the concept of an intelligible theory: a theory is intelligible to someone just in case she can make qualitative predictions with it, without recourse to detailed calculations. De Regt and Dieks use this to promote the following account of understanding: "A phenomenon P can be understood if a theory T of P exists that is intelligible (and meets the usual logical, methodological and empirical requirements)." (de Regt and Dieks 2005, p. 150) To understand why a phenomenon occurs, one needs to be able to derive it from a theory without calculation. Having an intelligible theory implies possessing a skill, which if properly developed provides a solution to the problem of jumping to false idealizations, raised by (Cartwright 1995). The intuitive grasp of qualitative consequences can serve as a sanity check.

In deriving this account, de Regt and Dieks make a claim about the nature of understanding that is independent of the details of their specific account. They claim that the intelligibility of a theory, its ability to produce understanding, is context-dependent. If we drop de Regt and Dieks's requirement that understanding must be based on scientific theories, their claim is that the conditions that can constitute understanding vary with context. For example, at one stage of scientific development, with one background set of tools and theories, an indeterministic theory might be able to provide understanding. If scientists have a good enough grasp of statistics and an intuitive sense of their implications, an indeterministic theory could be intelligible. At an earlier stage, the probability calculus might strike scientists as an opaque wall of math. Action at a distance might be easy to work with in one field of science, but not in another. They claim that ability for a true (or at least empirically adequate) theory to produce understanding is context-dependent.

There are also accounts of understanding that are less concerned with application of knowledge, but still reject the two horns of the dilemma above. These accounts claim that understanding is a distinctive cognitive connection to reality. Kvanvig says that this cognitive connection takes the form of having a body of true beliefs which are coherent, in the sense of 'coherent' employed by internalist coherentist epistemologies. This involves awareness of "how the various elements in a body of information are related to each other in terms of explanatory, logical, probabilistic, and other kinds of relations that coherentists have thought constitutive of justification." (Kvanvig 2003, p. 193). Disconnected beliefs do not constitute understanding, but once these beliefs are unified into a coherent world-picture, unified by a few basic principles, we do have understanding. This amounts to grasping the propositional structure of part of the world in a comprehensive way.

Zagsebski takes a similar line, and argues that "understanding is the state of comprehension of nonpropositional structures of reality." (Zagzebski 2001, p.242) Things like paintings and music have structure that is nonpropositional, and understanding a painting is a central case for her. She also speaks of part/whole relations as falling into the category of non-propositional structure. According to her, understanding comes from mastery of a craft or *techne*, and having this mastery allows us to "penetrate more deeply" into the nature of parts of reality related to the techne, in part by forming a nonpropositional grasp of the structure. Due to the nature of a *techne*, "it is unlikely that understanding is achieved by a single mode of reasoning, but it involves more-complex processes, including, perhaps processes that are noncognitive." (p. 241) She avoids going into more detail than this, but it is clear that her account of understanding makes it distinct from propositional knowledge. Further, the successful mastery of a craft puts us in epistemic contact with structures of reality, making understanding more significant than a mere subjective feeling.

This covers the main players in current debates about understanding. There is an emerging trend of treating understanding as an ability to unite and apply information. Still, there are several questions to ask about these theories.

To give an example of an issue that arises and how it is dealt with, one issue is whether understanding is factive. Understanding is factive just in case it must be based on mostly true beliefs. (Kvanvig 2003) claims that understanding must be factive to maintain adequate connection to the world. (Riggs 2009), on the other hand argues that understanding is not factive. He claims that one can understand a person's fear of water based on the completely false belief that she almost died in a boating accident as a child.

Another issue is luck. (Kvanvig 2003) argues that understanding can be acquired by luck using the Comanche case. Suppose you read a book about the dominance of the Comanche history that was based on no research, just guesses. However, suppose the author happened to guess correctly about everything. Kvanvig claims that in this case you would understand Comanche history.

On the other hand, (Pritchard 2008) rejects this conclusion. He presents his own example. Suppose you arrive at your house to find that it is on fire. Out front you see someone dressed as a fire fighter and ask her what happened. She tells you it was a short circuit in the kitchen. You believe her. However, what you do not realize is that she is not a fire fighter; she is simply on her way to a costume party and made up an answer just for fun. Pritchard argues that even if she is right, you do not understand why your house is on fire. Understanding is incompatible with this kind of luck.

The views described above illustrate most existing views: general rather than detailed. The debates about factivity and luck described above illustrate the debates that exist and how they are dealt with: through intuitions about cases.

1.4 My approach

My approach differs from previous approaches. It does not rely so much on intuitions about cases. It also aims for more detail, both in the account itself but also in the direction for future research suggested by the account.

My approach to understanding is not as much an exercise in traditional analytic philos-

ophy as most previous work on understanding has been. I am less concerned with seeking to accommodate all intuitions about hypothetical cases than I am with finding a unified natural phenomenon that fits closely with general ideas about understanding.[4] I look to human cognition to find a role for understanding. My methodological assumption is that understanding is a feature of cognition whose value and role can be independently described based on considerations of the needs of a cognitive system. I assume that the role of understanding is important enough that we need a term (in this case 'understanding') to refer to it. Examination of linguistic intuitions about 'understanding' are not enough to learn the features of understanding. We must look to the world to see what understanding does for cognition, which may not perfectly align with intuitions about the use of the term. I leave open the possibility that usage of 'understanding' succeeds in picking out a phenomenon while some of our intuitions do not perfectly reflect the reality of that which is picked out.

This dissertation is an attempt to take this methodological assumption as far as possible to form an account of understanding. It aims first and foremost to identify the role that understanding plays in our cognitive systems. I aim to explain what role it plays in allowing a cognitive system to engage in successful inquiry. I identify an independent problem that cognitive systems must solve in order to function as cognitive systems at all, and argue that the ability to solve this problem is understanding. This ties understanding into several important areas of research and interest. This approach opens new avenues in the study of understanding.

Not all subjects will necessarily be well-suited to this kind of approach. It depends on whether there is a united phenomenon nearby, or whether there is a unified role involved. However, if the procedure works, we are assured of an account that is firmly grounded in nature, lays bare the value of the thing, and still answers to our ordinary concept with at most some mismatch around the edges.

As a starting point, there is much we can say about the role of understanding in inquiry.

[4]For more about relying less on intuitions, see for example, (Weatherson 2003) and (Kornblith 2007).

Understanding is the ideal end state of education; it allows one to go forth into the world and apply what one has learned. It is more than memorization, and it is more than rote learning. Understanding is a robust state of achievement that allows one to achieve further cognitive success. Understanding can be applied; one cannot understand something and yet be unable to do anything with that understanding. Understanding is valued in part because it is rich and useful in future inquiry. In developing my account of understanding, I attempt to come to terms with these facts, and clarify the role of understanding. To look forward to my conclusion: I argue that understanding gives one the ability to find the answers to questions. That is its role, and why it is so important.

As a starting point in psychology, the aim is to consider what the cognitive system must do in order to be successful in inquiry. My method here avoids over-reliance on introspection or intuitions. Instead, I will engage in what Dennett calls 'hetero-phenomenology,' which he describes as thinking "about the purely informational demands of the task - what must be known by any entity that can perform this task." (Dennett 1987) I identify something that is necessary for success in cognition, an independently identifiable requirement that aligns in several significant ways with the concept of understanding. In proceeding this way, we find that a defensible account of understanding is already related to existing research programs in psychology.

While my approach differs from previous approaches to the topic, it is able to establish many of the general points that previous approaches were able to establish, though in most cases with more detail. It is able to provide insight into the debates that are being carried out in the literature on understanding. This approach is not a complete rejection of everything that has come before, but rather is a new method of studying the details of understanding. The approaches can complement each other.

By beginning this way, we can be assured of the value of understanding, which is one of the aspects of understanding that has been considered significant. By identifying an aspect of cognition that is necessary for inquiry, we identify something that has epistemic value.

In addition to showing the value of understanding, this approach prevents the pointless gerrymandering, as happened with analysis of knowledge in the wake of Gettier. We avoid the pressure to find ever more complicated conditions to cover ever more outlandish counterexamples. It is not a purely linguistic analysis. The goal is to find phenomena in the world that 'understanding' picks out by focusing on the phenomena first. It does not rely on any assumption about the reliability of intuitions to pick out reality, carve reality at the joints or any such thing. By keeping our focus on cognition and its requirements, we make sure that when we find something, it is natural. The rest of the dissertation is an attempt to realize the potential of this approach.

CHAPTER 2

A PLACE FOR UNDERSTANDING: MANAGING COGNITIVE RESOURCES

2.1 Introduction

Understanding is a cognitive achievement for which we strive. Teachers aim to impart understanding of the subject taught, and exams are designed specifically to test for understanding. Experts' judgments are sought due to their understanding, and amateur judgments are discounted due to lack of understanding. Understanding is central to epistemic activity. [1]

However, as we saw in the last chapter, understanding has been largely ignored in analytic philosophy due to the widespread belief that it is ultimately a derivative phenomenon, trivially reducible to concepts already well explored. For example, many have claimed that understanding is just a kind of knowledge, such as well-integrated knowledge, or very detailed knowledge, or knowledge of explanations. If this thesis is correct, then the bulk of philosophically interesting work related to understanding lies in getting a grasp on knowledge. An account of understanding will emerge as a by-product. This assessment can be found in work from (Hempel 1965) through (Salmon 1990b) and to (Trout 2002), (Strevens 2009), and (Khalifa 2012).

The general concern is that epistemology contains four basic types of evaluation: eval-

[1] As the last chapter explained, understanding is now the subject a growing literature in analytic epistemology, for example detailed discussions of understanding can now be found in (Kvanvig 2003), (Pritchard 2008), (Elgin 2007), (de Regt and Dieks 2005), (Grimm 2006), (Lipton 2009), etc. The central claim of this recent work is that philosophy would be abandoning important insight by neglecting the study of understanding in its own right.

uation of knowledge, justification, rationality and truth. Without a demonstration that understanding is not easily reducible to these established fields of study, at best understanding will be a part of one of the more important categories. There is no room in epistemology for additional fundamental categories of evaluation apart from these.[2]

In this chapter, I argue that the standard epistemic evaluations—evaluations of knowledge, justification, truth and rationality—do not exhaust epistemic evaluation. In particular, I show that they cannot be used to evaluate an agent's ability to apply the right cognitive abilities to the right cognitive states at the right time in order to achieve cognitive goals. Evaluating the ability to do this is an indispensable form of epistemic evaluation that is not reducible to any of the standard epistemic evaluations. I propose that the ability to apply the right cognitive capacities at the right time should be considered a form of understanding due to the fact that it plays many of the roles the ordinary notion of understanding plays. Understanding is an independent epistemic phenomenon.[3]

The argument proceeds as follows: In Section 2.2 I introduce cases that each involve two agents answering a question, where one agent's performances is epistemically superior to the other's. In Section 2.3, I argue that the differing epistemic evaluation seen in the cases are not reducible to truth, justification, knowledge or rationality. In Section 2.4, I argue that what does make the difference in the cases is how well agents manage cognitive resources, and I explain what managing cognitive resources entails. In Section 2.5, I describe the role that managing cognitive resources plays in teaching and learning. In Section 2.6 I I show the importance of cognitive management for long term memory. In

[2]Some have suggested that know-how and epistemic virtues are also irreducible, but these claims face skepticism just as the irreducibility of understanding has. A prominent theory of knowing-how is that it is just a kind of knowing-that, which implies that study of knowing-how is merely an application of the study of knowing-that. If, on the other hand, knowing-how is independent of knowing-that, and if understanding is a form of knowing-how, this would be unite two current subjects of skepticism and show their independent value. Similarly, some theories of knowledge in recent times have centered on epistemic virtues. If understanding turns out to be a kind of epistemic virtue that is not reducible to truth, justification, knowledge or rationality, that remains an interesting conclusion.

[3]In Section 2.9 I argue that cognitive management should be considered a kind of understanding. In the next chapter I build upon the concept of cognitive management to give a more complete analysis of the ordinary concept of understanding.

Section 2.7, I show the importance of cognitive management in expertise. In Section 2.8, I argue that evaluation of cognitive management ability is epistemic, against objections that it is a practical means/ends evaluation. Section 2.9 concludes by proposing that cognitive management should be considered a form of understanding.

2.2 Cognitive tasks

The argument that there is an additional form of epistemic evaluation begins with the following two cases:

Geometry Two students, Better and Worse, are trying to prove that there is no regular polyhedron with 14 faces. They both know the axioms of geometry and many theorems, and both are competent at constructing proofs. While working on the problem neither makes any mistaken inferences, and every proof they write down is in fact a geometric proof. However, Better immediately sees what she has to do to get the proof, while Worse spends days proving irrelevant conclusions before finally coming to the answer.

Division Better and Worse are given the homework question, "How many buses does the army need to transport 1,128 soldiers if each bus holds 36 soldiers?" They both start working on dividing 1,128 by 36. Better does the division, gets 31 remainder 12, and answers that 32 buses are needed. However, Worse gets 31 remainder 12 for the division but then moves on without thinking about whether she has answered the question. She looks over her answers checking for mistakes in calculations, but never considers whether she has finished getting the answer. Two days later, she comes back and notices that 31 remainder 12 is not a number of buses, and puts down the final answer 32.[4]

[4]This example is adapted from the results of an actual study of eighth grade students, reported in (Schoenfeld 1988). Approximately 1/3 of students tested gave 31 remainder 12 as the answer.

The cases above each have pairs of cognitive performances with one clearly superior to the other. In each case, Worse's performance is epistemically inferior to Better's performance. In Geometry, Worse works in circles before finding the answer. In Division, she does not pay attention to what she is doing until the last moment. The question is: when we evaluate the agents' performances, what kind of evaluation are we making? What failures justify assessing Worse as deficient? Did she make mistaken inferences, or come to have inconsistent or unjustified beliefs? Was she wrong about something? Our goal here is to determine by which form of evaluation Better and Worse differ in each case.

I argue that we judge Worse to be deficient by evaluating how well she tracks her cognitive goals and how well she deploys the right cognitive capacities. Further, this type of evaluation is not reducible to evaluations of truth, justification, knowledge, or rationality. Examples like the above reveal that there is a form of epistemic evaluation distinct from the standard forms of evaluation appealed to in epistemology. Epistemology cannot be complete without getting a handle on this evaluation.

Now let us show that in each of the cases given above, the variation exhibited cannot be explained in terms of truth, rationality, knowledge or justification.

2.3 Evaluating the evaluation

The cases above involve answering specific, predetermined questions. Answering specific questions is often how we learn specific facts. We start with a question that we need to answer. Does the Higgs boson exist? Does the death penalty have a deterrent effect? Where are my keys? What is over there? Then we work to answer that question. We engage in inquiry to answer it. When we find the answer, we have learned about the world.

Inquiry is the process of raising and answering questions. Once a question is singled out as requiring an answer, this can lead to further sub-questions that require answers, as when trying to answer the question of how much the Higgs boson weighs, one must answer the question how to measure the weight of such a particle. The ability to answer specific

questions is necessary for success at any kind of inquiry. The evaluation of the above cases relates to this necessity.

One might naturally expect that the difference in evaluations of Better and Worse in the examples from the previous section can be explained in terms of true belief, justification, knowledge or rationality. These four—truth, rationality, justification and knowledge—together constitute what I will call *the standard suite of epistemic evaluations*. The first step toward demonstrating there is space for understanding in epistemology is to show that none of the standard suite of evaluations is the one we are looking for.

2.3.1 Truth

One way for one's epistemic performance to fall short is for one to accept a proposition that is not true. Taking this as a starting point, a natural proposal is that in the examples above, the difference between Better and Worse is that Better has an advantage in true beliefs. Attempting to draw conclusions based on misinformation often leads to failure, which could potentially explain why we judge Worse to be worse. However, this proposal is inadequate to explain the cases above, as I will now show.

Geometry: Consider first the case of Better and Worse attempting to prove the geometry theorem. The difference is not that Better forms the true belief that X is a proof and Worse does not, since they both eventually find the proof. Also, neither starts out with true beliefs about the proof, so they both go from lacking the true belief at the start to having that true belief in the end. The difference in evaluation is due to how they acquire the true belief, not whether they have it.

Both agents make valid inferences from axioms and theorems of geometry. Neither agent has false beliefs about the inferences she makes. From true beliefs, Better and Worse make only proper, truth-preserving inferences. Finally, neither forms the false belief that she has the right answer when she does not; to start both agents know they do not have the proof, and at the end they know they do have it. False beliefs regarding geometry thus

cannot account for the difference between the two. The beliefs that play a direct role in their deliberations are not false.

One might still claim that even though false first-order beliefs cannot explain the difference, false meta-beliefs can. Perhaps Worse makes the inferences that she does due to false beliefs about those inferences. However, this also cannot give us the explanation we seek. First, Worse is correct that the sequence of inferences she makes will eventually lead to the proof, so that cannot explain the difference. She could have a false belief if she believes that she is using the optimal method of finding the proof; however, there is no reason to assume she believes she is acting optimally. She might realize she is not using the best method, but randomly making proofs is the only method she can think of at the moment. Further, if she does not have the intuitive geometry skill that Better has, then making random inferences without repetition until she stumbles across the proof may be the optimal method given her skill set. Then, if she believed she was using the optimal method *given the circumstances*, it would be a true belief. False meta-beliefs do not explain the differing evaluations.

The final suggested explanation in terms of truth is that lack of true beliefs, rather then possession of false beliefs, is the difference. However, this is inadequate as well. To start, both lack the true belief about what the proof is, and at the end, both have it. Again, if there is a difference here it is in meta-beliefs.

The most promising candidate is that the difference is that Better has a true belief about how to find the proof quickly, but Worse lacks this belief. However, simply having a true belief about methods to find the proof will not by itself bring Worse level with Better. Call the method Better uses to find the proof Method X. Suppose that Worse truly believes that Method X will give the proof. By itself, this belief does nothing; Worse must recall this belief at the right time, and use that belief to decide to employ Method X. Supposing she has the belief but does not call it to working memory, she still does worse. The example remains realistic and still exhibits the same difference in evaluation. No matter which additional true beliefs we give to Worse, the need to put them to use at the right time will

still apply. Thus, lack of the proper true belief is not the explanation we are looking for. Truth cannot explain the difference.

In addition, Better need not believe she is using an effective method in order for her performance to be better. She only needs to use the effective method. Again, true belief does not make the difference.

Division: Turning to the example of Better and Worse calculating the number of buses, both did their long division correctly based on true beliefs about the number of soldiers and the properties of the buses. Neither formed the false belief that *31 remainder 12* was a number of buses. Worse merely quit working on the problem after completing the long division. She was not thinking about whether she had the answer to the question yet. This was not due to any false beliefs; it was due to lack of attention to whether she had finished the problem. The difference is not that one had a true belief that the answer was 32 and the other did not. They both got the right answer.

Taking stock, both correctly inferred that the problem could be solved using long division. Then both correctly set up the long division. They both did the long division correctly. Then Worse moved on without addressing whether she had finished answering the question or not. She had in fact left it at a point where she did not have an answer to the question. Her performance is still problematic because she did not track the relevant question. She did this without relying on or forming false beliefs. Thus, false beliefs cannot be the explanation.

Furthermore, for the reasons given in our discussion of Geometry, adding one more true belief, no matter what it is, will not by itself bring about parity between Better and Worse. Any belief must be acted upon in the right way at the right time in order to help. Worse could be assumed to have any true belief and yet she could still fail by failing to act on it or recognize that it applies.

We can summarize the issues with appealing to truth by analogy with computers: one can input the exact same data into two computers, but if they use different methods of

accessing data and different methods of manipulating it, they can differ radically in performance. The same is true of epistemic agents: they can have the same true beliefs, but employ them better or worse. Appeal to differences with respect to truth fails to explain the differences in evaluation.

2.3.2 Justification and knowledge

If the difference is not due to truth, perhaps it is due to other features of beliefs, such as whether they are justified or constitute knowledge. Justifiedly believing or knowing the answer is not the issue: in each case, Better and Worse get the correct answer, and we can assume that at the end they know the answer and are justified in believing it.[5] If justified beliefs or knowledge make the relevant difference, it cannot be justified belief or knowledge of the answer. Since we have shown above that neither false beliefs nor lack of true beliefs explains the difference, the truth component of knowledge cannot be the relevant factor. Theories of justification and knowledge identify epistemically good features of the formation or grounding of beliefs, so the proposal here is that these features make the difference in Geometry and Division. However, this proposal fails to account for the difference.

Geometry: Better and Worse prove a geometry theorem, though Better does better than Worse. Is the difference that Worse forms new beliefs that are not justified or are not knowledge? No, for both eventually come to have a mathematical proof. We can assume that both have sufficient mathematical competence to allow their beliefs to be justified and to constitute knowledge. This does not explain the difference.The same is true no matter how many explanatory, probabilistic and inferential connections are required of beliefs before they can count as justified or as knowledge. With no way to argue that Worse does not actually know the theorems she proves along the way, there is no room to claim that

[5]By discussing knowledge and justification together, I do not mean to take a stand on the relationship between justification and knowledge. The similarity in how they must be dealt with would make separate sections redundant.

justification or knowledge make the difference.

One might claim that Better has greater justification or knowledge than Worse, and the features of her beliefs that confer greater justification/knowledge make the difference. However, like having further true beliefs, this cannot account for the difference either. No connection by itself guarantees that the right sequence of inferences and steps will actually be taken. Again, having these features among one's beliefs does not close the gap unless they are acted upon. We can even assume that Worse has the same knowledge and justified beliefs, justified to the same degree, but fails to exploit those particular connections when attempting to find the proof. The difference in evaluation is still present, unaccounted for by differences in justification and knowledge.

As long as Worse continues to draw upon suboptimal connections and make suboptimal inferences, the fact that her beliefs stand in the right logical relations does her no good. One can have a tightly connected body of knowledge yet still draw upon a useless portion of it. Justification and knowledge cannot explain the difference.

Division: Justification and knowledge cannot explain the difference in performance in calculating the number of buses, for reasons similar to those that lead to the same conclusion in Geometry. The deficiency of Worse's performance is not a matter of lacking justification or knowledge, it is a matter of failing to pay attention to the knowledge she has. Both Better and Worse knew what they were doing, but Worse simply did not pay attention to that knowledge, and moved on before getting the answer, only rectifying this much later. There can be as many connections among her beliefs as you want, but as long as Worse stops early before concluding that 32 buses are needed, it will not help. Merely having beliefs built on knowledge- and justification-conferring connections does not mean one will exploit those connections at the right moment.

Thus, evaluations in terms of justification and knowledge do not explain the differences in the cases. We must look elsewhere.

2.3.3 Rationality

We have shown that it is not the beliefs or their properties that make the difference, but that does not rule out the possibility that the difference is a matter of rationality. While this is not the place to fully explain epistemic rationality, the general idea is captured by Thomas Kelly's description of it as, "roughly, the kind of rationality which one displays when one believes propositions that are strongly supported by one's evidence and refrains from believing propositions that are improbable given one's evidence." ((Kelly 2003), p. 612) Another way to characterize epistemic rationality is that one is rational if one avoids incoherence in one's overall mental state (this characterization is prominent in Dutch Book Arguments, for example).[6] Rationality includes making only inferences that are in accord with the relevant logic (whatever that ultimately turns out to be), and how to maintain coherence dynamically. For example, Bayesians believe that in order to be rational, one's credences must conform to the axioms of probability and one should update on new evidence via conditionalization on the evidence.Whatever form rationality takes, however, it does not explain the difference.

Geometry: When Better and Worse prove there is no 14-faced regular polyhedron, they make mathematically sound, truth-preserving inferences from known premises to true conclusions. Mathematically sound inferences are not irrational. Their resulting total bodies of belief do not become incoherent as a result of making these inferences, for the new beliefs follow deductively from their initial beliefs. The inferences themselves do not differ in rationality.

Some theories of rationality (for example, standard Bayesianism) imply that lack of logical closure among one's beliefs is irrational (or at least is not perfectly rational). At the beginning, Worse did not believe all of the logical consequences of her beliefs, otherwise she would have had the proof already. However, this does not explain the difference

[6]Dutch Books arguments phrased in terms of incoherence can be found in (Hitchcock 2004) and (Christensen 1991) among others.

between the two, for Better also lacked logical closure among her beliefs. The main difference between them is the route they take to find the proof. However, models of rationality that consider lack of logical closure to be a defect do not dictate which particular sequence of otherwise rational inferences is a better way of more closely approaching the rational ideal.[7] Furthermore, by taking a longer route to the final proof, Worse acquires more true beliefs (irrelevant beliefs, but still true) about geometry along the way than Better does, so one might claim that she comes closer to matching the ideal of closure. Lack of closure does not explain the difference.

Perhaps the way in which Worse chose which inferences to make, rather than the inferences themselves, is irrational. One might suggest that Worse violates practical means-ends rationality. However, this does not explain the difference, either. While it is true that Better uses more efficient means to achieve her end, this does not equate to a difference in rationality. If Worse knowingly chose an inferior method over a superior one, she would be irrational, but we can assume that Worse knows that she lacks Better's ability to simply see what to do, so that for Worse the best way to ensure she gets the answer is to keep trying proof after proof. If she does this long enough, she knows she will eventually stumble across the proof. Knowing this, her approach is rational, it is the best means available to achieve her end. Again, rationality does not explain the difference.

Division: As in the Geometry case, when calculating the number of buses Better, and Worse make mathematically sound, truth-preserving inferences from known premises to true conclusions. At the point of calculating 31 remainder 12, everything the students have done is both practically and epistemically rational. They took the right approach to the problem, and they carried it out correctly. It was not incoherent for them to think their approaches would work, for they did work for both students. They both got the answer.

[7]There are debates about the nature of rational requirements and how to respond, such as the debate about the scope of rational requirements. This debate is orthogonal to the issue here. The scope debate asks whether the requirement is that given a belief that p, one is rationally required to believe that q, or whether the requirement is that one ought to be such that one either believes both or neither. Neither helps explain the difference. For a summary of this debate, see (Way 2010).

If rationality makes a difference, it must do so at the stage when Worse moves on to other problems before getting the final answer. Perhaps the fact that she was not thinking about whether she had come to a final answer is irrational. She did not ask herself whether she had answered the question, and therefore was able to move on before having the answer (though she did get the answer in the end). The issue here concerns the order in which she performed cognitive tasks. This is only a matter of epistemic rationality if epistemic rationality prescribes restrictions on which paths of otherwise-rational reasoning one is allowed to make when answering a question, and there does not seem to be any such limit.

Given that one cannot instantly draw every conclusion possible from every belief one has (assuming one is a finite being), one must organize one's reasoning, taking some steps before others. Comparing the current state of one's reasoning to the desired goal is a step that requires cognitive resources. In order to compare one's current stage in a calculation to the intended answer, one must act upon the beliefs regarding the nature of the question and what would answer it. The fact that Worse made other rational inferences between completing the long division and inferring the final answer is not a matter of irrationality.

While there is something deficient about moving on before having the answer, there does not appear to be any explanation of this deficiency in terms of rationality. Better and Worse correctly inferred that long division would solve the problem. They did their division correctly, but one of them did not have the right answer until the last moment. This does not violate any constraints of rationality. Thus, rationality—and therefore the standard suite—cannot explain the difference.

We are now in a position to determine in more detail what kind of evaluation allows us to conclude that Better does Better than Worse in the Geometry and Division cases. What does it mean to succeed and fail in the relevant ways demonstrated in the examples above?

2.4 Inquiry

The above analysis shows that deficiencies in how one answers questions cannot always be explained in terms of any member of the standard suite of epistemic evaluations. To find the explanation, then, we turn to the process of inquiry, the process of raising and finding the answers to questions. When we engage in inquiry, there are often specific questions that guide us, whether they are as mundane as where our keys are, as vague as what is over in some general direction, or as difficult as whether the Higgs boson exists.[8] Sometimes, we find answers without looking for them, but frequently the goal of inquiry is to seek and answer questions. Successful inquiry depends on the ability to answer specific questions.[9]

As illustration of how a particular question can guide inquiry, consider a police detective. Her primary goal is to find out who committed the crime. The question of how many seconds it takes to drive from the police station to the crime scene is not relevant. Even if she can answer this question without being distracted from the main question, answering this does not help her. The same holds in the sciences. When an archaeologist studies an artifact, among the questions to answer are how old it is, what it was used for and by whom. However, it would be foolish to become consumed with finding out how far it is from the nearest fire hydrant, how it tastes, or how long it will remain intact in a washing machine. Some questions are set apart as salient, and others are irrelevant; a skilled investigator must be able reliably to tell the difference.[10]

There are five modes in which one's method of answering a question can be deficient.

1. One can entirely lack the capacity to grasp the question or its answer;

[8]Defenses of different versions of the claim that inquiry is the process of finding answers can be found, among other places, in (Craig 1990), (Stalnaker 1984), (Schaffer 2004), (Hintikka 1988), (Hintikka, Halonen, and Mutanen 1999), (Levi 2003), (Williams 2004).

[9]Inquiry that is not guided by questions is also possible. For example, one might walk a specific pattern through the woods in order to gain knowledge, yet do so without having any question in mind. Though possible, the inability to engage in question-guided inquiry would be catastrophic to any cognitive system's use in finding out about the world, unless it had infinite capacity and thus could find every answer, and therefore the needed answer, instantly. This can be ignored since no human has infinite mental capacity.

[10]A being with infinite capacity could tackle all questions at once, but for those of us with limitations, issues of allocating resources are important.

2. One can come to the wrong answer to the question;

3. One can use fundamentally flawed methods;

4. One can use acceptable methods but in such a way that they do not lead to the answer.

5. One can lose track of the question and thereby cease to work toward the answer;

The first three of these are reducible to the standard suite of evaluation, but the last two are not. First of all, mode 1 is a consequence of finitude in conceptual capacities. Of the remaining, mode 2 is captured by true belief or knowledge: when one accepts an answer that is false, even if one's belief is justified or rationally formed, it is a false belief and is not knowledge. Next, 3 is captured by rationality or justification, since the use of fundamentally epistemically bad methods is what irrationality and lack of justification are.

However, I will show that 4 and 5 are the modes operating in the Geometry and Division cases, and are not failures of truth, knowledge, justification or rationality.

2.4.1 Deployment

Deficiency 4 can be rephrased as inadequate deployment of cognitive resources. When one deploys inferences that do not lead toward the needed answer, this is a problem, even if one keeps one's attention on the proper question. The methods one uses may be rational and capable of conferring justification; they may result in knowledge along the way; however, if they do not move one closer to the right answer to the guiding question it is still a deficiency in performance.

Consider, in the Geometry case, Worse tried to find a proof but spent days proving irrelevant theorems, though making mathematically sound inferences. She failed to employ the cognitive capacities she had in a manner that would take her to the correct answer to the question. She failed to deploy her cognitive resources optimally, given the question. There are inferences *that she had the ability to make* that would have taken her to the right answer, but instead of making those inferences, she made others that were not so helpful.

The difference between Worse and Better is that Better deployed her available resources more effectively.

Plato has gives us another illustration of the difference between those who can effectively deploy resources and those who cannot. In the *Meno* Socrates guides Meno's slave through a proof that a square whose side is the diagonal of a second square has twice the area of that second square. The slave produces every inference and belief along the way to the final answer. Socrates uses nothing that the slave does not at least implicitly know, and uses no inference that the slave is incapable of making. In fact, the slave makes each inference himself in response to questions. However, the slave could not have started with the original problem and worked his way to an answer. The slave lacked the ability to select those inferences out of all the inferences available. On his own, he would have failed via mode 4, but Socrates, with his skill, guides the slave to the right deployment of resources for success.

One potential concern about cognitive deployment skill is that it presupposes that the mind is organized in such a way that a sort of selection of capacities to deploy occurs. This concern is not fatal to the idea of cognitive deployment: research in psychology suggests that deployment is a real phenomenon. Many examples of deploying resources can also be found in the active research program in psychology studying what are known as *executive processes*. Executive processes modulate the operation of other psychological processes, for example by shifting between different cognitive tasks, shifting between different sets of background beliefs, updating the contents of working memory, and inhibiting automatic responses, among others (Miyake, Friedman, Emerson, Witzki, and Howerter 2000). The general direction of this research is to study how executive processes govern which lower-level psychological processes are deployed and how this relates to task success. There are processes in the mind that regulate which processes operate when. In short, the study of executive processes is a program directly studying cognitive deployment.

2.4.2 Monitoring

Mode 5 can be rephrased as the failure to monitor the cognitive task one is performing. One can use rational methods and come to have justified true beliefs that constitute knowledge, and yet if one loses track of what one is doing, one can fail to answer the question that guides inquiry.

While this failure does not occur in Geometry, it does explain the difference in Division, where the students calculate the number of buses the army needs. After reading the question, Better and Worse knew that to answer it they needed to set up a division problem. They reached the right answer to the division problem using justified, rational methods. However, by the time Worse finished the division problem, she was no longer tracking the initial question. If she had considered whether she had the answer, she would have realized she did not, but as it is she did not consider whether she was done. This was a defect in her performance. This shows a failure to monitor the cognitive task. If one does not keep in mind the relevant question, one can very well come to a true belief that is justified and constitutes knowledge, and yet completely fail to answer the guiding question. Worse failed to monitor her cognitive activity. Though eventually she made up for this mistake, the fact that it happened warrants an epistemic evaluation that she did worse than Better.

Another illustration of problems with monitoring is a variant of the Stroop task. When subjects are asked to report the color of a list of random words printed in different colors of ink, they generally do not have much difficulty. However, if asked to report the color of a list of color words that are printed in ink of a different color (for example 'red' printed in green ink), it is much more difficult. If forced to read a long list quickly, subjects eventually lose the plot and start reading the word instead of reporting the color of the ink. Evidence suggests the cause of this mistake is that subjects lose track of what they are meant to report. Saying color words out loud makes reading the color words salient, so it requires focused attention to recall that the task is to report the color, not read the words.[11] When

[11]The Stroop task is so difficult because reading a word is automatic while reporting colors is not. Thus,

attention falters, people start correctly reading the words, but this is not the task. They are correctly answering the wrong question, namely, "What word is this?"

The subjects have the abilities to identify words and colors. Using vision under the test conditions is a reliable method of forming beliefs both about what word is printed on a page and what color ink is on a page (or what color is displayed on a screen). It is not that subjects lack the capacity to recognize the words or the colors. It is not that they make irrational inferences, such as "This word is in red ink, therefore, the word must be 'red' ". However, they fail to answer the relevant question because the subjects have failed to monitor the task and keep the guiding question in mind.

The standard suite of epistemic evaluations does not provide the resources to evaluate cognitive monitoring. Each of the standard evaluations evaluate whether inferences are sufficiently truth-conducive to establish a conclusion, sufficiently justification-conferring to produce a justified belief, or are sufficiently rational. They look at issues such as which response to a given body of evidence is epistemically best. None of the standard evaluations evaluate the degree to which the steps one takes lead toward the answer to a particular question. How well one monitors cognitive tasks is a way in which we can evaluate one's epistemic functioning, and it is not reducible to the standard suite of evaluations.

2.4.3 Management and its evaluation

Collectively, monitoring and deployment make up the broader ability of *cognitive management*, the ability to use one's available cognitive capacities and background information to achieve cognitive goals. Cognitive management is the difference between actively answering a question and merely happening to find an answer. Deliberately answering a question occurs when one has sought the answer to the question, and unless it occurs through luck, this depends on managing cognitive resources.

---

one must focus enough attention to inhibit the reading response. As attention wanes, one simply correctly reads the word. See (Miyake, Friedman, Emerson, Witzki, and Howerter 2000), (Engle 2002), for details and analysis of the Stroop task and related tasks.

Let us consider how evaluation of cognitive management works. For individual tasks, the evaluation consists of answering a few questions. Is there evidence that the agent lost track of the task? Is there evidence that she was able to deploy the appropriate resources to achieve the answer? Did the agent take the most efficient cognitive route to the answer? The better the agent did in each of these categories, the better her cognitive management.

However, evaluation of cognitive management goes beyond performance in individual cases. We can evaluate a cognitive system as a whole. We can assess how reliable someone's cognitive system is at correctly answering a wide range of questions, including novel questions that are unlikely to be solved by rote application of a memorized procedure.

The most thorough means of carrying out such an evaluation requires knowledge of how the person under evaluation reaches conclusions. Relevant information includes: the reasoning she followed, the psychological processes employed, and how she directed her attention. This information is valuable in determining why someone fails when they do and in determining whether successes are due to cognitive management or whether success was due to bad methods happening to work.

The basic fact is, when an agent confronts a question and begins inquiry, cognitive activity occurs. This involves directing attention, recalling stored beliefs, and making inferences, among other things. Any cognitive system is arranged, by habituation or design, to respond to inputs with some sequence of cognitive activity. It cannot simultaneously perform every form of cognitive processing it has the capacity to perform; it is set up so that it starts with something specific. Systems can differ according to the kinds of cognitive moves they tend to make first. For example, one might be more sensitive to motion, another to bright color. The sequences of moves a system is disposed to make in different circumstances can have greater or lesser chance of leading to success. This is what evaluating cognitive management is all about: whether the system's arrangement works reliably within a range of inquiries.

For example, suppose that as a matter of habit or design a person's visual system tends

to focus on things that are red. Once he notices a red thing, it captures his attention and he will try to figure out what it is. After doing that, he will search for all visible non-red parts of that object and try to determine why they are not red. Only then will he move on to anything else. His cognitive system is designed/habituated in such a way that it will usually focus on irrelevant details (unless finding red things is the task). When this person is in a room full of red objects and he fails to determine why the refrigerator is not working, it may be due to the fact that he is so distracted by red that he cannot focus on the refrigerator. This need not be a failure of true belief, rationality, justification or knowledge; he just automatically gets distracted. His skill at cognitive management is poor across a wide range of inquiries.

For another example: suppose as a matter of design or habit, a person always applies modus ponens when it applies to two beliefs she has in working memory. Now suppose that she needs to apply disjunctive syllogism to two beliefs to solve a problem, but along with the two beliefs that will yield the conclusion, there are chains of beliefs to which modus ponens applies. As soon as she finishes one modus ponens, she moves to another. She will never reach her goal, not because of failures of truth, rationality, justification or knowledge, but because her cognitive system is poorly organized for cognitive management.

Cognitive management can sometimes be carried out via conscious decisions. However, some management must be due to sub-personal processes dictating cognitive activity. At some point, our minds must simply begin to work without a prior conscious decision to do so, on pain of vicious regress. Whether personal or sub-personal, the evaluation is always about how well the system is able to direct processes toward the answer to a question.

Unfortunately, we will rarely have detailed knowledge of the inner workings of a person's cognitive processes. Thus, we often must evaluate by observing patterns of success and failure in answering questions, observing the claims made, and observing patterns of attention. These can all provide evidence about how well a person is able to manage their cognitive resources over how wide a range of inquiries. Isolated cases of success or

failure in answering questions reveal the need to evaluate cognitive management, but the evaluation of management is a matter of assessing competence rather than individual performance. The underlying level of management ability revealed by patterns of success and failure is the primary locus of evaluation.

In the next three sections we take a tour through aspects of our epistemic lives that rely on cognitive management, situations in which evaluation of cognitive management is crucial.

2.5 Teaching and learning

Cognitive management is central to teaching and learning. Consider the goals a teacher has when teaching a group of students. While imparting propositional knowledge or rigidly defined procedures to perform tasks are often goals of teaching, this alone is not enough. Likewise, while teaching students to reason in accord with rationality (e.g. teaching students to avoid fallacies) or encouraging students only to form justified beliefs is important, this does not exhaust the goals of teaching either. Ideally, teaching should impart open-ended problem-solving skills. The overall goal is to teach students how to become better at inquiry, to teach them how to learn. Cognitive management is an indispensable component of problem-solving, and a necessary part of the goal of teaching.

Publications and discussions about education abound with laments about how exams too often only test for memorization, that textbooks too often present the subject as a series of disconnected facts and procedures to memorize, and that students lack the motivation to truly learn the subject, as in (Skemp 2006). These laments are about more than just a lack of knowledge; teachers are not complaining, "My students know these answers but not those ones." The complaint is that students lack the ability to work out the answers to questions in novel situations, or to work out solutions to problems using methods that they have not used before. The problem is that students cannot work out answers to novel types of problems even if the variation is tiny.

The following true story about a math teacher illustrates what is at issue.

> While teaching area he became suspicious that the children did not really understand what they were doing. So he asked them: "What is the area of a field 20cm by 15 yards?" The reply was: "300 square centimetres". He asked: "Why not 300 square yards?" Answer: "Because area is always in square centimetres."
>
> To prevent errors like the above the pupils need another rule [...] that both dimensions must be in the same unit. This anticipates one of the arguments which I shall use against [memorizing mechanical procedures], that it usually involves a multiplicity of rules rather than fewer principles of more general application. ((Skemp 2006), p. 90)

The students in this class know how to calculate the area of a rectangle when both dimensions are given in centimeters. However, they only have the ability to make one kind of inference in area problems, and they end up using it when it is no longer appropriate. The goal of the teacher is not just to teach another rule to cover the new case. While it is true that memorizing a longer list of rules to cover the new case will help one do better, if one's list of rules says what to do in situations A through T, this will not help you in situation U unless one has the skill to extend and re-tool current methods for novel problems. Remedying this shortcoming involves more than teaching students to make only rational inferences, or to make only justified inferences, or to maximize the amount of knowledge they acquire. Students also need to learn cognitive management. In principle, one could learn to solve different problems by building up an ever longer list of rules. However, as each new distinct rule is added, the difficulty of keeping track of the right one for any given situation increases. Cognitive management remains crucial.

In teaching a subject, one tries to instill—by training, inculcation, and encouragement—the habits and dispositions to respond to the available information in a way conducive to success in achieving cognitive goals. The brain is plastic to some degree, and patterns of inference can be trained. Success is not always a matter of imparting explicit knowledge of what to do in what circumstances. The capacity to recognize what is relevant, for example, cannot happen through explicit knowledge, for one would have to notice something as relevant before one could apply the knowledge that it was relevant. Effective

teaching must impart tendencies and dispositions for success.

The importance of cognitive management is also reflected in the manner in which students are tested. Ideally, students are not given problems they have seen before, or even structurally similar problems. They are given a range of novel problems that relate to the subject. This is to rule out the possibility that students can succeed merely through memorization of key facts and a few mechanical procedures. The test is to see if students have gained skills in managing cognitive resources in order to solve problems.

This covers teaching, but the learning side of the equation is not handled any better by the standard suite of epistemic evaluations. Learning only partly happens through inferences of the kind traditionally discussed. The formation of a belief is not the only important factor in learning. We must also consider how well knowledge can be recalled later, or whether information stays in memory or not. Forgetting can also be useful (for more detail, see (Michaelian 2011)). Research into learning has produced an extensive literature regarding aids and hindrances to learning, and what the goal of learning should be. For example, (Paik and Schraw 2012) studies the conditions under which animation helps students learn. They found that some animations hinder learning by encouraging students to pay less attention to the information, while other animations focus attention where it will improve recall. In other words, some forms of teaching invite mismanagement of cognitive resources, hindering the goal of giving students the ability to solve problems, while other improve it.

The issue here is reminiscent of the distinction between context of discovery and context of justification. In educational contexts, we are concerned with how one gets into a position to answer questions, in addition to whether answers already memorized are properly formed. It is an issue of cognitive management.

2.6 Long term memory

Cognitive management is also crucial to well-functioning memory. Two aspects of long term memory are closely tied to management: first, whether some information makes it into long term memory in the first place, and second whether it can be recalled later on. Without either of these functions, memory would be worthless. Without the ability to transfer some information to long term memory, we would never be able to learn from experience, and without the ability to recall stored information, we may as well have never stored it in the first place.

Given the amount of information we have available to us at any time, we cannot transfer all of it into long term memory. Some information must be ignored and forgotten. Since some facts are more likely to be relevant in future investigations than others, a well-functioning cognitive system ought to privilege information more likely to be useful over information that is not as likely to be useful.[12] To a large extent, whether information is retained in memory is not a matter of conscious choice, so this is often a subpersonal process. In assessing these processes, we must consider whether the cognitive system is organized in such a way as to preferentially retain information likely to be useful.

Information can be useful by being likely to be relevant in future investigations, or by fitting into current investigations. Investigation into a question can extend over time. To succeed in answering a difficult question that may take days or years of study to answer, one must be sure to store beliefs that represent genuine progress toward an answer. Keeping records can help, but one cannot at every step of the investigation consult all records of significant information. In order to continue, one needs long term memory of significant results. This involves both the ability to monitor the current level of progress, to prevent covering the same ground multiple times, and managing resources to make sure that the necessary information is retained.

[12] Just as an example of how this could be achieved, a cognitive system could be sensitive to cues that, in one's environment and given one's cognitive goals, tend to play a role in answering questions that tend to arise.

To get an idea of what is under evaluation, consider some empirical research in the psychology of long term memory. Many details are known about the conditions under which information is stored in long term memory. For example, devoting greater attention to a fact makes it more likely to be stored. Greater rehearsal makes long term recall more likely. On the other hand, interference from past memory can hinder storage. For example, memory of your previous address can hinder recall of your current address, as in (Kane and Engle 2000). These features of memory can be evaluated in how well they are conducive to success.

The standard suite of evaluations does not play a significant role in assessing how well one manages which information to store long term and which to forget. As for true beliefs: those items that get stored are already our beliefs, so their truth or falsity is not at issue. Nor will it help to have true beliefs about what to remember and what to forget, for almost all of the "decisions" about what to remember or forget are made automatically by the organization of the cognitive system, not by the conscious choice based on explicit beliefs. For similar reasons, knowledge and justification are not at issue either. One may know something, but the question is whether one should hold this knowledge or forget it.

Assessments of memory are also not exhausted by rationality. Regarding remembering, transferring a belief from short to long-term memory is perfectly truth-preserving, going from p to p. Since it is just a change in manner of storage and not a new belief, the transition from short to long term memory will not affect overall coherence. If one had beliefs about what one should remember, or what one should call to working memory, then it could be irrational to act contrary to those beliefs. However, the function of memory cannot be entirely guided by beliefs about what one should remember. One must call those beliefs to mind and act on them in order to make a difference. Further, no one has beliefs that cover every possible act of memory, so this form of rationality cannot cover all of the evaluations of long term memory.

Rationality does not capture all evaluations of forgetting, either. Losing a belief cannot

cause incoherence. Even if one considers forgetting to be irrational in itself,[13] this cannot account for the fact that some acts of forgetting are superior to others. Forgetting the arrangement of pencils on your desk is almost always better than forgetting where you are, but both involve forgetting.[14] If one has contradictory beliefs, forgetting one of them can make one's mental state rational, but again, this does not help us in cases where forgetting does not remove incoherence. We can still evaluate whether one act of forgetting is cognitively better than another even when no change in coherence results. Rationality does not cover memory evaluations.

After information is stored in memory, recalling stored information is the second essential memory function. The amount of information stored in memory vastly outstrips the amount of information we can work with at one time in working memory. Somehow, we need to be able to search out and call to mind the information that will help us succeed at the moment.

Suppose you are trying to identify a fish. It has one eye on each side of its head. You know that flounders have both eyes on one side of their head, but you do not recall that fact at the time. You cannot apply it to the task of figuring out what kind of fish this is. This is not a failure of belief or knowledge. Consider our responses in such cases. If someone asks, "Is there some feature of this fish that rules out flounder?" The reply to such a question is more likely to be, "I can't think of anything," rather than, "For all I know, it might be a flounder." We understand that the failure to recognize a significant feature, or the inability to think of a relevant fact, does not necessarily demonstrate that we do not know any relevant facts. It frequently occurs that we know something that distinguishes between hypotheses, but we can't bring it to mind at the moment. We can evaluate people in terms of how well they can bring to mind the right facts at the right time. This is a subject

[13]This is true of some versions of Bayesianism; see (Arntzenius 2003) for discussion of this.

[14]There will be exceptions where the arrangement of pencils is of paramount importance. Effective management is required to be able to recognize this, as well. This is closely related to the frame problem, (Dennett 1987). I do not claim to know the details of a solution to the frame problem, but somehow our minds do solve it and doing so is tied to cognitive management.

of study in psychology, what factors memory draws upon to recall beliefs.

A phenomenon illustrating how human minds manage recall is that it is easier to recall facts or words when one is in an environment similar to the environment in which one originally learned it. For example, in one study subjects were told to memorize lists of words while either on land or under water. There was a significant difference in recall depending on whether the test environment matched the learning environment (Godden and Baddeley 1975). This is the sort of phenomenon that is up for evaluation when human cognitive management is at issue.

Another example: when one item is called to mind, items similar to it are primed so that they are easier to call to mind. For example, if one is given a list of letter strings and tasked with sorting them into English words and non-words, seeing the word 'nurse' makes one's reaction time better on words like 'doctor' seen shortly after (Schacter, Wig, and Stevens 2007). These are ways in which our minds are organized to manage recall, and it is possible to evaluate how well these processes serve us.

This is not an issue we can address in terms of the standard suite. We are dealing with beliefs (often knowledge) that the agent already has. We do not have exhaustive beliefs about what to remember in any given situation, so rationality does not capture it either. The issue remains: when should a particular bit of knowledge be recalled? Evaluations of truth, knowledge, justification and rationality cannot allow us to evaluate the difference between calling the right belief to working memory and failing to do so.

2.7 Experts and degrees of belief

There are a variety of questions for which we consult experts for answers. The importance of expertise has been enshrined in the Expert Principle, the principle that given that the experts have credence x that A is true, one's own credence in A ought to be x. This treatment of experts, and the attractiveness of the principle, can be explained partly in terms of cognitive management.

So far I have not explicitly addressed the relationship between cognitive management and degrees of belief. Non-deductive reasoning is not monotonic; a conclusion supported by epistemically relevant considerations $C_1 \ldots C_n$ may not be supported by epistemically relevant considerations $C_1 \ldots C_{n+1}$. Some considerations weigh more heavily on some conclusions. For example, regarding the question how likely one is to die in a house, the presence of a raging fire in the house has a bigger impact than the color of the curtains. Due to our limitations, we cannot account for every possible bit of evidence there is, but our degrees of belief will be closer to optimal if we account for the most significant evidence first. A cognitive system that functions well will be more likely to attend to the more significant factors first. Thus, just as finding answers to specific questions takes management ability, so does approaching optimal degrees of belief.

To the issue at hand, the reason it might make sense to adopt an expert's credence in a proposition is because when they settle on a credence, it is likely to be close to optimal. There is less chance that their credence arose by failing to account for the most significant evidence. They have been able to work through the significance of evidence due to their skill in managing resources. Thus, they are more likely to settle on the ideal response. Their cognitive management maximizes the information extracted. That is part of the skill of being an expert. Thus, when their credence in a proposition is x, that is reason to think the chance the proposition is true is approximately x.

In our world, humans usually gain this ability through things like education, training, experience, etc. This has allowed them to develop a variety of skills to determine what is relevant, and to find answers within their chosen subject. Because of their training and ability, they are able to identify the evidence that makes the biggest impact on likelihoods of various hypotheses.

To take an example, to be an expert doctor requires more than just a detailed knowledge of medicine and related sciences. One must be able to manage that knowledge. In order to make diagnoses in time to help patients, doctors must have developed the skill of sorting

through all available information to find facts that allow for an accurate diagnosis. When we ask for a doctor's advice, we never hope merely for someone whose medical beliefs reliably turn out to be true. A doctor could be rational and knowledgeable yet only reach conclusions about unimportant matters. We want someone who can answer quite specific medical questions, such as "Why is my arm numb, and what can I do about it?" We want a reasonable assessment of the chances of diagnoses being correct and reasonable assessments of the chances of treatments working. Being able to do this requires efficient cognitive management, in addition to knowledge and rationality.[15]

Some of the differences between expert and non-expert cognition are known. One advantage that experts have in answering questions is that they have the ability to recognize patterns that lay people do not recognize. For example a chess master can recognize the state of a game in progress as exemplifying a pattern without memorizing each piece's position separately. Thus, while working memory capacity is not increased, their means of storing information is more efficient. The chess master can see and remember the board as a few clear patterns, whereas I can only see it as individual pieces (with some exceptions, like standard game starting position).

Experts notice what is more important, and thus can remember important details more reliably. An expert will be used to checking for certain kinds of common and easy mistakes, whereas a non-expert, even if they know of those mistakes, will be less adept at watching for them (especially if there are several kinds). An expert has developed the skill of noticing which pieces of information are most relevant, most diagnostic, and most likely to yield a conclusion that would not be overturned by the ignored data. This is important, since one must always ignore some of the available data. Each skill conspires with the others to rig a conclusion that is likely to be true. This is why we do not go to a trivia master, or a well-informed amateur with our questions. A trivia master's long list of memorized facts need not give them the ability to use those facts effectively, whereas the expert has the advantage

[15] I am ignoring oracles who magically have the answers and doctors with infinite mental capacity, neither of which exist anyway.

in this respect. Cognitive management is one of the key defining traits of the expert.

2.8 Is cognitive management epistemic?

I have claimed throughout that evaluation of cognitive management is a form of epistemic evaluation. However, one might object that it is practical rather than epistemic. In many examples presented above, the specific question that the agent was trying to answer was singled out for practical reasons. Further, the idea that there are some questions that one *ought to* answer appears to be a distinctly practical notion, a matter of deciding what investigations to undertake in order to further one's ends. While cognitive management would still be important if it were practical, I here argue that it is epistemic.

The reason to consider cognitive management epistemic is that inquiry is the paradigm of epistemic activity. What it is to be an epistemic agent is to engage in inquiry. In turn, a fundamental part of engaging in inquiry—the posing, investigating and answering (or forming credences in answers to) specific questions—is carried out only due to effective management of cognitive resources. If a being lacked the ability to manage cognitive resources, it is unclear whether such a being could be an epistemic agent; it would not direct its epistemic life. Someone who cannot manage cognitive resources cannot answer particular questions, and without the ability to do so they do cannot set out to generate a picture of the world they live in; they just acquire beliefs piecemeal. Cognitive management is integral to the constitutive epistemic activity, inquiry, which is reason to consider evaluation of cognitive management to be epistemic.[16]

Further, the references to practical concerns when illustrating cognitive management do not make cognitive management itself practical, for appeal to practical factors is ubiquitous in epistemology. Epistemic evaluation unites with the practical because practical decision-making depends on epistemic success. The fact that X can make a pragmatic difference does not prove that X itself is pragmatic. If this were a sound inference, then belief,

[16]This evaluation specifically relates to how well one works with the cognitive limitations due to our finite nature.

knowledge, justification and rationality would all be pragmatic concepts.

Finally, accepting that management is epistemic does not imply that practical reasons are reasons for belief. In fact, the necessity of the ability to answer specific questions, and the management that this requires, implies nothing about how questions guide inquiry. Nothing implies that specific questions can only be singled out due to practical considerations. It might be that some questions bring one closer to that which is intrinsically epistemically valuable, for intrinsically epistemic reasons. The selection of questions does not imply that management is practical.

There is good reason overall, therefore, to classify the evaluation of cognitive management as a kind of epistemic evaluation. It is the central aspect of the defining epistemic activity, inquiry.

2.9 Conclusion and a way forward

I have shown that there is an epistemic evaluation, the evaluation of cognitive management, that is not reducible to any combination of evaluations in the standard suite. To end, I propose that we are justified in considering cognitive management to be a form of *understanding*, though perhaps a semi-technical precisification of the ordinary concept of understanding. The identification of understanding with cognitive management is justified due to several ways in which management aligns with common usage of understanding.[17]

First, the identification is justified because the roles of understanding and cognitive management in teaching and learning match. As argued above, improving students' cognitive management abilities is a crucial part of teaching. At the same time, publications about teaching frequently emphasize that understanding is the most important goal. In addition the claims made about understanding in these publications closely resemble the above discussion of cognitive management. For example, the story from (Skemp 2006)—cited because it shows the importance of cognitive management in teaching—is taken from

[17] I will not here go so far as to argue that this is a complete analysis of the ordinary concept of understanding. That is for the next chapter.

an article extolling the importance of understanding in teaching. Their roles align.

Also, the identification is justified because the roles of understanding and cognitive management match with respect to expertise. As argued above, one of the important reasons to accept experts' judgments is that they have superior cognitive management skills. This is what allows them to discern what is relevant. A generally accepted defining feature of an expert is that they understand the subject in which they are an expert, and that understanding is what allows them to discern what is relevant. This is another reason to identify understanding with cognitive management.

In addition, cognitive management is in part the ability to use knowledge. It is the ability to sort the relevant from the irrelevant. These are also marks of understanding; one who understand does not merely have knowledge but is able to apply it effectively. Understanding is frequently measured by the ability to answer novel questions without groping, without going around in circles. This is exactly what cognitive management is. Again, the two align.

Finally, both understanding and cognitive management ability come in degrees. One can understand better or worse. The evaluation of cognitive management likewise can be better or worse. Thus, they are structurally a match.

For all of these reasons, it makes sense to consider cognitive management to be a form of understanding. Thus, the above arguments demonstrate the importance of understanding in epistemology. In conclusion, then, we have shown that epistemology ought to include the study of at least this form of understanding, for it is a necessary component of epistemic agency and it is not reducible to any of the standard epistemic evaluations. This will be explored further in the next chapter.

CHAPTER 3

UNDERSTANDING: THE ART OF COGNITIVE MANAGEMENT

3.1 Introduction

Understanding is often considered by teachers, scientists, lay people, and philosophers to be a necessary part of intelligence. If one does not understand, one necessarily falls short of epistemic excellence. However, there is no ready to hand explanation for why understanding is so important. We can explain why knowledge is a necessary part of epistemic excellence: knowledge means getting things right in the right way. It satisfies our need for information about the world.[1] We can likewise explain why rationality is a necessary part of epistemic excellence: rationality means reaching the best conclusion given the information available. We want to know what the best conclusion is with the available information, and so we seek an account of rationality. On the other hand, understanding has until recently been neglected in philosophy, in large part because the need it satisfies is not so easy to express. Understanding something means that you "get it" or you "grasp" something, but it is not clear exactly what this means or why it is important.

Still, we have enough of a working knowledge of understanding to tell that it is important. It is easy to find articles and speeches about the importance of understanding. This has spurred the creation of several accounts of understanding, including (Kvanvig 2003),

[1] At least, this thought was obvious enough to inspire centuries of work on knowledge. In the last couple of decades, *the value problem* has questioned whether knowledge is more valuable than its constituent parts. What matters for our purposes is that the need knowledge was thought to satisfy was easily articulable, though later called into question by some.

(de Regt and Dieks 2005), (Elgin 2007), (Boylu 2010), (Wilkenfeld 2013), and more.[2] Unfortunately, none of these accounts goes far enough in clearly capturing and explaining the value and unity of understanding.

In this chapter I address this issue by defending a novel account of understanding with three valuable features: (1) it shows that understanding is necessary for epistemic success, (2) it provides clear guidance for future research, and (3) it unites different forms of understanding (e.g. understanding-why, understanding-that, understanding a person, and understanding an object or work of art). The first of these features of my account shows that understanding is in fact essential and deserves closer study in epistemology. The second of these features will help turn the study of understanding into a more productive research program. The third feature allows us to clarify the essential features of the various forms of understanding and accounts of understanding.

The account defended in this chapter is the *management account* of understanding. The management account says that understanding is the ability to solve a version of the epistemological frame problem that I call the cognitive management problem. Briefly put, understanding is the ability to extract and exploit the relevant cognitive resources to answer questions. This account achieves (1) because without the ability to find what is relevant, we could not be cognitive agents. Thus, the account shows that understanding is necessary for epistemic success. The account achieves (2) because it relates understanding to the ongoing research program of the frame problem, so we know where future research should be directed (even though this is a difficult problem). This account can also, as per (3), be extended to many forms of understanding. Given this account, existing accounts of understanding can be seen as attempts to provide more or less detailed hypotheses of how we manage to solve the cognitive management problem.

In this chapter, in Section 3.2, I present the cognitive management problem and explain its significance to epistemology. In section 3.3, I present the management account of un-

[2]See Section 3.5 for a more in depth survey of existing accounts of understanding.

derstanding in detail. In section 3.4, I show that the management account can explain a variety of facts about understanding. Understanding is the ability to solve the cognitive management problem. In section 3.5, I address concerns about the account. Finally, in section 3.6, I show that all major existing accounts of understanding can be seen as attempts to give more specific versions of the general account I present and defend.

3.2 The cognitive management problem

There is a problem any cognitive system must solve in order to be effective: it must find what is relevant to answer specific questions, both what information is relevant and what cognitive operations are relevant. Without the ability to find what is relevant, it is impossible to answer specific questions. Since inquiry is the process of answering questions, the inability to find what is relevant renders one incapable of successful inquiry. Somehow, we must find what is relevant. This is a problem I will call the *cognitive management problem*. To solve it is to manage one's cognitive resources.

The cognitive management problem arises from the fact that there is an array of information available to the mind in perception and memory. However, this information needs to be extracted before it can take part in problem-solving.[3] Whether extraction involves recalling a sentence in the language of thought from the belief box to working memory, or transducing perceptual signals, or employing a properly-tuned connectionist network, or whether it involves some other form of organization, the information available does not help answer questions if it is not extracted for use by the cognitive system.

The problem does not end there. It is possible to extract the relevant information to answer a question but fail to exploit it properly. Cognitive systems exploit information to reach conclusions. This may be through classical inference, or through other means. But if the right exploiting moves are not undertaken, the problem is not solved. If you extract

[3]Here, 'extraction' refers to the cognitive activity of the cognitive system to extract information that is already available to the mind in perception or memory. Looking up a fact in a book or performing an experiment is not the relevant kind of extraction.

the information that all men are mortal and that Socrates is a man, but do not perform modus ponens, you won't reach the conclusion that Socrates is mortal. We must find the relevant operations to perform on the relevant information. Without doing so, inquiry is not possible.

The extent of the cognitive management problem becomes apparent when one considers how much information is available and how much could be done to exploit it. Given how much we believe, how many inferences we could make from those beliefs, and how much information is implicit in perception, finding just the right information at any given time is difficult. For example, I know the rules of chess, but I frequently suffer unexpected defeat by my opponent's move. The possibility of the surprising move is implied by the rules and previous positions, but I did not extract that information to infer the potential for loss. Another example, there is an English word that ends in 'mt.' We all know this word, but most people cannot think of it even given the clear description. The word is 'dreamt.' Despite how common this word is and the fact that it obviously ends in 'mt,' it just does not come to mind. This information was available to us all along, but extracting it is another matter.

The problem cannot be solved by including beliefs about what is relevant in a situation (e.g. "If the question is how many cats there are, the color of the moon is not relevant"), for that only compounds the problem. Either these beliefs are among the other information we are trying to search for the relevant information, or they are partitioned off from the rest in a kind of meta-memory. If they are among the other information, then they only add to the problem of finding what is relevant, for now one must find the relevant meta-beliefs to determine which beliefs are relevant. If the meta-beliefs are partitioned off, the same problem will arise within meta-memory, and given the number of situations we would need meta-beliefs about, there is no reason to think it would be easier to solve the problem at the meta level. Thus, the solution must include some processing at a subpersonal, unconscious level. The processes that realize our cognitive activity must work in such a way to solve

the problem.

Solving the problem can be a matter of degree, for one can be able to manage one's cognitive resources over a wider or narrower range of circumstances. One who can answer a broader range of questions from a broader range of starting information has greater cognitive management abilities. If one is able to manage one's resources to answer central questions, one is better off than one who can only answer peripheral questions (see section 3.3.1). On the low end of management ability one can only extract and exploit the relevant information by chance, while on the high end one is omniscient with the ability to draw every conclusion instantly. To function as a cognitive system, one must be able to solve the problem to a sufficient degree.

Note that this is a version of the epistemological frame problem, which Dennett describes by noting, "Even if you have excellent knowledge [...] about the changing world, how can this knowledge be represented so that it can be efficaciously brought to bear?" (Dennett 1987, p. 442). The broadest statement of the problem is from Glymour: "Given an enormous amount of stuff, and some task to be done using some of the stuff, what is the relevant stuff for the task?" (Glymour 1987, p. 65).[4] In our case, the stuff is cognitive resources (cognitive operations and information available to the cognitive system), and the task is to find the answers to questions.

Recognizing that the cognitive management problem is a version of the epistemological frame problem highlights the difficulty of solving it, for it is not yet fully known how our minds solve the frame problem. While it is clear that we often can find what is relevant, it is not clear how.[5] The fact that it is so often obvious to us what is relevant hides the fact

[4]The original frame problem from (McCarthy and Hayes 1969) dealt with issues like: how can one make an artificial intelligence that knows that when it closes the door the roof does not change color, and do this without explicitly including the belief that door closing does not affect roof color? An AI is built from scratch, and does not know which changes are relevant to which other changes. Philosophers have since identified related problems that apply beyond updating in response to change.

[5]That is, for limited minds like ours it is not known how; infinite minds can solve the problem by exhaustive search. Some have argued that connectionist networks avoid some versions of the frame problem, but they do not avoid this broad version, for example, see (Dreyfus and Dreyfus 1987), (Dennett 1987). (Xu and Wang 2012) claim to have solved the epistemological frame problem, but I will not here assess whether they have.

that the cognitive management problems needs solving, lest we be incapable of inquiry.

3.3 Understanding is cognitive management

My proposal in this paper is that understanding is the ability to solve the cognitive management problem. Thus, understanding is essential for epistemic success, for without it we would not be able to engage in inquiry.

First, we must address the fact that understanding is understanding *something*, such as understanding statistics or understanding your parents. I will begin with understanding a subject matter, and then use this to give accounts of other forms of understanding. I will follow (Lewis 1988) and assume that a subject matter is a set of questions. Roughly, the subject matter of X is the question "What is the full truth about X?" along with questions whose answers partially answer this. The subject matter of physics includes the question "What is the whole truth about physics?" as well as questions like, "What are the laws of physics?" and other subquestions about physics.[6] The account of understanding is:

The Management Account Understanding a subject matter is the ability to solve the cognitive management problem in that subject matter, i.e., to extract the relevant information and exploit it using the relevant cognitive operations to answer questions within the subject matter.

If one has the ability to solve the cognitive management problem in a subject matter, then one can find the needed information and capacities to answer questions and engage in inquiry within that subject matter. Understanding is important because without it, the information available to us would be all but useless. Understanding allows us to use our cognitive resources effectively. Given this account of understanding a subject matter, we can give accounts of other forms of understanding.

[6]This formulation depends on a notion of aboutness that does not imply that 'Physics is such that Harry loves Ginny" is about physics. If such an account is impossible, the limits of a subject matter will then rely on a contextual notion of aboutness, similar to what is described in section 3.3.1.

Understanding why: Understanding why p singles out the question why p. This defines a subject matter that consists of the question why p, and related related questions that make up the whole truth about why p. Thus, we can extend the account of understanding a subject matter to an account of understanding why: one understands why p when one has the ability to manage cognitive resources in order to answer questions related to why p. Note that this requires more than merely knowing why p, for in addition to knowing why p, one must have the ability oneself to identify what is relevant to answer questions within that subject matter, not just the ability to recite an explanation. Understanding is a matter of cognitive management ability, not simply having the answer.

Understanding an object/person: A person or object also forms a subject matter, summarized by the question "What is the whole truth about this person/object?" There are questions that relate to the person or object, and understanding requires being able to answer those questions. Essentially, understanding a person or object just is understanding the subject matter of the person or object. There are questions about the objects, and if one can answer those then one understands the object.

Understanding a theory: Understanding a theory is the same as understanding an object, it is just that it is an abstract object. This allows us to give a single account of understanding a true theory and understanding a false theory. Understanding a theory means having the ability to answer questions about the theory, such as what it implies, whether it is an accurate or inaccurate theory. Understanding a false theory like phlogiston theory means being able to answer questions about the theory, such as what it implies or whether it is true. Note in particular that this is distinct from understanding a phenomenon that a theory is about. Understanding phlogiston theory does not imply understanding combustion, which phlogiston theory is a theory of.

Understanding-that: The fact that p also defines a subject matter with the question: What is the whole truth about the fact that p? The whole truth about the fact that p is how it relates to everything else. Consider that one makes the claim that one understands that

p to express that one recognizes how p relates to the current situation. One says that one understands that the door is open in order to express that one realizes how the fact that the door is open relates to what one is doing. In other words, there are questions that, in context, are related to whether the door is open. This includes whether there will be a breeze, whether the air conditioner will be effective, etc. The questions raised by the truth of the proposition then define a subject matter. Understanding that a proposition is true requires one to be able to answer these questions.

Starting from an account of understanding a subject matter, we have accounts of different forms of understanding. The management account unites the different forms of understanding; they are all fundamentally a matter of solving the cognitive management problem. Before moving on, let us make explicit some implicit features of the account.

3.3.1 Degrees and context

The management account correctly implies that understanding comes in degrees. Since the ability to solve the cognitive management problem is a matter of degree, the management account allows for comparing degrees of understanding by comparing degrees of cognitive management ability. The better one can manage cognitive resources, the better one understands. The Nobel laureate understands physics better than the physics graduate student. Good parents understand their child better than the child's teachers do. Further, these determinations can vary with context.

First, some questions within a subject matter are more central than others, depending on context. In theoretical contexts, questions about the general laws of nature can be more salient to understanding Newtonian Mechanics, while in experimental contexts, questions about specific predictions may be more salient. The ability to answer more central questions in a context leads to better understanding. Level of centrality also gives us a weighting to determine one's degree of understanding based on the range of questions one is able to

answer.[7] The variability of centrality means that the weighting also varies with context.

Given a contextually appropriate weighting, greater reliability in answering questions in a subject matter means greater understanding.[8] The more reliably one can find the answer to a question in a subject matter, the better one's understanding. There can be a trade-off between range/centrality and reliability. Someone who memorizes the central facts about a topic can answer those reliably, while someone who relies on skillfully working out the answers as they go will be less reliable over the central questions, but can answer questions she has not memorized. Lower reliability over a wider range may imply better understanding. (I will not here attempt to detail the exact nature of the trade-off, just note that there is one.)

Finally, in addition to comparisons of degree, there is all-or-nothing understanding. As with other graded predicates that have an absolute form—like 'tall' and 'flat'—there is a threshold in degree of understanding above which an agent counts as having all-or-nothing understanding. Also, like 'flat' or 'tall,' the threshold depends on the context. Just as a surface can be flat for the surface of a road but not for surface of an air hockey table, one's degree of understanding can be enough to count as understanding physics in a high school class setting, but not in the setting of researching fundamental particles. The threshold degree of understanding an agent must meet before she counts as understanding depends on context.

With the nature of context-dependence, we now have an account of understanding in hand. Let us see what this account can do for us.

[7] Level of centrality can be represented as a measure defined over sets of questions. Sets with more central questions will be assigned a higher value.

[8] It is not always a straightforward matter to determine exactly what reliability means, see e.g. (Alston 1995) and (Conee and Feldman 1998). However, without asking how, we will assume that a measure of reliability is possible.

3.4 The work of understanding

The management account matches everyday expectations about what understanding entails. We generally expect that someone who understands can find the relevant among the irrelevant. Consider how strange it sounds to say, "She understands, though she can't tell what is relevant." We expect that understanding allows one to operate effectively with the topic that is understood. One of the marks of someone who understands is their ability to cut through to the heart of the matter. Someone who understands cars, for example, can quickly identify what factors are relevant for finding what is wrong with a car. These generalizations about understanding all describe cognitive management. Cognitive management is the ability to find what is relevant. Being able to find and exploit what is relevant allows one to operate effectively within a topic.

In addition, a sign that one lacks understanding is that one spends time focused on the wrong area, floundering in irrelevancies. Someone who does not understand a topic will insist on irrelevant details. For example, those who do not understand photography will pour over photographs of Mars, marveling at straight lines and what look like faces. They do not recognize that lines and the vague "face-iness" of a land feature is not relevant to identifying what the feature is. Someone who does not understand may have the relevant information available but focus on the irrelevant. We desire understanding precisely because it prevents us from falling into such traps. These descriptions of lack of understanding describe lack of cognitive management ability. Focusing on irrelevance is a failure to manage resources.

The management account explains these features of understanding. It provides a good explanation of others, as well.

3.4.1 Education

Understanding is not memorization. One cannot understand a subject just by memorizing a long list of facts about the subject. One cannot understand a person by memorizing

her biography. One needs to show a greater familiarity of how the facts fit together. This feature of understanding is well explained by the management account. One cannot manage resources simply because one has a great deal of knowledge. The difficulty of finding the relevant knowledge *is* (part of) the cognitive management problem, so the possession of knowledge will not solve it. In fact, a greater amount of knowledge can in some ways make matters more difficult, because there is a greater amount of information to sort through before finding the relevant information. Thus, the management account offers an explanation for why understanding is more than memorization.

Similarly, one cannot gain understanding by learning rote procedures. If one can solve problems only in virtue of the fact that one has memorized a procedure that works given specific inputs, this does not demonstrate a high degree of understanding. For example, there is a formula to calculate the orbital speed of a of a body in circular orbit around a central mass, given the orbit radius and central mass, and assuming the orbiting body has negligible mass. Memorizing this without being able to derive it from Newton's laws does not demonstrate understanding. The management account explains why. One cannot solve the cognitive management problem with rote problem-solving procedures. First of all, one still must be able to employ the right memorized procedure; the fact that one has memorized it does not guarantee this. Thus, one still needs to realize that it is the relevant procedure, which means solving the cognitive management problem.

Furthermore, the procedures we are willing to call rote procedures work only given specific circumstances and inputs, and tend to be useless otherwise. Consider again the formula for orbital speed. If the orbit is not circular, the formula is useless. If the orbiting body has mass comparable to the central body, it does not work. Rote procedures have limited application, which means that to achieve a wider range of success, one needs to memorize many. To answer questions given a battery of rote procedures, one must be able to find the relevant one. The ability to do this is cognitive management. Finally, rote procedures cannot extend to novel questions. If one can manage cognitive resources, one

can find relevant information and inferences and answer novel questions, whereas reliance on memorized procedures will leave one powerless. The inability of rote procedures to solve the cognitive management problem explains why they cannot provide understanding.[9]

The fact that understanding is more than memorization leads us to the importance of understanding in education. Understanding is the ideal goal of education. Education should not consist simply of memorization. When teachers lament that students do not understand, the issue is not that students have not memorized enough. The goal of providing understanding is that students will leave with the ability to think for themselves regarding the subject.[10] This is why tests, as much as possible, present students with novel questions to answer, forcing students to demonstrate their understanding by answering novel questions.

This is explained by the management account. In order to answer novel questions, one must be able to identify what is relevant to answer the question. One must be able to determine what are the relevant inferences to make from the relevant information. Doing this successfully just is managing one's cognitive resources. Thus, the management account explains the role of understanding in education.

3.4.2 Expertise

Experts understand, and this is why we seek them for answers. Their understanding means not only that they have knowledge, but that they can use that knowledge. A trivia buff can answer from a long list of memorized answers, but for questions not on the list, they have little to offer. In contrast, an expert understands the subject, which makes her adept

[9]Likely, at some level of analysis, our minds operate using procedures just as robotically as one follows a rote procedure. However, a procedure is called a "rote procedure" in virtue of being limited in the ways described, not in virtue of being an algorithm. If our minds operate using algorithms, they are not rote procedures.

[10]Resources to help teachers teach for understanding repeat this frequently. To take an example at random, the website Teaching for Understanding says, "the only way to know with certainty how much students understand is to ask them to carry out some task that requires them to go beyond what you have told them or what they have read in a textbook," learnweb.harvard.edu/alps/tfu/info3e.cfm. Also see (Wiggins and McTighe 2005).

at looking at a body of information, finding what is relevant, synthesizing it and reaching a conclusion that reliably fits the evidence. Give the non-expert enough information to answer the question, and she won't necessarily be able to exploit the information to find that answer. For example, imagine providing a non-expert with a paper about a medical experiment, along with detailed descriptions of the kinds of errors that can occur in experiments. Now suppose we ask her whether the experiment is flawed. Even if she has memorized the kinds of errors, since she is not an expert she will likely struggle with the task, going line by line and trying to remember kinds of errors that might have occurred on any given line. In contrast, the expert who understands medical experiments will likely recognize the error on sight, for she is able to make the relevant identifications, extracting the relevant information and exploiting it, by skill. That is the benefit of understanding that experts have. This is captured by the management account, for the ability to immediately infer which type of error occurs requires extracting and exploiting the relevant information, which is what separates the expert from the novice.

Experts notice what is more important, and thus can remember important details more reliably. An expert will be used to checking for certain kinds of common and easy mistakes, whereas a non-expert, even if they know of those mistakes, will be less adept at watching for them (especially if there are several kinds). An expert has developed the skill of noticing which pieces of information are most relevant, most diagnostic, and most likely to yield a conclusion that would not be overturned by the ignored data. This is why we do not go to a trivia master or a well-informed amateur when we are trying to extend human knowledge. Their long list of memorized facts need not give them the ability answer questions that have not yet been answered. Give them all of the relevant data, and they will not know what to do with it.

The expert has the advantage in this respect due to her understanding. An expert is skilled at determining what data is relevant, and she is skilled at figuring out what to do with that data. This is what gives her the advantage, and it is why we seek experts to

answered unanswered questions. The ability to do the relevant thing with the relevant data is cognitive management. That which makes an expert superior in understanding is the ability to manage cognitive resources. The management account explains the role of understanding in expertise.

3.4.3 Degrees of belief

Another characteristic of understanding is that, given a body of evidence and a hypothesis, one who understands will reliably adopt the optimal (or near optimal) degree of belief in the hypothesis.[11] Someone who does not understand will be much more likely to have degrees of belief too high or too low. What accounts for this difference? First, non-deductive reasoning is not monotonic; a conclusion supported by epistemically relevant considerations $C_1 \ldots C_n$ may not be supported by epistemically relevant considerations $C_1 \ldots C_{n+1}$. The addition of new evidence can completely change the supported conclusion (e.g. when finding out that someone you thought was a friend is actually an enemy spy). Furthermore, some considerations weigh more heavily on some conclusions. For example, regarding the question how likely one is to die in a house, the presence of a raging fire in the house has a greater impact than the color of the curtains. Due to our limitations, we cannot account for all of the evidence there is, but our degrees of belief will tend to be closer to optimal if we account for the most significant evidence first. This will decrease the chances that evidence not yet considered will change the supported conclusion.

In a situation where we are dealing with degrees of belief and chances, relevance also comes in degrees. The factors that have the greatest impact on the rational degree of belief are those that are most relevant to the question. Thus, the ability to account for the evidence that has the greatest impact on the conclusion is the ability to find what is most relevant. Therefore, the ability to approximate the optimal response to a large body of evidence

[11] This leaves open whether the evidence makes just one degree of belief optimal, or whether whether the optimal response depends on the agent (as subjective Bayesianism implies, for example), or whether multiple degrees of belief are optimal. One who understands is still expected nearly optimal.

is a part of the ability to manage cognitive resources. One who understands thus forms near optimal degrees of belief through cognitive management, by focusing on what is most significant first. Again, the management account explains this feature of understanding: she who understands is better at forming optimal degrees of belief by finding the most relevant considerations.

3.4.4 Explanation

The management account also makes sense of the close relationship between understanding and good explanations.[12] One who understands is more likely to know an explanation, or is better able to produce explanations, or has more ready access to explanations. Nearly every account of explanation claims that having an explanation provides understanding. Kitcher is representative in saying, "A theory of explanation should show us how scientific explanation advances our understanding," (Kitcher 1981, p. 508).

First, consider the kinds of explanations we offer and accept in practice, both in everyday life and in research. A good explanation only cites what is relevant; any explanation is made worse by the inclusion of irrelevancies. If you ask why the door is closed, and I tell you that there was an election in France and the wind blew it shut; that is worse than the explanation that the wind blew it shut. The election in France is irrelevant, and its inclusion in the explanation suggests that somehow it is relevant, undermining the explanation. Or consider the famous example of the male who takes birth control pills and is not pregnant. A good explanation for why he is not pregnant is that he is male. A bad explanation is that he took birth control pills, for given that he is male this is irrelevant. Prominent accounts of explanation assert that explanations include only relevant information—from causal accounts such as (Strevens 2009) and (Woodward 2004), to unification accounts like (Kitcher 1989), to pragmatic accounts such as (Achinstein 1983).

The management account captures this by identifying understanding with the ability

[12] See Section 3.6.1 for more discussion of what other philosophers have said about the connection between understanding and explanation. See chapter 6 for more in-depth discussion of explanation.

to find what is relevant. The ability to produce good explanations is in part the ability to find what is relevant, which in turn is understanding. The ability to recognize good explanations is in part the ability to recognize which facts are relevant to a given explanatory why-question, which again is understanding. If one knows a good explanation, then one knows what is relevant. Each of these relate explanations to understanding according to the management account.

The above considerations apply to accounts that restrict explanations to relevant factors, but some accounts of explanation imply that ideal explanations are not constrained by relevance. For example, some accounts claim that the ideal explanation of an event cites the entire causal history of the event, as in (Lewis 1986).[13] This might appear to be a problem for the management account, since without any emphasis on relevance the management account cannot explain why explanations would have anything to do with understanding. However, humans never know these ideal explanations or offer them as answers to why-questions; we only deal with truncated explanations that are constrained by relevance. Since understanding is a feature of an agent's psychology, it is only these latter explanations that relate to understanding. Thus, proponents of accounts with no relevance constraints claim that the explanations we actually deal with are partial explanations that are useful given our background. Once we have taken this step of admitting that we deal with relevant portions of ideal explanations, the management account can again explain the connection between understanding and (partial) explanations. Partial explanation that we actually use again appeal only to that which is relevant, and the connection between this kind of explanation and cognitive management is secured.

The above considerations show how the management account brings together and explains many of the features of understanding. This gives us a sense of how the management account provides a coherent, working picture of what understanding is and what role it has to play in our cognitive lives. Now we must consider some concerns about the management

[13] Such accounts still rely on a limited form of relevance, for only causally relevant facts are admitted.

account and whether it is an adequate account of understanding.

3.5 Simple-minded understanding?

One concern is that the account above implies that one can understand using simple-minded methods that intuitively cannot provide understanding. Consider a machine, call it Mechano, that has every question about a subject paired with its answer in a simple input-output mechanism. Given a question, Mechano can give the answer. It appears Mechano is thus able to solve the cognitive management problem through exhaustive input-output pairing. However, it also appears that Mechano's method of answering questions is not understanding.

First, a clarification. In order for Mechano to solve the cognitive management problem, it must be a genuine cognitive agent with genuine mental states. One cannot find the relevant cognitive resources if one does not have cognitive resources. The management account does not say that if Mechano makes sounds we can interpret as an answer, then Mechano understands. Mechano must actually be able to answer the question through genuine mental activity. The list look-up procedure describes how Mechano's mental abilities are realized. If such a procedure does not qualify as mental activity, then the Mechano example is not a counterexample to the management account. The management account is not providing a reductive account of the mental.[14]

Given this, the task Mechano performs cannot be simply to look up the question and spit out the answer. The questions we ask include 'How does this work?' and, 'Why is it doing that?' Even if Mechano has all the answers to all questions memorized, semantically speaking, it still has to determine which question is being asked, which is not uniquely determined by the sentence expressing the question. For example, Mechano has to recognize that the answer to whether it is raining depends on the relevant location, and since location is not explicitly stated, Mechano has to determine which location. If it can only provide

[14]Though, if an extreme behaviorist account of the mind is true, the considerations in the rest of the section still apply. Mechano is still not a counterexample.

the answer when the question is expressed in a way that avoids all possible ambiguity, that does not look like understanding, and it is not effective cognitive management. Finding the relevant question and answer pair manifests some level of sophistication, so it is not as absurd as it may seem initially to conclude that Mechano understands.

There is no clear reason to reject the claim that Mechano does understand. First of all, the ability to call forth the right answer from a database that includes every answer to every possible question is not a simple task. Mechano requires infinite mental capacity to solve it. The fact that we can easily describe the subpersonal activity of Mechano's mind does not undermine that.

Further, Mechano is not an example of understanding purely by memorization; Mechano understands based on memorization combined with infinite search capacity. Humans cannot acquire understanding through memorization because it only allows us to answer a limited range of questions. Once we encounter a question that is not memorized, we have no way to answer it. Contrast this with Mechano, which by hypothesis knows all answers and can find them infinitely quickly. Not only can Mechano answer how long it takes the moon to go around the Earth, it can answer why it takes that long, which experiments could confirm this and why, and so forth. In normal circumstances, memorization fails to provide understanding because it gives no depth to one's ability to answer questions; one can sometimes give the right answer but cannot explain why it is the right answer. In contrast, Mechano can give the answer and tell you why this is the right answer. Its knowledge is deep and it can provide all of the information that one who understands can provide.

Finally, consider: What is the maximum degree of understanding? The maximum degree of understanding comes from having infinite cognitive capacity and omniscience. How would a being with maximum understanding answer questions? It is omniscient, so it will already know the answer. It has infinite capacity, so it can find the answer immediately. Capacities other than search for known answers will be idle. Its operation is almost the same as Mechano's, and yet it fits the intuitive idea of perfect understanding. It also has perfect

understanding according to the management account, since the perfect being has the ability to perform every applicable operation to all information at once, and thus a fortiori applies the relevant operations to the relevant information. It solves the cognitive management problem by its infinite capacity.

This illustrates the fact that the appearance that one's mind operates without intelligence does not equate to lack of understanding. In some cases, the appearance is due to the possession of such prodigious mental capacities that simple problem-solving methods are effective. We humans use complicated systems of heuristics and computations to achieve some measure of success, but that does not imply that such complexity is a necessary part of understanding. It is just how we do it. There is good reason to think that a properly understood version of Mechano teaches the same lesson, and thus does not undermine the management account.

3.5.1 The boundaries of the mind

A further natural question is whether external devices can take part in effective management. For example, many people carry smartphones that have access to modern search engines. Consider a time in the future when search engines are even more effective, and can answer almost any question humans can ask them. Does someone have good cognitive management skills if they always turn to their smartphone to find answers? Does this mean that they understand by having a smartphone? After all, the discussion of knowledge-producing actions—including looking up answers—has long been a part of the literature on the frame problem, of which the cognitive management problem is a species.[15] Intuitively it appears that one who can only answer questions by using a phone does not understand. Yet it looks like the management account implies that they do understand.

To deal with this concern, consider two possibilities: either the use of an external device is a cognitive activity (as in the extended mind theory), or it is not.

[15]For example see (Moore 1985).

If the use of an external device is genuine cognitive activity, then it is not problematic that someone can understand in virtue of the use of an external device. It is no objection to point out that one's mind can play a role in understanding. If some kind of extended mind theory is true, the location and construction of the mechanisms employed do not make a difference. Thus, if the use of such devices is cognitive, the objection fails; it is no counterexample.

On the other hand, if the use of an external device is not cognitive, then we have a principled reason to reject it as part of understanding. Understanding is the ability to manage cognitive resources; it is a cognitive ability. It has to do with the function of the mind. As long as there is a distinction between mental activity and non-mental action, then the use of a device that is external to the mind is non-mental action. Once one needs to employ such a device, one is no longer engaged in cognitive management, one is managing other resources. These activities may allow one to gain new understanding, but they themselves are not characteristic of or constitutive of understanding.

3.6 Implicit in existing accounts

So far we have seen that (1) since understanding is the ability to solve the cognitive management problem, it is necessary for epistemic excellence. (2) The fact that understanding is the ability to solve the cognitive management problem tells us how to discover the details of understanding. (3) The management account can be extended to give a unified account of many forms of understanding. Now, I will show that no other account can do all of this. Accounts either do not demonstrate the necessity of understanding, or they are too vague, or they are restricted to only some forms of understanding.

In addition, the management account can explain what is attractive and what is inadequate about existing accounts of understanding. We can view existing accounts of understanding as attempts to provide more specific versions of the management account. Whenever an author working on understanding asks what leads to understanding, the answer they

propose at least partly addresses the cognitive management problem. Existing accounts of understanding try to explain how we overcome the cognitive management problem.[16]

3.6.1 Explanation accounts

Many philosophers have defended the view that S understands why p if and only if S bears the right relation to a true explanation of why p. Different accounts identify different "right relations" to an explanation, and different accounts of explanation produce different accounts of understanding. Yet the appeal of each of these accounts is explained by the management account.

First, as shown above, the management account captures the close relationship between explanation and understanding. According to the management account, understanding is a matter of finding that which is relevant. Since explanations single out what is relevant the management account explains why an account that ties in with explanation would be appealing.

In addition, the simplest explanation-based account of understanding is that S understands why p iff S knows a true explanation of why p. According to the management account, understanding why p is having the ability to manage cognitive resources in order to answer the question why p. The answer to many why questions is an explanation. If one needs the ability to answer a question, knowing the answer looks like a good way to do it. Thus, knowing the explanation why p works as a first pass hypothesis of how we solve the cognitive management problem for why questions.

However, merely knowing an explanation is not enough for understanding. One can be told what the explanation of an event is, and thereby come to know what the explanation is, but still not understand why it happened.[17] For this reason, explanation-based

[16] I do not claim that this was the intention behind the creation of the accounts of understanding listed below. Most likely the goal was to find an account that aligned well enough with our intuitions about understanding. However, since the intuitions are about understanding, they happen to lead authors to accounts that address cognitive management.

[17] Even (Grimm 2006), who argues that understanding is a species of knowledge, agrees that mere knowledge of the explanation does not provide understanding; he argues that understanding requires "thick" knowl-

accounts that have been defended in the literature claim that in addition to knowledge of an explanation, understanding requires more. Accounts have been defended claiming that understanding requires grasping the explanation (Strevens 2009), (Trout 2007), (Grimm 2006); using the explanans to derive the explanandum (Hempel 1965); possessing a unifying system of explanation-generating inference patterns (Friedman 1974), (Kitcher 1989); knowing that the explanation is the best explanation (Khalifa 2011); and being able to produce explanations (Grimm 2010).

The management account explains why understanding requires more than mere knowledge of the explanation: one cannot solve the cognitive management problem just by having knowledge. At a minimum, one must be able to recall that knowledge when it is needed to answer a question. The possession of knowledge does not guarantee that one has this ability. Furthermore, the subject of why p encompasses the question 'What is the whole truth about why p,' which arguably includes more than just the question why p. Thus, knowing that one answer is not enough. The management account explains why explanation-based accounts almost universally demand more than mere knowledge.

However, even explanation-based accounts that demand more than mere knowledge are imperfect. One shortcoming is that they all only cover understanding-why; they are not extensible to accounts of understanding-that, understanding a subject matter, or understanding a person or object. There is no such thing as the explanation of Barack Obama, so understanding him cannot just be knowing his explanation. There is no explanation of the proposition that p, so explanation accounts are ill-equipped to account for understanding that. Although it is possible to explain a subject matter, there is no such thing as *the* explanation of a subject matter. These accounts fall short, as examination of their details shows further.

Consider causal accounts of understanding. For example, Strevens' account of explanation is that explanations cite causal facts, but only the causal facts that are difference

edge, where this includes grasping the explanation the way a priori mathematical knowledge requires grasping the mathematics involved. See (Grimm 2006, p. 531).

makers, not the entire causal history (Strevens 2004), (Strevens 2009). (Woodward 2004) also claims that only part of an event's causal history is relevant for explanation, namely those facts whose manipulation would change the explanandum. While it makes sense to think one can understand why something happened if one has a good grasp of its causal history, one does not understand a work of art, or a person, or a subject matter by knowing its causal history (what is the causal history of a subject matter, anyway?).

Next, consider the contrastive account of explanation, defended in (Lipton 1991) and (Lipton 2004). On this account, the fundamental form of explanation is explaining why p *rather than* q. In context, some foil q is relevant when explaining why p. While it makes sense to say that one understands why p when one can rule out alternatives to p, it does not make sense to say one can understand a person by ruling out other people or understand a subject by ruling out other subjects.

Consider the Deductive-Nomological account, defended by (Hempel 1965). The Deductive-Nomological account says that an explanation of why p is a derivation of p from laws of nature plus initial conditions. While perhaps one can understand why something is true if one can derive it from laws of nature and initial conditions, one cannot understand a subject matter or a person by deriving them from universal laws and initial conditions. It is not clear what it would mean to derive such things from universal laws. The account is restricted to understanding-why.

The unification account, introduced in (Friedman 1974) and expanded in (Kitcher 1981), (Kitcher 1989), and elsewhere, requires more detailed consideration to rule out. The unification account of explanation starts with a unifying system of inferences that allows one to derive one's entire body of knowledge from the smallest set of inference patterns and background assumptions. The unification account says that a good explanation of why p is an argument that conforms to one of the patterns in the unifying system and whose conclusion is p. For example, Newtonian Mechanics showed how to derive phenomena as disparate as the behavior of billiard balls and the motion of comets from the same basic equations.

Given the unifying system of Newtonian Mechanics, one could then explain events using arguments based on Newtonian Mechanics.

About the relationship between unification and understanding, Kitcher says, "Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again, and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute)," (Kitcher 1989, p. 432).[18] We can take this as a statement of a general account of understanding: one understands X when one's beliefs about X are unified. X can be a person, a theory, or a fact and the account still makes sense.

The problem is that this is too restrictive. There are more ways to understand than just through unification, and unification does not always produce understanding. First, consider Anna, who has a high level theory of physics that unifies her beliefs. However, she can only apply the theory in simple cases. Despite the unification of her beliefs, Anna does not understand based on this unification. The management account explains this counterexample, for despite the unifying theory, she does not have the cognitive management skill to make the relevant inferences from the relevant parts of the theory.

In the other direction, consider Sally, who is skilled in using two (accurate) theories to answer questions in two domains, chemistry and orbital mechanics. According to the unification account, Sally understands worse than Anna, due to Anna's unification. However, the opposite is true. Effective but disparate theories give more understanding than unified but unusable theories.

The management account explains this failing: the ability to manage one's resources does not prejudge which methods one must use. Their effectiveness is all that matters. The unification account errs in suggesting that one method of alleviating the cognitive management problem—reducing the number of inference patterns and brute facts to work with—is the only way. This is not the only way to organize a cognitive system to solve

[18]Friedman makes a similar statement at (Friedman 1974, p. 195).

the cognitive management problem. The mind can manage resources with multiple, easily used methods that are not unified but effectively solve problems and answer questions in their respective domains.

Thus, the management account shows us what makes explanation-based accounts attractive, and where they fall short.

3.6.2 Applicability accounts

Some accounts of understanding claim that understanding is the ability to apply one's beliefs or knowledge, as in (de Regt and Dieks 2005), (Grimm 2010), and (Wilkenfeld 2013). The ability to apply beliefs depends on cognitive management; one must be able to determine which beliefs are relevant and what to do with them. Thus, the management account can explain why applicability accounts are attractive.[19] However, the applicability-based accounts that have been presented in the literature are either too restrictive or fail to explain what is valuable about understanding.

First, consider the account in (Grimm 2010). Grimm claims that to understand why a plane can fly is to "grasp Bernoulli's principle, and to be able to apply the principle to the relevant details about the plane," (Grimm 2010, p. 341). In general, understanding why a phenomenon occurs is the ability to apply the relevant principles and facts to infer that the phenomenon occurs. The ability to apply one's knowledge is the ability to determine which knowledge is relevant to the question at hand, and what to do with it. In other words, this is a case of cognitive management. The value of the account is captured by the management account.

However, like explanation accounts, this only covers understanding-why. It does not explain what unites different forms of understanding. The account as Grimm formulates it cannot be thus extended. With people, subject matters, and propositions there is no

[19]The management account, with its emphasis on being able to answer questions, could be considered an applicability account. However, I will use this heading to cover just the other applicability-based accounts that have been offered.

particular event to understand and to which one must apply the relevant principles. The reliance on principles hinders the ability to extend this account to cover other forms of understanding. It is not clear which principles must be applied to understand a person. Grimm's account based on the application of principles is thus too restrictive.

Even if Grimm can overcome this difficulty, Grimm's account does not imply that understanding is essential. While the ability to apply principles is useful, it is not clear that it is essential for effective cognition. Based on his account, one might conclude that understanding is a good trait to have, but that one could get by without it. One could get by without the ability to apply general scientific principles to cases. This account does not imply that understanding is indispensable, and does not demonstrate that it requires further study.

Next consider the account of (de Regt and Dieks 2005), which says, "A phenomenon P can be understood if a theory T of P exists [... and] scientists (in context C) [...] can recognise qualitatively characteristic consequences of T without performing exact calculations." (de Regt and Dieks 2005, pp. 150–151) One understands when one has a theory that one can draw intuitive consequences from without much effort.

First of all, the strength of the account is that it captures the fact that someone who understands a theory is adept at using that theory. The ability to draw qualitative consequences without calculation requires one's mind to be organized in a way that allows for the recognition of shortcuts. One must be able to recognize some relevant features that can be gleaned without calculating. The ability to do this depends on the ability to find what is relevant; only some relevant features of a situation and some relevant features of the theory are amenable to inferring qualitative consequences. The ability to find these is an instance of cognitive management. The management account also explains why this account relies on an un-analyzed ability to recognize consequences: lacking a detailed solution to the cognitive management problem, we are often left with black boxes in places whre cognitive management occurs, as is the case here.

However, de Regt and Dieks' account is too restrictive. Understanding a theory consists of more than just the ability to quickly draw some qualitative consequences from the theory. Even restricted to mathematical theories, their account leaves out the ability to set up the equations correctly in order to perform rigorous calculations. Someone who understands should be able to do this, and it does not involve drawing qualitative consequences without calculation. The management account avoids this problem, for cognitive management plays a role in answering any question, not just qualitative questions about theories. In addition, the management account allows us to explain the inadequacy: the ability to draw qualitative consequences cannot solve the cognitive management problem in general. The ability to solve a problem by exact calculation also depends on cognitive management in bringing to mind the right equations and performing the right calculations.

Wilkenfeld attempts to fix the difficulties with de Regt and Dieks' account. He says, "The difference between understanders and non-understanders is that the former, but not the latter, can utilize the understood effectively." (Wilkenfeld 2013, p. 1002) In explaining how one can utilize the understood effectively, he claims that understanding means having a representation that "enables efficacious (according to standards relevant in context) inferences pertaining to, or manipulations of, [the object understood]," (Wilkenfeld 2013, p. 1003-1004). First of all, the management account explains what is attractive about this account: inferences are efficacious in virtue of allowing one to answer questions. If one has a representation that allows efficacious inferences, then that representation allows one to solve the cognitive management problem, for otherwise it could not help one answer questions.

The trouble is that the ability to manipulate representations is not explained. In this case, the lack of detail undermines the account, for representation manipulation in this account is a place holder, defined as what one does with representations when one understands. The importance of this place holder is not clear, and thus the account does not show us where future research should go and it does not show that understanding is essential. The

management account, in contrast, identifies understanding with the solution to a problem that is independently studied and that cognitive systems must solve.

3.6.3 "Grasping" accounts

There are also accounts of understanding that claim that understanding is based on a relation called "grasping." The primary example is Kvanvig, who claims, "Understanding requires the grasping of explanatory and other coherence-making relationships in a large and comprehensive body of information," (Kvanvig 2003, p. 192).[20] The appeal of this account is that it is plausible that the coherence-making connections among beliefs establish relations of relevance. Thus, grasping such connections could help find what is relevant to a given topic. Just follow the connections. That is why the connections must be grasped, for the mere presence of connections does not ensure that they will play a role in one's reasoning.[21]

The problem is that it is not clear what grasping is. To the extent that the relevant use of 'grasping' is familiar, it is because it is a near synonym to 'understanding.' In order to show that grasping accounts are not circular, we need an account of grasping.[22] An account of understanding in terms of un-analyzed grasping is thus not informative.

In contrast, the management account does not have the same problem of vagueness. The cognitive management problem, as part of the frame problem, is understood well enough that research programs dedicated to solving it are being carried out. Thus, though the

[20]Elgin's account is similar. She states, "I suggest that understanding is a grasp of a comprehensive body of information that is grounded in fact, is duly responsive to evidence, and enables non-trivial inference, argument, and perhaps action regarding that subject the information pertains to," (Elgin 2007, p. 39).

[21]Other accounts identify understanding with grasping the criteria that make something the kind of thing that it is, or that make it a good instance of that kind. Boylu says, "what is involved in understanding is not merely grasping the relations among a bunch of facts [...], understanding is the criteria-grasping epistemic pursuit," (Boylu 2010, p. 603). For example, understanding a painting means grasping those features that make it a good or bad instance of a painting. This account is too restrictive, applying only to understanding objects that have criteria for being of the kind they are. This cannot be extended to understanding why, understanding that, or understanding a subject matter, or even to understanding objects that do not have well-defined criteria of being that object.

[22](de Regt and Dieks 2005) presents and account of understanding described as giving an account of grasping. However, given their account, we can easily replace all references to grasping with the ability to apply a theory. For this reason, I classified it as an applicability account.

management account does not provide all the details of what understanding is, there is a clear direction in which to direct further research.

Furthermore, the management account explains why authors have turned to grasping despite its vagueness: without a solution to the cognitive management problem, the exact mental relationships involved in understanding cannot be precisely specified. Thus, accounts sometimes have black boxes where cognitive management occurs. The management account shows us why this happens.

3.6.4 Summary

In all cases, the strengths and weaknesses of accounts of understanding can be explained by the management account of understanding. Existing accounts of understanding can be seen as hypotheses about what it takes to be able to manage our cognitive resources. The accounts differ in how much background they take for granted and how adequate they are, but they all identify pieces of the puzzle of how we manage to solve the problem.

3.7 Conclusion

The management account is a general account of what understanding is and why it is important. Understanding is the ability to manage cognitive resources, an ability without which we could not function as cognitive agents at all. It is what allows us to answer questions. That is why understanding is so crucial in our epistemic lives. The management account also shows us what other accounts of understanding are trying to explain. It makes sense of the details of these accounts. More than that, it vindicates and explains many of the intuitions and commonplace claims made about understanding. For all of these reasons, this is a fruitful account of understanding.

Filling in the details of understanding in humans requires work in philosophy, cognitive science, psychology, AI, and computer science. This does not mean that there is no philosophical work left to be done. The fact that understanding is cognitive manage-

ment has implications for several philosophical questions relating to understanding, such as whether it is possible to acquire understanding based on false beliefs, whether it is possible to acquire understanding by luck, and what is the relationship between understanding and explanation. There is a rich field to be explored, and the remaining chapters of this dissertation continue this work.

CHAPTER 4

BETTER UNDERSTANDING THROUGH FALSEHOOD

4.1 Introduction

Can false beliefs improve understanding? A common intuition is that they cannot, that false beliefs always hinder understanding. There is an expectation that understanding and the truth go hand in hand. However, in this chapter I argue that sometimes false beliefs can increase understanding. In addition, one can lose understanding by gaining true beliefs. The relationship between understanding and the truth is not as simple as one might have thought; it is not simply a matter of false beliefs hindering understanding while true beliefs improve it.

The fact that false beliefs can improve understanding is inimical to reductive accounts of understanding that treat it as a species of knowledge. According to reductive accounts, to understand a topic is just to have the right kind of propositional knowledge about that topic. Since knowledge that p can only exist if p is true, the reductive view requires that understanding can only be based on the truth. The beliefs must constitute knowledge. This chapter will provide reason to reject these reductive views.

In addition, this chapter demonstrates that by answering the question of whether false beliefs can improve understanding, we also answer the question whether understanding is factive, a question debated in (Kvanvig 2003), (Elgin 2007), (Elgin 2009), (Pritchard 2008), (Riggs 2009), and elsewhere.[1] This debate has largely focused on understanding

[1]Note that this is an extension of the original use of the term 'factive'. The debate is not whether that which you understand must be true, but whether understanding of something must be based on true beliefs.

as an all-or-nothing state, rather than as coming in degrees. I show that if one's degree of understanding can be increased through false beliefs or hindered through true beliefs, then understanding is not factive in the sense used in these debates.

Truth is often taken to be the only, or highest, epistemic good. On the other hand, understanding is often taken to be the goal of inquiry. The argument in this paper undermines the connection between understanding and the truth, thus creating a tension between these two claims. This provides occasion to reassess the roles of truth and understanding in inquiry.

What role does truth play in understanding, then? I argue that understanding means having the ability to find the truth. It does not always require already having true beliefs. In particular, some false beliefs can aid this ability, and some true beliefs can hinder it, but the ability to find the truth is what matters. Thus, understanding retains an important connection to the truth, just not the connection many have suspected.

The argument presented below does not depend solely on the management account of understanding (which I defended in Chapter 3), but also employs two methods by which we measure and compare degrees of understanding in real life. The first method, used in classrooms, job interviews and psychology experiments, involves finding answers to questions, especially novel questions. If one has a greater ability to answer novel questions about a topic, then one has greater understanding of that topic. The second method, also employed in evaluating understanding, is reasoning ability. If one can reason better about a topic, then one has a better understanding of the topic (for more detail, see section 4.2.2).

Using these methods of comparison, I identify two broad classes of cases where false beliefs can improve understanding and true beliefs can hinder understanding. The first class of cases involve approximation and idealization. Approximations and idealizations are false, but believing them to be true can sometimes increase understanding. The second class of cases involves misleading true beliefs. Misleading true beliefs are true, but believing them can decrease understanding. Understanding can be increased through the acquisition

See section 4.5.

of false beliefs and decreased by true beliefs.

The rest of the paper is structured as follows: in section 2, I clarify the nature of the question about the relationship between understanding and truth. Next, section 3 provides two classes of counterexample to the factivity of understanding. In light of these counterexamples, Section 4 considers what role truth plays in understanding. Section 5 shows how the argument here implies that understanding is not factive, and explains how the present argument avoids objections to past arguments against factivity. Section 6 explains why there is no reason to impose a truth requirement on understanding.

4.2 Understanding and the truth

Consider these two questions one can ask about the relationship between understanding and truth.

Object Question Must that which one understands be true?

Basing Question Must the beliefs upon which understanding is based be true?[2]

An example illustrates the difference between the two: Suppose John understands why boats float. The object question is whether John's understanding why boats float implies that boats float.[3] On the other hand, the basing question is whether John's understanding why boats float implies the truth of the beliefs that allow him to understand why boats float, such as his beliefs about boats, water, floating, etc.

We can dispense with the object question briefly. On one hand, there is good reason to think that understanding why p implies that p is true. One cannot understand why rocks

[2]This is closely related to the question whether understanding is factive in the sense employed by (Kvanvig 2003), and those adopting his use of 'factive', such as (Elgin 2007) and (Riggs 2009). See section 4.5 for details.

[3]Technically, the object of understanding in the example is the question why boats float. However, following convention, for the object question we decompose 'understanding why boats float' into the verb *understanding-why* taking the proposition that boats float as its object. For the basing question we need not make this assumption.

float in air, because rocks do not float in air. In the same vein, it also appears that understanding that p implies that p is true.[4] For these kinds of understanding, one can only understand what is true.

However, not all understanding requires the object of understanding to be true. One can understand a false theory. For example, I understand several JFK assassination conspiracy theories that are false in almost every detail. Further, the object question does not make sense when applied to understanding a person, or a painting, or a symphony, since these things are not truth apt. Relatedly, one can understand a subject matter, yet if subject matters are sets of questions as argued in (Lewis 1988), it does not make sense to ask whether a subject matter is true or false. For these kinds of understanding, there is no requirement that what you understand must be true. This covers the object question.

On the other hand, the basing question—whether one can understand based on false beliefs—applies to understanding theories, subjects, objects, and people, even in cases where the object question makes no sense. Though one can understand a false theory, it is still an open question whether that understanding must be based entirely on true beliefs (perhaps true beliefs about what the theory implies). Even though a person or a symphony cannot be true or false, it is still an open question whether one's understanding of a person or a symphony must be based entirely on true beliefs (perhaps true beliefs about the person or symphony).

The basing question arises because many beliefs can be implicated in understanding. In understanding why boats float, one must draw on beliefs about buoyancy, density, pressure and so forth. As a first approximation, the basing question is whether these beliefs must be true.[5] In each case, we can ask whether the beliefs that are implicated in understanding

[4] A potential problem here is that one can say to a colleague, "I understand that you got a promotion," as a way of expressing that one has reason to think she got a promotion but that one is not sure and is seeking confirmation. However, it is plausible that this is a different use of 'understanding that' that does not affect the use that concerns us here.

[5] There is an analogous debate regarding knowledge, but this debate does not occur under the heading 'Is knowledge factive?' but rather under the heading 'Can knowledge be based on false lemmas?' A defense of the 'No false lemma' thesis can be found in (Harman 1973), and objections can be found in (Klein 2008).

must be true.

Which beliefs are implicated in understanding? Without taking a stand on any specific account of understanding, we can say that if understanding is constituted entirely by beliefs as reductive theories claim, then the beliefs implicated in understanding are simply the beliefs that constitute understanding. If, on the other hand, understanding is a kind of ability, then the beliefs implicated are the beliefs that contribute to the ability. Some cognitive abilities depend on the right basis of beliefs from which to draw appropriate inferences. For example, the ability to calculate a projectile's path depends on beliefs about Newton's laws, or similar beliefs. The beliefs are implicated in understanding to the degree that they contribute to the cognitive ability that is understanding.

Given the above, the simplest way to clarify the basing question is to ask whether the following strict truth condition is met:

Strict Truth Condition If S understands a topic, then every belief implicated in S's understanding that topic is true.

I use 'topic' as a general term for the different objects that understanding can take as complement, including propositions, why-questions, persons, works of art, subject matters, and theories. According to the Strict Truth Condition, a single false belief relating to X precludes understanding X.

However, this way of defining the issue is too strict to be interesting; understanding almost trivially fails the Strict Truth Condition. To illustrate, suppose Sally has a single false belief about a minor aspect of the circulatory system of bats. Other than this, Sally has only true beliefs and is an expert on the bat circulatory system. She can explain in detail how every part of it works (except for the one minor error), and she can successfully experiment on it and manipulate it. By any sane standard, Sally understands the bat circulatory system despite her one false belief. Thus, the Strict Truth Condition is false.

To locate a more interesting, less strict way of addressing the basing question, it helps to recall that understanding comes in degrees. We are concerned with whether one can ac-

quire understanding by gaining false beliefs. Since understanding is not just all-or-nothing, we should not restrict ourselves to all-or-nothing truth requirements. What matters is the degree to which the beliefs implicated in understanding are connected to the truth, in terms of how many true beliefs versus false beliefs one has, and how closely one's false beliefs approximate the truth. Let us call the degree to which a body of beliefs approaches the truth the body's *degree of veracity*.

Now, the motivation for asking the basing question is to figure out whether increasing understanding requires approaching the truth by increasing the degree of veracity of one's beliefs. Does greater understanding of a topic imply believing more truths about the topic? Does less veracity among one's beliefs lead to a lower degree of understanding? In short, can false beliefs bring about greater understanding?

To summarize, understanding is based on background beliefs and (perhaps) cognitive capacities. If we hold capacities fixed, the question is whether false beliefs are capable of increasing understanding. The issue can be encapsulated in the following condition:

Covariance Condition The degree to which one understands X varies in the same direction as the degree of veracity of the beliefs that are implicated in understanding X, holding cognitive capacities fixed.

The Covariance Condition is motivated by the idea that the success involved in understanding is having true beliefs, so that degrees of veracity and degrees of understanding always vary in the same direction. If one can gain understanding by lowering the degree of veracity of one's beliefs, then the answer to the basing question is 'No, understanding need not be based entirely on truth.' Thus, to answer the basing question, we just need to be able to compare degrees of veracity and degrees of understanding to determine whether the covariance condition holds.

4.2.1 Comparing degrees of veracity

In order to find counterexamples to covariance, we do not need a full account of degrees of veracity, we only need sufficient conditions for when a body of beliefs A has greater veracity than another body of beliefs B. Three such sufficient conditions are:

Body of beliefs A has a higher degree of veracity than body of beliefs B if

1. A and B contain the same beliefs, except for additional true beliefs in A,
2. A and B contain the same beliefs, except for additional false beliefs in B,
3. A and B only differ in that A contains beliefs that are closer approximations to the truth while B contains inferior approximations.

Whatever other complications are present in determining degrees of veracity, adding true beliefs and nothing else increases the degree of veracity. Similar considerations apply to the other two conditions. If we deal only with cases involving the above three conditions, we avoid the need to deal with trade-offs between more false beliefs in one area compensated by more true beliefs in another, and other trade-offs. We also avoid the problem of determining how to count beliefs. These sufficient conditions for greater veracity will enable us to show that one can achieve a greater degree of understanding with a lower degree of veracity, in violation of the covariance condition.

4.2.2 Comparing degrees of understanding

Assessment of putative counterexamples to the Covariance Condition also requires a way of comparing degrees of understanding, in addition to a way of comparing degrees of veracity. We have already identified three sufficient conditions for comparing degrees of veracity of beliefs. In the literature, previous descriptions of degrees of understanding are in terms of degrees of "breadth and depth" (Kvanvig 2003), or the degree of one's ability to "connect, synthesize, and grasp a body of information," or "unify a body of information," (Elgin

2007, p. 41). While these are ways of describing degrees of understanding, it would be preferable to have measures that are more specific. This section identifies two such methods of comparing degrees of understanding.

In everyday situations as well as in more formal endeavors such as psychology and education, we often need to make comparisons of degrees of understanding. The management account of understanding also allows for comparisons of degree of understanding. The methods we will use to compare degrees of understanding are:

1. The ability to answer questions correctly, including novel questions, and

2. Reasoning ability,

3. The management account of understanding.

First, the greater one's ability to answer questions about a topic, the greater one's understanding of that topic. Second, the greater one's reasoning ability on a topic, the greater one's understanding of that topic. Third, the management account of understanding identifies a greater degree of understanding with a greater ability to manage cognitive resources. These allow us to make comparative judgments about degrees of understanding sufficient to identify counterexamples to Covariance.

The first method of comparison is to compare the ability to correctly answer questions regarding a topic. This is the method that teachers use to test students, employer's use to test job candidates, and psychologists use to test subjects. Novel questions are important to such tests, because verifying that someone understands a subject requires more than giving them a series of questions or problems that they have already encountered, to which they could have simply memorized the answers. To test whether someone understands quantum mechanics, one asks novel questions about quantum mechanics. To test whether someone understands why solar eclipses occur, one asks novel questions regarding why solar eclipses occur. The wider the range of questions one can answer, the greater one's understanding.[6]

[6]Like any method of measurement, interfering factors can prevent the use of this method from giving an

In general, to test for understanding, we present subjects with a range of questions from the familiar to completely novel. This requires them to go beyond what they may have memorized and demonstrate their understanding.[7]

When using this measure, it is important to be clear about which topic is under consideration. For example, when we ask, "How well can Susan answer questions about morality?" we might be asking how well she can answer questions about what is right and wrong. On the other hand, we might be asking how well Susan can answer questions about *why* actions are right or wrong. One can be able to answer questions about what is right or wrong without being able to answer any questions about why actions are right or wrong. Suppose utilitarianism is true, but that one can better determine what is right and wrong in specific cases using Kantian ethics. Then one who believes Kantian ethics is correct will, according to this measure, understand what is right and wrong better than one who believes utilitarianism is correct. On the other hand, she has worse understanding of *why* actions are right and wrong.[8] Clarity about the topic is important.

The rationale behind this method is that the ability to answer questions, especially novel questions, expresses the underlying state of understanding. It captures the fact that understanding contributes to our grasp of reality through allowing us to find the truth. Genuine understanding is not idle. Either understanding is itself an ability that contributes to answering questions, or it is constituted by beliefs that are readily applied to allow one to answer questions. Because understanding plays this role, the ability to answer questions reveals degree of understanding. If there are no interfering factors, such as intoxication or uncharacteristic performance errors, then the test of correctly answering questions reveals the degree of understanding. While individual tests may not perfectly implement the strat-

accurate result. This practical concern about application will not affect the counterexamples to covariance given below.

[7]This is nearly universally accepted in the education literature and in psychology. For examples from education, see (Skemp 2006) and (Wiggins and McTighe 2005). For examples from psychology, see research on the illusion of understanding, such as (Rozenblit and Keil 2002) and (Fernbach, Rogers, Fox, and Sloman 2013).

[8]I thank an anonymous reviewer at *Pacific Philosophical Quarterly* for this example.

egy, the principle that greater understanding produces greater ability to answer questions is all we need here. The basing question, in these terms, is whether one can increase one's ability to answer novel questions solely through adopting an overall less true set of beliefs. If one can do this, then the Covariance Condition fails.

The second method of comparing degrees of understanding is by comparing levels of reasoning ability. When someone understands a topic, they are able to reason well about that topic. Greater reasoning ability implies greater understanding. Either understanding is an ability that takes part in reasoning, or it is constituted by beliefs that enable better reasoning. There are various aspects of reasoning ability regarding a topic: the ability to distinguish the relevant from the irrelevant, the ability to bring the right facts to mind at the right time, and the ability to evaluate how evidence relates to conclusions within the topic. For example, if one can tell which equation to apply to a physics problem, one understands physics better than someone with the same mathematical abilities but who struggles to figure out which laws to appeal to. (As with the previous measure, clarity about the exact topic is important.)

This method can be used to assess understanding when one is not in a position to ask the subject questions. Suppose we wish to assess whether John or Larry understands Susan better by studying essays they wrote about her in the newspaper. Suppose each essay only contains true statements, but John relates facts and anecdotes that capture the central features of Susan's character while Larry relates facts and anecdotes that give a completely misleading picture of her character. We are not in a position to ask them any questions about Susan, but we can still compare their degrees of understanding. Assuming it was not Larry's goal to be misleading, he does not understand Susan as well as John does, for he displays worse reasoning abilities about Susan. John can properly assess what is relevant or important about her, while Larry cannot.[9] Thus, reasoning skills provide another

[9]Reasoning ability is closely related to the ability to answers questions, since reasoning ability is what allows one to answer questions. However, they are different methods. One can assess novel questions without directly addressing reasoning skills, and one can assess reasoning ability without asking questions.

method upon which we have an independent grasp, and which can provide information about degrees of understanding.

Both measures are intended to measure understanding, and thus they should agree on cases. There might be concern that they do not in fact measure identical states. Rather than arguing that they do measure the same underlying state, consider the consequences if we assume they do not. What would be missing from the ability to answer questions to prevents it from being a complete measure of understanding? The missing component is *how* one answers questions. Some ways of answering questions, such as memorizing a list, are effective but not enough for understanding. Understanding is not just answering questions correctly; it also requires answering them in the right way. In other words, one must answer them using proper reasoning ability. This suggests, then, that increasing both ability to answer questions and reasoning ability increases understanding, and likewise for decreases. When the measures agree, that is sufficient.

On the other hand, what is missing from reasoning ability that prevents it from being a complete measure of understanding? Here, the missing component is proof that the reasoning in question is not just idle, mental wheel-spinning. Considering reasoning alone could lead to the conclusion that understanding is merely a useless sense of satisfaction with how one can reason about a topic (see (Trout 2002) for more about this objection). One must be able to apply this reasoning ability for it to be understanding. Thus, the potential deficiency of this measure is addressed by the first measure. Even if the the ability to answer questions and reasoning ability do not coincide, together they are sufficient to measure changes in understanding. If both measurements give the same verdict, as in the cases below, they are adequate to identify counterexamples to Covariance.

Finally, the management account of understanding, defined and defended in Chapter 3, defines a measure of degrees of understanding. According to the management account, one understands a subject just in case one has the ability to manage cognitive resources successfully to answer questions in that subject. Managing cognitive resources implies being able

to identify the relevant information and exploit it using the relevant cognitive capacities. It allows one reliably to find valuable information within that area. There are several dimensions along which degree of understanding can vary: there is degree of reliability, breadth of reliability, and degree of centrality of questions one can answer reliably. If one does better along all of these lines, then one unequivocally has a higher degree of understanding, though other cases will be more difficult to solve. We can use the management account of understanding as a measure of degree of understanding.

The above measures of understanding provide sufficiently precise measurements of degrees of understanding. The counterexamples to Covariance below involve agents for which the differences in degrees of understanding are larger than any uncertainty in the applications of the measures.

4.3 Cases of understanding through falsehood

We now have enough information to compare degrees of veracity among bodies of beliefs and to compare degrees of understanding. In this section, we employ these methods to examine a series of cases in which an agent with greater veracity among the beliefs implicated in understanding nevertheless has a lesser degree of understanding.

4.3.1 Approximation and idealization

The first counterexample to Covariance is approximation.[10] Approximation is the use of a value, equation or statement that is not exactly correct but that is nevertheless close to the truth.[11] Approximations are strictly speaking false. If it is exactly 45.2 degrees Celsius,

[10]Elgin also discusses approximations and outdated scientific theories in her argument against factivity in (Elgin 2007). However, she does not use them as her counterexamples to factivity; she allows that they might provide only honorary understanding, not genuine understanding. Idealization is the only example she defends as a counterexample to factivity. Thus, by identifying approximation as a counterexample, the argument in this paper goes further. For more details about Elgin's arguments and where they differ from mine, see section 4.5.1.

[11]The notion of "close enough" is difficult to define exactly, but for our purposes we can rely on a working sense of approximation. Approximation has been explored for example in (Elgin 2004).

45 degrees Celsius can often serve as a good approximation.[12] The temperature is not 45 degrees, but it is approximately 45 degrees.

Like approximation, idealization is the use of falsehoods in problem solving. An idealization is a simplified version of a concrete situation that abstracts away from some features for the sake of clarity.[13] Some idealizations are approximations, but some are not. As an illustration of this, in formulating linguistic theories, linguists employ the idealization that speakers have infinite capacities to perform certain kinds of operations (e.g. embedding), abstracting away from human limitations. However, humans do not even approximately have infinite capacity.

While approximations and idealizations are strictly false, they are employed in reasoning as if they are true. One makes derivations from them as one does from any belief. While we often use approximations with the explicit knowledge that they are approximations, suppose one believes that an approximation is literally, exactly true. Then one has a belief that is strictly speaking false. If this false belief were removed and replaced with the exact truth (or a better approximation), the overall body of beliefs would have greater veracity. There are cases where believing a closer approximation to the truth leads to lesser understanding.

Consider physics, a field where approximations are ubiquitous. Today, we know that Newtonian Mechanics (NM) is not the exact truth. It is an approximation that works well in many of the situations where we need to use physics. Strictly speaking, it is false. General Relativity (GR), while probably not exactly true either, is a better approximation that gives accurate predictions in a much wider range of circumstances. Thus, believing NM rather than GR is a decrease in degree of veracity; a shift from a closer approximation to a worse approximation. Let us assume for the sake of argument that GR is exactly true. Consider a

[12]The statement "The temperature is between 45 and 46 degrees," is not an approximation in the same situation, since it is a strictly true statement about the temperature. The fact that it does not give the exact temperature does not make it an approximation.

[13]There is some controversy regarding just how accurate this account is, for example see (Cartwright 1995) and (Kowalenko 2009), but it is good enough to give a working sense of what idealization is.

student who has been taught GR, and has enough background in math to grasp the meaning of the claims. Consider also a second student who has the same background but has learned Newtonian Mechanics and believes it to be the exact truth. The second student has fewer true beliefs (by lacking belief in GR), and more false beliefs (by believing in the exact truth of NM). Otherwise, they are the same. Thus, the first student's beliefs have a greater degree of veracity.

Suppose both students consider the subject of the motions of heavenly bodies. The student who only knows GR spends hours fruitlessly trying to solve the field equations. She does not come close to describing the motion of any heavenly body in any form, no matter how approximate.[14] On the other hand, the student using NM with the same level of mathematical skill is able to derive equations for orbits that are approximately correct for several bodies. As she acquires more true beliefs about celestial motion, she continues to use NM fruitfully in answering further questions. With this in mind, let us turn to their degrees of understanding.

First, compare their degrees of understanding using the ability to answer questions. Given a series of novel questions about celestial motion, the student who believes in NM, and whose beliefs have lower veracity, is able to answer many correctly. On the other hand, due to the difficulty of solving Einstein's Field Equations, the student who knows GR cannot answer any. Finding solutions to Einstein's Field Equations is so difficult that solutions are named after their discoverers (e.g. the Schwarzschild solution). Since the subject is celestial motion and not the universal laws governing such motion, by the measure of answering questions, the student whose beliefs have less veracity has greater understanding, which implies that lower veracity among beliefs can provide greater understanding, in violation of the Covariance Condition.

Next, we compare degrees of understanding by comparing reasoning ability. Return to our two students before they learn physics, whether NM or GR. They each contemplate the

[14] Anyone who has learned GR should recognize the plausibility of this happening to a student who has more than enough mathematical skill to master NM.

motions of heavenly bodies, but neither one has the ability to reason through the topic of planetary motion. They have no sense of what keeps everything moving in the observed orbits, or how hypothetical bodies would move. Then one student learns NM, taking it to be the exact truth, and the other learns GR. The first student can identify which information is relevant, which factors play a significant role in determining celestial motion and which do not. When confronted with features of the motion in the heavens, she can often explain what is going on; she can successfully navigate the topic. However, the student who learns GR is not in the same position. Relying on GR, the student knows the equations governing the universe, but cannot use them to work out even general constraints on how heavenly bodies move. Though possessing new true beliefs, the student is still confused. If this student comes to believe that GR is false and that NM is true, she will thereby increase her ability to reason about celestial motion and thus she will have greater understanding. Both measures give the same verdict, and the Covariance Condition fails.[15]

Finally, let us compare degrees of understanding using the management account. Like the above analyses, the management account delivers the same verdict. Our ability to manage cognitive resources, to find useful information in the subject of shock formation and drop breaking is increased by believing NM,even though it is strictly peaking false. Belief in the idealization provides a method for solving these problems that is not present when strictly more accurate models are all one has. Thus, according to the management account, increased falsehood via idealization can increase understanding.

One might object that the above does not show that falsehood can increase understanding; there are true beliefs in the vicinity of the false belief (belief in NM), and these true beliefs do the work in understanding. The student who believes that NM is exactly true presumably also believes that NM is accurate to within practical limits, which is true. Perhaps it is only this true belief, and not her false beliefs, that are implicated in giving her

[15]The same analysis can also be applied to idealizations that are believed to be true. The example of ideal gases is used in (Elgin 2007) to argue that understanding is not factive, though her argument has a weaker conclusion, as shown in section 4.5.1.

understanding of celestial motion.

However, this objection fails to undermine the case. First of all, in realistic cases, she reasons from belief in NM, not from a hypothetical belief that NM is true enough. The fact that we can find true beliefs she has that she could have used is not relevant. In addition, one has greater understanding by the first measure when one has a greater underlying ability to answer questions. Understanding contributes to the ability to apply one's beliefs, whether understanding is constituted by beliefs or is itself an ability. In this case it is NM that contributes to the student's abilities. Once the student gains her false belief that NM is true, she has the ability to answer novel questions about celestial motion. She does not need to infer anything weaker before she actually has the ability. While she could infer that NM is true within error tolerance, she need not do so. Even if she uses NM to build up a large body of true beliefs about celestial motion, the false NM beliefs are still implicated in her understanding. It is not a ladder she has thrown away; it continues to play an active part in the success of her reasoning.

One has greater understanding by the second measure when one has greater reasoning ability. Again, once she comes to hold the false belief in the truth of NM, she then has the greater ability. She need not infer weaker beliefs to increase her skills. Furthermore, when actually making derivations, predictions and explanations, the actual belief she draws upon is false. Her derivations do not start from the premise that NM is true within certain tolerances; they start from Newton's laws. The belief about error tolerance need play no role in her reasoning. The false beliefs are central and are implicated in her understanding. They continue to play a role in her reasoning. Thus, approximation does provide an example of gaining understanding through falsehood.

The same is true of the management account: the false belief in NM is what allows her to organize her reasoning to find what is relevant to answer questions about celestial motion. Appealing to possible true beliefs that she could have used does not undermine the case.

Thus, by all measures of understanding—novel questions, reasoning ability, and the management account of understanding—approximation is a counterexample to the factivity of understanding. One can gain understanding by lowering degree of veracity.

4.3.2 Misleading truths

Misleading true beliefs are another class of counterexamples to Covariance. A misleading true belief is a belief that inclines one to draw false conclusions. By adopting such a belief, one's overall body of beliefs gains a higher degree of veracity: either one has added one true belief and changed nothing else, or one has removed one false belief and replaced it with a true one. And yet, as I will now show, by all measures, one's degree of understanding decreases.[16]

As an example, consider someone who starts out knowing the basic facts of the JFK assassination, such as where it happened, the number of shots, kind of rifle and so on. Now suppose she learns that JFK's brain went missing in the 1960's. This is a new true belief. However, without information about the context, it suggests a nefarious plot that is not there. There is strong evidence that Robert Kennedy took his brother's brain so it could be buried rather than stored as a curiosity in the national archives. The missing brain is not evidence of a cover-up, but in isolation, the fact that the brain is missing can give rise to suspicions that can lead one off track. It is a misleading fact about the assassination.

First compare degrees of understanding before and after learning the misleading truth in terms of the ability to answer novel questions. A misleading truth just is a true proposition that leads one through good inferences to false conclusions. This is the sense in which the truth misleads: when one treats it as one treats other truths, one is led astray. It is the phenomenon summed up in the famous, "A little learning is a dangerous thing."

To return to our example, from the fact that JFK's brain went missing, one can correctly infer that it did not fail to go missing, but this is not much help. On the other hand, the false

[16]The significance of misleading truths is widely discussed outside the literature on understanding. For example, it figures prominently in the debate about the epistemology of disagreement.

conclusions about conspiracies are more consequential. Given an array of questions about the Kennedy assassination, one would do worse in answering them as a result of the true belief (except those questions relating directly to the location of the brain). The primary effect is doubt, suspicion and inability to draw conclusions. This hurts one's ability to answer questions about the assassination, questions that one could answer if one lacked the misleading true belief. While this will eventually lead one to false beliefs, the hindrance in one's ability to answer questions occurs before false beliefs are inferred. Thus, a true belief can decrease understanding, as measured by the ability to answer questions, in violation of Covariance.

In terms of reasoning ability, misleading true beliefs can decrease reasoning ability in a topic, thereby lowering understanding. If one understands a subject then later adds a new true belief that seems to imply much of what one believes about the subject is wrong, one's abilities to reason go down. With the misleading truth, the effect can be to leave one unsure of what to do, cognitively speaking, perhaps even giving one the false sense that one is doing well when in fact one is confused.

In the Kennedy example, learning about the missing brain leads someone who had a basic sense of how to reason about the assassination to harbor vague suspicions of conspiracies. Worse, the true belief itself provides no clear idea where the suspicions lead or what to do with them. The belief can make irrelevant details seem relevant, as if they are part of a brain-concealing plot. The person does not know what inferences to draw, since the obvious inferences from the misleading belief conflict with the implications of other beliefs. The result is someone who is stuck, confused, and lacking understanding. This effect occurs even before one forms false beliefs as a result of the misleading truth. Measured by reasoning ability, adding this true belief can lower understanding. Both measures show the increase in veracity decreases understanding, so the Covariance Condition fails.

Finally, when we analyze misleading truths in terms of the management account, we find the same verdict. In the above examples, adding the new true belief took an agent

from a state where she had the ability to find valuable information in a subject matter to one where she lacked that ability. The agent originally had the ability to infer some truths about the Kennedy assassination. After gaining the new true belief, she now would instead infer false beliefs. She no longer can determine which facts are relevant, or what to do with them, due to her suspicions. Her degree of ability to find this information has been decreased by the acquisition of a true belief. This is a decrease in understanding induced by gaining a true belief, in violation of the Covariance Condition.

Any misleading truth one cares to mention can give rise to similar counterexamples. As long as the misleading conclusions one is inclined to draw are central to understanding a topic, misleading truths hinder understanding.

The objection raised against approximation and idealization—that the difference in understanding is due to the veracity of nearby beliefs—does not apply here. Suppose John has a misleading true belief. Susan is otherwise in the same position but lacks that misleading true belief. We cannot attribute the difference in understanding to true beliefs that Susan has but John lacks. The difference in the degrees of veracity of their beliefs goes precisely in the opposite direction: John has more true beliefs. The objection fails. Misleading true beliefs stand as another counterexample to Covariance.

4.3.3 Summary

The above counterexamples show that the Covariance Condition does not apply to understanding. One can gain understanding by adopting an overall less veridical body of beliefs, and one can decrease understanding by adopting an overall more veridical body of beliefs.

The above examples are not simply outliers that one can set aside. Cases such as these are prevalent. For one, the history of science contains a long succession of people believing approximations or idealizations to be the exact truth. The limits of the approximation were not yet discovered. This alone shows that counterexamples are not insignificant outliers.

Further, misleading truths crop up frequently: they are deliberately promulgated in

politics and advertising, and they are a common result of our cognitive limitations. It often happens that we think some fact points to a conclusion, but it turns out not to. The failure of the Covariance Condition is not a mere logical trick; it is a central feature of understanding.

4.4 The role of truth in understanding

Yet surely there must be some negative effects due to having lower veracity of beliefs. Truth must have some role to play in understanding. In this section I consider in more detail what effect the degree of veracity of beliefs has on understanding.

I have shown above that understanding can be gained through false beliefs. However, the lesson is not that truth plays no role in understanding.[17] Instead, the lesson is that understanding is not entirely a matter of already having the truth, but is also a matter of *being able to find* important truths. Beliefs take part in understanding via how they contribute to skill in discovery.[18] In order for a belief to increase understanding by the measures used above, it need only have the right inferential profile. That is what allows it to contribute to reasoning skills, the ability to answer questions correctly, and cognitive management ability.

That said, having a true belief that p at least allows one to correctly answer the question whether p. If the question whether p is true is part of the topic under consideration, then the true belief will help understanding. Further, a true belief has no false deductive consequences, whereas all false beliefs do. Thus, there must always be limits to how much a false belief can increase understanding. Consider the physics students described above. The student using NM understands celestial motion better than the student using GR because given their level of mathematical competence, neither student can use GR. They are not capable of dealing with those questions where GR gives correct results and NM does not. However, with greater mathematical acumen, the student using GR can do better than

[17] Indeed, if one accepts the argument in (Klein 2008) that false beliefs can play a role in producing knowledge, it does not prevent understanding based on false beliefs from playing a role in producing knowledge.

[18] This does not imply that before a belief can contribute to understanding one must actually use it to discover information. Understanding is an ability, and one can have an ability that one does not exercise.

the student using NM. The student using GR may even be able to derive NM as an approximation, at which point she will have a complete advantage, by being able to use NM while recognizing that it is but an approximation. In the examples above, only cognitive limitations prevent the truth of GR from producing greater understanding to the student who believes it is true.

There are some falsehoods that we do not have the capacity to use badly. Given our state of technological and scientific development, some falsehoods will only have observably false consequences in situations we will rarely or never experience (consider NM in the eighteenth century). In cases such as these, even though agents' beliefs are false, it is not the falsehood of the claims that plays a role in limiting degrees of understanding. The limitation in the degrees of understanding is due to limitations of skill, know-how and technology available. With these limitations, one's understanding is rarely hindered by the fact that the claims are false. When these limitations are removed, the truth once again makes a difference to understanding.

This also illustrates the importance of clarifying the topic at issue. While cognitive limitations prevent the students in the example from using GR to answer questions about celestial motion, the student who believes in GR can still answer many questions about the theory itself. While NM did not produce detectably false predictions (aside from a couple of puzzles) in the eighteenth century, it was still a false theory. Thus, if the topic is the general laws of nature rather than simple motion, then the person using NM no longer has the advantage over the person using GR, even if NM is usually more useful.

The problem with false beliefs is that they cannot be entirely isolated. Even useful falsehoods can only have a restricted usefulness in finding information about the world. A measure of understanding is the ability to answer questions correctly, and any false belief will suggest some false answers. For some false beliefs, the restrictions might be insignificant—like NM in the 18th century—but there are restrictions nonetheless. If one continues to use an approximation or idealization in a situation where greater precision is

needed, then failure is the expected outcome.

For this reason, one has an advantage if one realizes that an approximation is not strictly speaking true, or that an idealization is merely an idealization. To summarize the lessons of these examples and how they relate to each measure of understanding, the following list ranks states of belief and skill in terms of their degree of understanding, holding cognitive capacities fixed. The first has the greatest understanding, the last has the least.

1. One has only true beliefs and the ability to use them effectively to solve problems;

2. One has true beliefs that one cannot effectively reason from, but one also has access to useful falsehoods that one can use to solve problems, with the knowledge that they are falsehoods;

3. One has false beliefs that one can use to solve problems.

4. One has true beliefs that one cannot use to solve problems.

5. One has false beliefs that are not useful.

State 1 occurs when one has true beliefs about a topic and can reason from those beliefs directly. State 2 occurs when one has true beliefs about a topic, but cannot reason from them directly (for example being unable to use the equations of GR), yet one knows of an approximation or idealization that one can use to solve problems and answer questions. This is a lesser degree of understanding than 1, since with 1, in addition to the ability to use approximations, one has the ability to answer further questions about how to derive solutions to problems from first principles and one has greater skill in reasoning from these beliefs. With 2 one cannot answer these questions, relying as one does on useful falsehoods to make derivations rather than being able to derive from first principles, and lacks the skill to reason from the beliefs.

States 3 and 4 are the two states that are at issue in the counterexamples to Covariance. State 3 is a state of greater understanding than 4 due to the possession of useful false beliefs.

In these terms, the counterexamples in the previous section show that 3 is correctly placed above 4, which demonstrates the failure of Covariance.

Finally, state 5 is a state of pure error. At least in state 4, one can correctly answer the questions about the true beliefs one has, though one cannot infer anything important from those beliefs. One in state 5 lacks even this ability since the beliefs are false.

The hierarchy shows that we do not need to deny the importance of truth in order to deny the Covariance Condition. The highest degree of understanding results from true beliefs paired with the ability to reason from those beliefs. Truth is important to understanding, but there can be trade-offs due to our finite cognitive capacity. Sometimes we can trade off accuracy for efficacy.

4.5 Covariance and factivity

The Covariance Condition is not only interesting in its own right. This section shows that if the Covariance Condition fails, then understanding is not factive, in the sense of 'factive' used in recent debates in (Kvanvig 2003), (de Regt and Dieks 2005), (Elgin 2007), (Pritchard 2008), (Riggs 2009), and elsewhere. In this debate, the relaxed version of the Strict Truth Condition is a condition known as 'factivity'—first proposed in (Kvanvig 2003):

Factivity Understanding X is *factive* just in case understanding X implies that all of one's central beliefs about X and most of one's peripheral beliefs about X are true.

Factivity avoids the problem of over-strictness that afflicts the Strict Truth Condition. Sally, with her single false bat belief, can understand the bat circulatory system without being a counterexample to factivity. Unlike the Covariance Condition, however, factivity treats understanding as an all-or-nothing matter. It does not address the fact that understanding comes in degrees. Nevertheless, with suitable assumptions the failure of the Covariance Condition implies the failure of factivity.

First, factivity allows that one can understand despite some minor false beliefs, a concession made to avoid being excessively strict. However, it is against the spirit of factivity for false beliefs to benefit understanding. As Kvanvig puts it in his defense of factivity, "When the falsehoods are peripheral, we can ascribe understanding based on the rest of the information grasped that is true and contains no falsehoods. In such a case, the false beliefs are not part of the understanding the person has," (Kvanvig 2003, p. 201). According to factivity, false beliefs are deviations from the ideal; they are analogous to the microscopic deviations from ideal flatness of a flat table. These deviations do not contribute to the flatness of the table; rather, they are imperfections that we can ignore. Likewise, one can understand despite minor false beliefs, for they are imperfections that we can ignore. However, the false beliefs are imperfections that cannot contribute to understanding any more than microscopic bumps contribute to a table's flatness. Therefore, if false beliefs that are implicated in understanding can increase understanding, then factivity falls. In this way the Covariance Condition is an extension of factivity that accounts for degrees of understanding.

We can also show that failure of Covariance implies failure of factivity by relating degrees of understanding to all-or-nothing understanding. The Covariance Condition relates changes in degrees of understanding with changes in degrees of veracity: they must vary in the same direction. One understands in an all-or-nothing sense when one's degree of understanding is above some threshold. If Covariance does not hold, problems emerge at the threshold.

Given that there is a threshold, it is possible for there to be an agent who does not understand but has a degree of understanding exactly at the threshold. (This assumes that understanding requires being above the threshold. An analogous argument can be given if we assume that one at the threshold understands. In that case, assume the agent gains a true belief that removes her understanding. The same conclusions follow.) If Covariance does not hold, the addition of a false belief that is implicated in her understanding can increase

her degree of understanding, putting her above the threshold. The only way this could be avoided would be if every violation of Covariance occurred away from the threshold. While this is not logically impossible, the widespread counterexamples above render it sufficiently unlikely to discount the possibility. Furthermore, if, as thresholds for flatness and tallness suggest, the threshold for understanding is context-dependent, the hypothesis that Covariance violations never occur at the threshold is even less plausible. In Chapter 3, I argued that the threshold for understanding is dependent on context. In other words, merely acquiring a false belief that is implicated in understanding can take her from a state of not understanding to a state of understanding. This violates factivity.

In short, factivity is the claim that understanding must be based on true beliefs. While this deals with understanding as an all-or-nothing state, Covariance is the natural extension of factivity to understanding as a graded state. Further, due to the relationship between thresholds and degrees of understanding, violations of Covariance will produce violations of factivity. Thus, the arguments presented above not only show that understanding violates Covariance, they show that understanding is not factive.[19]

Finally, while failure of Covariance implies failure of factivity, the converse is not true. For example, (Riggs 2009) claims that one can understand a topic even if some of one's central beliefs about the topic are false. Thus, if he is correct, factivity is false. However, he still accepts Covariance, claiming that the threshold for understanding can be low enough to violate factivity, but veracity and understanding vary together. The denial of Covariance is thus a stronger claim than the denial of factivity: there are theories of understanding that violate factivity but not Covariance, while the argument above shows that no plausible

[19]There may be a response Kvanvig could give to the above argument. His primary concern in (Kvanvig 2003) is to find a state that is like knowledge but that is not subject to the value problem, the problem that knowledge is not more valuable than its constituents. Thus, he could claim that what he calls 'understanding'—the state of having a tightly interconnected body of true beliefs—is such a state, though it is not what we normally refer to as 'understanding.' Such a response may not succeed, however. It is not clear that interconnectedness in itself provides a distinctive epistemic value, unless the interconnectedness allows for better ability to apply those beliefs. Application is what the measures appealed to in my argument above capture, which suggests that understanding of a nonfactive variety is the epistemically valuable state that Kvanvig seeks. There is not space to fully adjudicate this matter here, but these considerations are suggestive.

theory of understanding can violate Covariance without violating factivity.

4.5.1 Elgin's argument against factivity

The most detailed argument against the factivity of understanding so far is given by Catherine Elgin in (Elgin 2007) and (Elgin 2009). Unfortunately, her argument is subject to serious objections. I will now describe her argument and show that it faces objections that the argument presented in this chapter avoids.

Elgin's counterexample to factivity is the use of idealization in science. While Elgin discusses examples similar to those I present in Section 3.1, examples she refers to as "felicitious falsehoods," she does not employ these as her counterexamples to factivity. She allows for the possibility that the false beliefs in cases of approximation only contribute to honorary understanding, not genuine understanding, and they do so only in virtue of the fact that belief in approximations was part of a process that eventually led to accurate beliefs. We had to go through stages of false beliefs to achieve true understanding (e.g. we had to first discover NM before we could get to GR), but only the final state is genuine understanding (Elgin 2007, p. 37). The argument she presents is thus neutral on whether approximations or past false theories are counterexamples to factivity.

Her counterexample is idealization. She claims idealizations, though false, function as exemplars of relevant features of the system under study, "affording epistemic access to the features and making their significance manifest," (p. 41). Some theories are so complex we cannot recognize some of their significant features, while the simplifications of idealizations can bring these features out. By exemplifying relevant features, idealizations allow us to better grasp and unify information about a theory and, if the theory is true, to better grasp and unify information about the world.

For Elgin, this does not yet show that idealizations provide understanding, for her argument depends on showing that they are not just felicitous falsehoods. To demonstrate that they are not, she appeals to a second premise: that the falsehood of idealizations is not a de-

fect. In defense of this, she argues, "The ideal gas is a fiction ..." (p. 40). "Fictive sentences neither are nor purport to be true. They function in other ways. It is no defect in ideal gas descriptions that there are no gases that instantiate them" (p. 41).Thus, idealizations are false, but are a central part of improving understanding, in violation of factivity.

This argument faces objections. First of all, as Elgin herself says, the use of idealization need not involve false beliefs. In fact, Elgin emphasizes that in her example idealizations are knowingly used as fictions to illuminate a true theory. Thus, her example of idealization in science is not an example of false beliefs contributing to understanding—scientists don't believe that the idealizations are true—and thus is not a counterexample to factivity.[20]

Similarly, Elgin's argument depends on the claim that idealizations work by exemplification. (Strevens 2013) objects to this, claiming that the use of idealizations is governed by different standards of correctness. When something is treated as an idealization, it is understood that we need a "translation manual" to determine what the idealization actually says about the world. Further, he claims, successful sciences uses idealizations whose translations are true. That is when they produce understanding. Thus, if he is correct, Elgin's argument depends on a mistaken account of idealizations. The use of idealizations does not involve false beliefs, because when properly understood idealizations are not false claims about the world. Idealizations are no longer counterexamples to factivity.

Finally, (Mizrahi 2012) argues that scientists continually improve their idealizations, rejecting old ones in favor of new ones that work better. This suggests that, just like "felicitous falsehoods," idealizations are intermediate steps of falsehood that allow us to work eventually to true theories. If Elgin is neutral about felicitous falsehoods, by parallel reasoning she should be neutral about whether idealizations are a counterexample to factivity. It may only provide honorary understanding in virtue of leading us toward more accurate theories.

The argument in this paper avoids these objections. The counterexamples I provide do

[20]A similar point is made in (Mizrahi 2012, p. 247).

not rely on controversial claims about how idealization works in science. The counterexamples are not restricted to scientists who know that they are using idealizations as fictions, and thus lack false beliefs about the topic; they involve agents with straightforwardly false beliefs. The measures of understanding show that some felicitous falsehoods can improve understanding. Thus, my argument is not subject to the objections Elgin faces, and factivity remains overturned.[21]

4.5.2 Riggs' argument against factivity

(Riggs 2009) argues against factivity based on a case that he claims elicits the intuition that factivity is too strict. The case is this: Wayne's wife is afraid of water, and he understands her fear of water based on the belief that she was in a boating accident at the age of three. However, she was never in a boating accident. Since his false belief about the supposed accident is almost all he believes about her fear of water, it is a central false belief about the topic of her fear of water. Riggs claims that intuitively, despite the falsehood, Wayne still understands his wife's fear of water. The false belief allows him to understand the nature of the fear, if not its origin.

This argument relies on intuitions about the case; Riggs does not present a metric by which we can measure whether it is a case of understanding. Unfortunately, the intuition that forms the basis of the counterexample is controversial. It is not clear that Wayne understands his wife's fear of water based on completely false beliefs about a childhood accident. Since she is his wife, it is likely that he has witnessed the effects of her fear and spoken to her about it, so these true beliefs, rather than the false belief about an accident, may be central. If the example were changed to exclude additional true beliefs, it is less plausible that he understands her fear. It is still plausible to conclude that he does not understand her fear of water, though there is some intuitive pull in the other direction.

Further, it is not clear whether the false belief is implicated in understanding her fear of

[21]This is not to say that my argument is in conflict with Elgin's, only that my argument goes further in demonstrating counterexamples to factivity.

water. Unlike NM, which is used in many calculations of celestial motion and thus clearly plays a role in understanding, it is plausible that Wayne's belief about a boating accident plays a role only by allowing him to infer that his wife is deeply afraid of water. This latter belief is true. The conclusions that require inference from the false belief will largely be false, such as the conclusion that a boating accident was the precipitating event. Without a more precise way of evaluating the case, it thus appears inconclusive.

On the other hand, the argument in this paper avoids appealing to controversial intuitions about specific cases. The argument above is based on measures with which to test cases, and the cases are not isolated but are broad classes of counterexamples. The specific measures of understanding presented above are independently motivated. They are the measures that we actually use when we need to test for understanding in the world. The argument above avoids the limitations of Riggs' argument.

The argument in this paper not only refutes Covariance, but shows that factivity fails using arguments that avoid the pitfalls of existing arguments.

4.6 Falsehood is here to stay

One person's modus ponens is another's modus tollens. One could take the above arguments to show that the measures of understanding, while often good enough for many purposes, are inaccurate in some cases. In particular, the cases above may be due to a breakdown in these methods of measuring degrees of understanding. In that case, the above arguments may just reveal places where these measures go systematically wrong; they do not reveal the failure of Covariance or factivity. In this section, I resist this line of reasoning and argue that there is good reason not to impose the Covariance Condition.

First of all, I have already defended the management account of understanding in detail in Chapter 3. There is good reason to think that it is correct. It will take a powerful argument to overcome that. I claim that there is no such argument, and the weight of reasons is against the Covariance Condition.

The key observation is that understanding can be applied. One who understands has the ability to produce explanations, or the ability to apply knowledge to cases, or the ability to sort the relevant from the irrelevant. This is what makes the measures compelling; it is why we use them to test students, job applicants, and psychology test subjects. It is anathema to the very idea of understanding that it could simply be idle. One could not at the same time understand physics while lacking the ability to do anything with physics. That is what the measures capture, the fact that understanding translates into ability, either by being an ability itself or by being the kind of state that is conducive to producing an ability given our cognitive make-up.[22] The need to apply understanding motivates the measures used in this paper, and the measures show that the core features of understanding that give it its value do not imply Covariance.

Furthermore, the conclusion above already allows for the importance of the truth for understanding; one must be able to find correct answers. Having true beliefs contributes to that, securing a role for truth in understanding. An additional Covariance or factivity requirements would be ad hoc.

In addition, we should expect violations of Covariance. The close tie between understanding and the ability to find the truth suggests the possibility of trade-offs. One ought to expect the possibility of trade-offs, where a decrease in truth can be compensated by benefits elsewhere to produce an increase in the ability. While knowledge requires truth because knowledge means having the truth, understanding gives the ability to find answers and reason well, which does not imply already having the answers. This explains why there is no parallel truth requirement for understanding. The conclusion that one can gain understanding via false beliefs does not overturn a defining characteristic of understanding, as it would for knowledge. There is no need to resist the conclusion that Covariance fails.

Even if understanding is not an ability (though the management account implies that it is), the fact that understanding comes in degrees opens the possibility that there can be

[22]For arguments that understanding is an ability, see (de Regt and Dieks 2005), (de Regt 2009), (Grimm 2010), (Wilkenfeld 2013), among others.

trade-offs between truth and other factors conducive to understanding. Some have claimed that beliefs constitute understanding when there is a web of strong epistemic connections among the beliefs, as in (Kvanvig 2003). Such an account leaves open the possibility that a loss in veracity can be compensated by an increase in connections (and these connections improve one's ability to find correct answers).

Finally, the denial of Covariance does not eliminate truth from playing an important role in understanding, for it is only at intermediate steps of reasoning to a conclusion that false beliefs can increase understanding. The ability to answer questions correctly means finding true beliefs, for that is how understanding puts us in contact with reality. Though one can draw upon falsehoods to do this, truth must remain in the picture. In short, there is no rationale for imposing the Covariance Condition.

4.7 Conclusion

Understanding can be increased by the acquisition of false beliefs; understanding is not factive. This undermines claims that understanding is simply a kind of *knowledge that*. Insofar as understanding is taken as an essential epistemic goal, it casts doubt on the claim that truth is the only epistemic good. While true belief still plays a role in understanding, sometimes the right false belief is better for understanding than the truth.

CHAPTER 5

UNDERSTANDING BY LUCK

5.1 Introduction

I have defended the importance of understanding for epistemology. Acquiring understanding is epistemically valuable. Acquiring understanding is an achievement. Since it is not an achievement to acquire something due to luck, this suggests that understanding cannot be acquired due to luck.

Not only that, but in the decades since the introduction of Gettier problems, epistemology has reached a consensus that knowledge cannot be acquired via luck. Understanding is generally taken to be a more robust epistemic success than knowledge; even those who claim that understanding is merely a kind of knowledge allow that it is deep, robust knowledge (e.g. knowledge of a true explanation). This suggests that understanding likewise cannot be acquired via luck.

However, in this chapter I argue that understanding can in fact be acquire by luck.[1] The key to the argument presented here is the fact that understanding is an ability or skill, rather than merely a kind of belief. In short, the argument is that the origins of a skill do not undermine the skill itself. Thus, as long as one has the cognitive skills that are characteristic of understanding, it does not matter how they arose; one still understands. The skill could arise from the luckiest miracle of all time, and it would still be understanding.

While the claim that understanding is a skill is controversial, the management account

[1] For others who argue that understanding is compatible with luck, see for example (Kvanvig 2003, ch. 8) and (Morris 2012).

of understanding, defended in earlier chapters, implies that understanding is an ability, specifically the ability to manage cognitive resources. In addition, arguments based on other accounts of understanding also support the conclusion that understanding is an ability, for example (de Regt and Dieks 2005), (de Regt 2009), (Grimm 2010), and (Wilkenfeld 2013). If any of these views of understanding is true, then understanding is an ability and the argument thus shows it can be due to luck.

In addition to showing that understanding can be due to luck, we must consider why this is a surprising conclusion. The reason it might be difficult to accept this conclusion is multi-faceted, but two reasons stand out. The first reason is that in real life cases understanding is almost always an achievement not due to luck. While understanding is not essentially an achievement and is logically compatible with acquisition by luck, this only rarely happens. Thus, our experience and familiarity with understanding can suggest the mistaken conclusion that understanding cannot be due to luck.

The second reason is that understanding is what allows us to avoid luck in the rest of our epistemic lives, though understanding itself can be acquired through luck. Understanding is where luck ends. The operation of understanding is incompatible with luck, though its acquisition is not. We must be careful when assigning exactly where and how luck can and cannot be present with regard to understanding.

Section 2 begins with a discussion of luck and knowledge, since luck has been studied in much more detail in relation to knowledge. In Section 3, I adapt the varieties of luck as they relate to knowledge to explain the varieties of luck that are at issue with regards to understanding. Section 4 presents the argument from the claim that understanding is a skill to the claim that understanding can be acquired by luck. Some might consider this argument a reason to reject the claim that understanding is a skill. However, in the remaining sections, I argue that the consequences of accepting this conclusion are not as bad as one might suppose. Section 5 explains how understanding remains connected to the facts even if it is due to luck. Section 6 addresses the worry that luck is incompatible with

achievement, and understanding is an achievement. Section 7 argues that understanding is where many kinds of luck stops, so it can be due to luck, but it prevents that which is based on it from being due to luck. Sections 8 and 9 respond to objections.

5.2 Knowledge and luck

The issue of luck in relation to knowledge has been well-studied in the decades since Gettier.[2] While there are some forms of luck that are universally considered to be incompatible with knowledge, there are other forms of luck that are universally taken to be compatible with knowledge. This section surveys luck and its relationship to knowledge to provide a guide to which kinds of luck might be relevant when it comes to understanding.

To begin, when a true belief that p is formed, it was formed using some process or method. If that process or method is unreliable and is unlikely to lead to true beliefs, then the fact that the method was used and led to a true belief is a matter of luck. (Pritchard 2005) refers to this as *veritic luck*.[3] This form of luck is inconsistent with knowledge. Methods and processes that only by luck lead to the truth are not knowledge-producing. Knowledge requires being directed successfully at the truth.

One special case of veritic luck is Gettier luck. In a Gettier case, one's belief is based on a reliable method that normally makes the belief likely to be true, but the factors that make the process reliable are not in play in this situation. The reliable connection between belief and the fact of the matter is broken. Yet by luck, the belief turns out to be true. There is a fortuitous compensation for a break between belief-forming process and the truth. A classic example: one is in a land full of convincing fake barns. One looks at the only real barn and form the belief that it is a barn based on its appearance. Normally this is a reliable process, but in these circumstances it is not. Yet by luck, the belief is true anyway. This

[2]For an in-depth examination of the varieties of luck in the epistemology of knowledge, see (Pritchard 2005).

[3]Pritchard's definition of veritic luck occurs when one uses a process P that in the actual world leads to the truth, but that in all nearby possible worlds leads to false beliefs. I prefer to leave open whether this is the best analysis of epistemic reliability.

form of luck is widely taken to be incompatible with knowledge.[4]

A belief-forming method need not be unreliable for a belief formed by it to be true by luck, however. Even if the method is reliable, if the agent has good reason to think the method is unreliable in the circumstances, then there is still a sense in which the belief is true by luck. The truth of the belief was not foreseen by the agent. The agent bore no responsibility for its occurrence and is not responsible for its occurrence. (Pritchard 2005) refers to this as *reflective luck*.[5] This form of luck is also arguably incompatible with knowledge (though some forms of reliabilism might allow it to count as knowledge as long as the process is actually reliable).

Veritic and reflective luck have to do with whether the belief-forming method is likely to lead to the truth. Other forms of luck have to do with the circumstances. It can be improbable that one would find oneself in the circumstance of being in a position to learn that p. One can gain a true belief by luck when, against the odds, one happens to be in a position to acquire information. For example, suppose in a drawing with thousands of entries, Ann alone wins the grand prize of a trip to Hawaii. Ann travels to Hawaii, and upon entering her hotel room, she happens to drop her keys and they happen to slide under a cabinet. When she looks for her keys, she finds a famous stolen painting hidden under the cabinet. It is entirely a matter of luck that Ann came to have the true belief that the painting was under that cabinet in that room at that time. However, that does not prevent her from knowing that the painting was there. She was lucky to be in a position where that information was available to her, but once she was in that position, it was not merely a matter of luck that the belief was true. Luck of being in the right place at the right time, or *positional luck* for short, is compatible with knowledge.

It can also be a matter of luck that one has any true beliefs at all. Suppose one is

[4]The Gettier problem is not universally accepted as a problem. For example, (Weatherson 2003) questions whether theories of knowledge need to account for Gettier cases, and (Sosa 2001) allows for a kind of knowledge that agents in Gettier cases possess.

[5]Pritchard defines reflective luck as when one uses a process that leads to the truth in the actual world but in nearly all possible worlds consistent with what one can know by reflection alone, if one formed that belief it would be false.

standing on the slope of a mountain when a rockslide occurs. All around, millions of tons of boulders crash down the mountain. Luckily, one is not struck by any of them, though many miss by mere millimeters. After this, one is lucky to be alive to believe anything at all, let alone anything true. One would have no true beliefs if not for the amazing luck of surviving the rock fall. However, one still knows what one knew before (leaving aside forgetting). This is another variety of positional luck that does not prevent one's beliefs from being knowledge.

A special case of positional luck has to do with the formation of our reliable cognitive systems in the first place. Suppose it turns out that intelligence was only able to evolve due to a series of unlikely coincidences. In that case, it would be a matter of luck that we had any beliefs, and thus that we had any true beliefs. However, this would not preclude knowledge, according to most theories of knowledge. Positional luck, luckily being in a position to gain true beliefs, is compatible with knowledge.

This is how matters stand with knowledge. The question, then, is whether analogous kinds of luck are compatible with understanding.

5.3 Understanding and luck

The switch from knowledge to understanding introduces some complications. True belief is a component of knowledge, but not identical to knowledge. We can consider the ways in which a belief can be true by luck without assuming the belief constitutes knowledge. We can ask if a true belief acquired by luck can be knowledge. However, as I have argued in previous chapters, understanding is not a species of true belief. The question whether one can acquire understanding by luck is not only a matter of whether luckily true beliefs are consistent with understanding. Understanding is above all an ability.[6] We must ask

[6]In previous chapters, I present and defend the management account, which implies that understanding is an ability. (de Regt and Dieks 2005), (de Regt 2009), (Grimm 2010), (Wilkenfeld 2013) and others also argue that understanding is an ability. For arguments against this claim, see (Trout 2002) and (Khalifa 2012), and for replies see Chapters 2 and 3 of this dissertation. For the remainder of the chapter, I will assume that understanding is an ability.

whether acquiring the ability characteristic of understanding—the ability to extract and exploit relevant information to answer questions—by luck is compatible with understanding. This introduces new kinds of luck.

To start, though, understanding could be due to luck if understanding requires true beliefs and the beliefs are true by luck. If there is such a truth requirement, then we must consider whether any of the forms of luck identified in the previous section are compatible with understanding. We must determine whether understanding does require true beliefs, and then whether it is compatible with acquiring true beliefs by luck.

In addition, even supposing that understanding can be based on false beliefs (as argued in the previous chapter), that does not imply that understanding can be due to luck. Not just any false belief can produce understanding; only the right kind of false belief is conducive to understanding. Thus, understanding could be incompatible with luck in the sense that one must acquire understanding-conducive beliefs by design, not by luck.

To illustrate, consider the physics expert who reliably finds understanding-conducive approximations and idealizations to apply. The fact that she can use strictly false approximations and idealizations to good effect is not an accident; it is something she has the skill to do. If understanding requires this, then it is incompatible with some forms of luck.

One also may be in a position to suspect that acquiring a given belief will increase understanding, without having a clear sense of whether the beliefs are true.[7] This shows that the truth of the beliefs that play a role in understanding may not be the best place to look to determine whether understanding can be acquired by luck. Since more than truth may be at issue, the matter requires some work. The truth of the beliefs involved is orthogonal to (or at least linearly independent of) whether understanding can be due to luck. Thus, we cannot simply consider the kinds of truth-related luck that are important to debates about knowledge; we need to find something analogous.

Since understanding is an ability, in addition to truth we must question in which ways

[7] If one is concerned about apparent implications of doxastic voluntarism, substitute 'accept' for 'believe' in the description.

an ability might be acquired due to luck. What is the issue regarding whether understanding is compatible with luck? Having the ability to Φ means having a special kind of disposition to Φ. Dispositions are thin. An ability to Φ is a stable set of dispositions that allow one, in the right circumstances, to Φ. To illustrate, one can have the disposition to punch by virtue of having muscle spasms, but this falls short of having punching skills. That requires that the disposition is a stable character trait tied to one's agency. This introduces a potential layer of luck: when one has a disposition to Φ due to luck, can that qualify as the ability to Φ?

Understanding is the ability to find what is relevant to answer questions. Is gaining the disposition to find what is relevant due to luck incompatible with having the relevant ability that is understanding? To be neutral, when there is doubt about whether one genuinely has an ability, I will refer to it as a disposition, with the understanding that it is the kind of disposition that would be an ability if it had the right origin. The disposition characteristic of understanding is the disposition to extract and exploit the relevant information to answer questions. The question is whether one can acquire this by luck and thereby understand.

First, there are analogues to the forms of positional luck discussed in the previous section. For example, suppose the person who won the trip to Hawaii encountered something that gave her the disposition characteristic of understanding, and which they would not have encountered anywhere else. This would be acquiring the disposition by luck, by fortuitously being in a position to gain the disposition.

Also, when one survives a rock-slide by luck, then the fact that one has any dispositions at all is a matter of luck. As an example of a different kind of positional luck, suppose Mary is one of 100 equally qualified candidates to work on a top secret military guidance system, learning from the world's only experts on it. To fill the position, they pick her at random. As a result, she comes to have the disposition characteristic of understanding relating to top secret military guidance systems. It is in a sense due to luck that she has this disposition, since without being picked at random, she would not have acquired it. The same general

kind of luck is at play if it is only due to luck that we evolved to have cognitive systems capable of the dispositions characteristic of luck.

Another way in which the formation of the relevant disposition could be due to luck is if the disposition characteristic of understanding is formed by a process that almost never leads to that disposition (analogous to veritic luck). In all nearby possible worlds, the process leads to dispositions that are not characteristic of understanding, or to abilities that are not understanding-like. However, this time it did. Call this *process luck*. For example, suppose that one disposition conducive to understanding is the disposition to recognize goats as mammals. If one acquires the disposition to infer *mammal* from *goat* based on the fact that those words always happened to be next to each other in a list of random words, it is only a matter of luck that one gained a disposition to answer questions correctly. In nearby possible worlds, forming dispositions based on words being near each other does not lead to helpful dispositions. (On the other hand, if one gains the disposition while trying to learn about the world and skilfully gaining the disposition through a process that filters out spurious correlations, then one the acquisition of the disposition that is helpful is not a matter of luck.)

The analogue of Gettier luck for understanding occurs when a process that reliably leads to understanding is disrupted, but by luck the process leads to the characteristic disposition of understanding anyway. The result of the process is the disposition to answer questions correctly, but the reliability of the method is not responsible for the disposition having the right features. For example, a physicist makes useful approximations using her skill to find understanding-conducive approximations, but this is an anomolous exception where the normal method of approximation inexplicably does not work. However, another approximation does work. In this case, the two approximations happen to give the same result, though in general they do not. If the formation of a skill is at issue, a Gettier case would involve using a method that normally produces a skill, but this time the conditions are wrong for it to work, but it does anyway. Collectively, the beliefs and dispositions that

contribute to understanding constitute *understanding-conducive* states.

The relevant form of luck is forming the dispositions characteristic of understanding through means that are not connected to that which makes it understanding in successful cases. This generalizes the notion veritic luck to cover understanding. Since Gettier luck is less extreme than non-Gettier process luck (in Gettier cases at least something goes right on the agent's end), we will focus on non-Gettier instances of process luck. If understanding is compatible with process luck then it is compatible with Gettier luck.

Is it possible to gain understanding through a process that only by chance produces understanding-conducive states? Could one undertake to use a method that almost certainly will not produce anything like understanding, yet thereby come to acquire understanding? Is it possible to gain understanding using a method whose success is completely outside of one's own responsibility? I will now argue that in all of these ways, understanding can be due to luck.

5.4 Gaining understanding by luck

Let us see whether understanding can be based on luck. First of all, understanding is compatible with beliefs that are true by any form of luck identified in Section 5.2. As shown in Chapter 4, understanding can be based partly on false beliefs.[8] Understanding is consistent not only with beliefs that aren't true by design, but it is also consistent with beliefs that aren't true at all. Beliefs that are true merely due to luck do not preclude understanding.[9] Understanding is compatible with veritic, Gettier and reflective luck.

Further, understanding is compatible with the various forms of positional luck. Being in a position to develop the relevant dispositions only due to luck does not make the dispositions any less genuine understanding. The disposition to answer questions correctly can

[8]The argument in Chapter 4 assumes three different abilities serve as accurate measures of degrees of understanding. However, even if understanding is constituted entirely by beliefs, there is reason to think those beliefs need not all be true, as argued in (Elgin 2007), (Elgin 2009), (de Regt and Dieks 2005), (Riggs 2009) and elsewhere.

[9]This does not necessarily imply that one can have understanding despite having no beliefs that are true by design, only that understanding is compatible with the presence of some lucky beliefs.

be as stable as you like, consistent with positional luck. Understanding is compatible with by luck finding oneself in a position to have or gain understanding.

It remains to be determined whether understanding is compatible with the characteristic disposition arising from unreliable means. We can show that understanding is compatible with luck solely from the thesis that understanding is a cognitive ability; we need not settle anything else about the nature of understanding.

In previous chapters I have presented and defended a detailed account of understanding, namely the management account.

Management Account S understands X when S has the ability to manage cognitive resources to answer questions in the subject matter of X.

According to the account, one's degree of understanding is the degree to which on has the ability to manage one's cognitive resources. Beliefs play a role in this ability by serving as the basis for inferences, recall and so forth, but a great deal of managing cognitive resources is a matter of deploying the right inferences and attention at the right time. At a fundamental level, understanding is an ability.

The argument that understanding is compatible with luck is simple: one has the ability to do X if one has a stable disposition to do X through one's agency. Whether one can do X is independent of how one acquired the dispositions that lead one to do X. Witnessing a person lift a box shows she has the ability to lift a box, without any information about the origin of that ability. This is the short version; let us look in more detail.

5.4.1 Luck and abilities in general

Rather than drawing upon the details of the management account of understanding, let us focus on just the claim that understanding is an ability. Abilities in general can be acquired by luck. As illustration, suppose John sees an advertisement in the newspaper proclaiming, "Learn to pay the violin like a virtuoso from world-renowned teacher Betty 'The Blade' Walker!" John signs up for lessons, not realizing that Walker knows nothing about playing

the violin, does not know how to read music, and has never even tried to play the violin. Yet through one of the most miraculous flukes of all time, the instructions that Betty makes up on the fly are effective in giving John the ability to play the violin well. He can effectively read music and play pieces of music like a professional.

Despite the luck that was required for John to end up in this state, he does in fact have the ability and skill to play the violin. It does not matter that the method he used—namely following the instructions of someone who has no idea what she is doing—is unreliable and only successful due to luck. It does not matter that the success of the lessons had nothing to do with his agency. While there might be certain negative judgments one can make about John's ability in this case (perhaps that John does not deserve to have the ability), he undoubtedly has the ability. Given his ability, the origin of it does not affect the degree of skill. Someone whose violin-manipulating dispositions were identical to John's but whose dispositions were gained from a genuine violin teacher would not be more skilled. Luck makes no difference.

However, one might object that we are willing to grant that John has an ability only because his success was not entirely due to luck. He acquired the ability to play the violin using a method that he believed would work, and furthermore, he was justified in so believing. He was justified in believing that what Betty was doing would teach him how to play the violin. After all, he saw her ad in the paper claiming she could teach the instrument. She acted like she could teach and play it. She sounded like she knew what she was talking about (she would have to if her lessons actually gave John the ability to play the violin). He was justified in believing he was taking reasonable steps to achieve his goal. Therefore, it was due to his rational choices, not luck, that he learned to play.

This response is limited, however, for we can change the example to make it moot. Assume that the ad in the paper that leads John to Walker is dubious to begin with. Assume that before he goes to his first lesson, his friends tell him they have gone to her and found she is a fraud. Then, when John arrives for his first lesson, Walker tells him that she has

no idea what she is doing. Even with all of that, if he follows through with the lessons and ends up with the disposition to make the right movements and to interpret the notes on a page in a way that allows successful playing of the violin, then he has the ability to play the violin, even though he was never justified in thinking that the lessons would work.

The general point about skill acquired by luck does not require a teacher. Suppose that Mary's plane crashes in China (at a time when not so many people there speak English) and she needs to learn how to speak Mandarin. She has no idea how to speak it, but she makes up her own training regime anyway, expecting that it will fail. 'As long as I'm trying,' she thinks to herself. Yet, somehow, she happens to train herself to speak Mandarin. It is just a fluke. This still does not undermine the fact that she has the ability to speak the language. She has a robust, underlying capacity to carry out the function. Abilities are perfectly indifferent to whether they were acquired through luck.

This thesis is true of abilities in general. An ability is a synchronic disposition of an agent. The origin of the disposition does not prevent it from being an ability. Thus, if understanding is in fact a cognitive ability, then its origin does not affect whether or not an agent understands. It does not matter whether the ability characteristic of understanding was acquired through dedicated training or luck. It is still understanding, for the ability identical with understanding is there.

5.4.2 Luck, ability and understanding

At this point, we turn our attention to understanding and apply these lessons there. Consider an adapted version of the Comanche case from (Kvanvig 2003). Suppose Sally reads a book about Henry VIII that she knows was not based on scholarship but was simply made up (perhaps she knows the author well enough to be sure that the claims were just made up). After reading the book, Sally remembers most of what the author wrote in it. She thinks about it in detail and develops a highly complex understanding-like set of dispositions to answer questions about everything in the book. However, over time, she forgets that these

details were made up and falsely "remembers" them as facts about Henry VIII. Yet the beliefs nevertheless turn out to be true, or at least enough of them are true that Sally's disposition to answer questions about Henry VIII correctly matches understanding. Her beliefs were badly formed, and only true by an incredible stroke of luck. However, this does not undermine Sally's abilities. Given that understanding is an ability, she understands Henry VIII.

According to the management account, understanding is an ability, and we have shown that an ability can be acquired by luck. Thus, understanding can be acquired by luck. The Henry VIII example above can be filled out to give an example of this: all we need to suppose is that based on the beliefs acquired from reading the book, Sally has the ability to manage one's cognitive resources to find useful information about Henry VIII. One can effectively sort the relevant from the irrelevant. The fact that the beliefs arose from a completely unreliable method do not undermine the presence of the ability.

In addition to lucky guesses, one can understand on the basis of lucky biology. For example, suppose Erin is electrocuted, and the resulting trauma affects her cognitive abilities. Miraculously, the organization of her capacities now allows her to effectively cut right to the relevant details in tax policy. Give her a problem about tax policy to solve, and she can reliably pick the right beliefs and inferences to solve the problem.

The management account implies that Erin understands: by supposition she has the ability to manage cognitive resources in the subject of tax policy, and according to the management account, that is sufficient for understanding (more accurately, if her ability is above some threshold, then her degree of understanding is above the threshold for understanding in the all-or-nothing sense). Thus, assuming the management account is correct, the origin of her ability plays no role in determining whether she understands, even if the origin is just luck. The same is true if Erin's tax policy ability results from an innate feature of her psychology resulting from a novel mutation. According to the management account understanding is an ability that can be acquired by luck.

Thus, the management account shows us in more detail that understanding can be acquired by luck. Note that although we have referenced the management account, the case only depends on the claim that understanding is an ability, without relying on the specifics of the management account.

5.5 Connection to the facts

One might object to the idea that one could come to understand Henry VIII by forgetting the nature of a book and mistakenly believing random guesses were scholarship. This leads to the worry that the connection between the facts about Henry VIII and the disposition to answer questions correctly is not strong enough. Does understanding require a more robust connection to the facts?

One response to this objection is to compare this case of understanding Henry VIII to understanding a false theory. Consider a person who believes the George W. Bush administration carried out the 9/11 attacks. They have a detailed theory that explains it all. While this person does not understand the 9/11 attacks, they understand their own theory. This means that they are able to correctly answer questions about the theory itself. They can tell you what the theory says, what evidence it is based on, and what its implications are. Sally in the above example has at least this level of understanding of Henry VIII. She has an understanding of the story, the collection of propositions that make up our current body of knowledge about Henry VIII. She can answer questions about the internal structure of this story. Intuitively, she understands the story. We have at least one way in which Sally uncontroversially understands Henry VIII.

With this in mind, note that the internal structure and relations within a topic seem to play a larger role in understanding than does the origins of the beliefs involved.[10] The ability to work within the topic plays a bigger role than how one got the ability. In the case in question, Sally has all of those beliefs in her head and she can work with the connections

[10]Compare to the view of understanding found in (Kvanvig 2003) and (Zagzebski 2001).

among them to find the right answers in the subject matter. She has assigned them relative importance based on the internal structure of the narrative. All of these abilities are significant steps toward epistemic success. The idea that they constitute understanding is not far-fetched.

We must be careful when we describe what one allegedly understands. If one understands what the claims mean, that only requires a close connection to the claims themselves. One does have the proper connection to the claims, since that is what you believe. The only part lacking is the proper connection between the beliefs and whether they are true. There is reason to hold that understanding is a synchronic relation between an agent and a topic that cannot be undermined by infelicitious origins. The agent happens to have true beliefs and true meta-beliefs about how the first order beliefs relate to the world. It is just that those true beliefs do not amount to knowledge because of their problematic origins. However, the fact remains that she has a firm grasp of the internal connections, and has the ability to use the information going forward.

Finally, in support of the claim that Sally understands Henry VIII, she does grasp the connections among the facts about him. If given a question about him, she will be able to answer the question; the origin does not affect this. If given a new piece of information about Henry VIII, she will be able to integrate it effectively into her other beliefs; the origin of her dispositions does not affect this. Sally's performance and abilities are flawless (we can suppose). The only concern is over the origin, but given an ability like this, we can evaluate it without reference to its origin. Judgments about justification and knowledge deal with evaluations of origins of beliefs. Understanding is not the same kind of evaluation. It has to do with the cognitive abilities that we have and how much success we can expect regarding a topic.

5.6 Understanding is (usually) an achievement

Understanding is often taken to be an achievement, and the very idea of an achievement is incompatible with acquisition by luck. Understanding is heralded as the highest cognitive success, the pinnacle of cognitive achievement. When one understands something, one deserves credit and praise for doing so. Yet, one cannot deserve credit or praise for something acquired purely by luck.[11] We must reconcile the fact that understanding is an achievement with the fact that it can be acquired through luck. The way in which these two facts are reconciled is that in actuality, understanding is almost never acquired through luck. It is safe to assume that if someone understands something, it is an achievement.

Most of the situations in which it is important to determine whether someone understands, understanding by luck would be near impossible. Consider understanding why the allies won World War II. Though we lack the specific details about the cognitive architecture and processes required to answer WWII questions correctly, they must be remarkably complex, consisting of many facets that must work together. The chance that one could simply make stuff up and thereby achieve such a state are so small as not to be worth considering. This is true for all but the most narrow and trivial subjects. Thus, most of the time, when we assess whether someone understands, we either find that they have achieved understanding through non-lucky means (excluding positional luck) or that they do not understand at all. The fact that this is so common could lead one to think that it is in fact impossible to acquire understanding by luck, and that the nature of understanding disallows luck. Yet this only shows that understanding is complex, not that understanding by luck is impossible, just as the fact that we never see anyone roll 100 Yahtzees in a row does not mean it is impossible.

Consider some situations in which we are concerned about understanding. For example, you are teaching quantum mechanics to upper level undergraduate physics students. You

[11] For example, if you attempt to catch a pass in basketball, and bungle it so badly that the ball flies through the air and goes in the basket for two points, this is a good thing for your team, but not an achievement of yours. Also, see Sosa's discussion of apt and adroit performances in (Sosa 2010).

want to know whether they understand the material. Proficiency in quantum mechanics requires the knowledge of several branches of mathematics and the ability to apply them. It requires knowledge of a large number of laws and the ability to apply them it requires mastery of several methods of problem-solving, and more. This is usually the result of more than a decade of training in mathematics and science, and at least months in the specifics of quantum mechanics itself. It would be almost impossible to gain all of this through luck. Other topics where understanding is an important issue are similarly complicated, and unlikely to be due to chance. Thus, acquisition of understanding by luck is not conceptually ruled out, but it is statistically unlikely to occur in all but the simplest of cases. In most cases, the possession of understanding is due to persistent hard work and training, and is an achievement. This is not built into the very nature of understanding, however.

There is another aspect of the relationship between understanding and achievement that explains how understanding can be an achievement and still be compatible with luck. Whether understanding is due to luck or not, it is responsible for further epistemic achievement, including further understanding. Understanding is what allows us to acquire true beliefs without relying on luck, even when that understanding was the result of luck.

As an analogy, suppose someone is born and happens to be in an environment that leads him to develop remarkable flexibility. As a result, he is able to crawl down a narrow tunnel and rescue some trapped children. The rescue is an achievement due in large part to his flexibility. However, his flexibility itself is not an achievement; he just happened to be that way due to genetic and environmental luck.

Likewise, even in cases in which understanding is due to luck, the feats of cognition one performs with understanding are genuine achievements that are due to understanding. The fact that understanding is an ability implies that necessarily its proper use can produce cognitive achievements. This is what it is to be a cognitive ability. Thus, understanding is closely associated with cognitive success, even in those cases where its acquisition is not a cognitive achievement. The intuition that understanding is necessarily an achievement

can be party diagnosed as a misattribution of properties of the products of understanding to understanding itself.

One last aspect of the relationship between understanding and achievement is due to an insight of (Zagzebski 2001): part of gaining understanding is gaining the ability to recognize (at least implicitly) success. Part of acquiring understanding is being able to recognize methods that work, being able to recognize what is relevant and useful in a topic. One who understands is in a position to distinguish achievement from luck. Given sufficient introspective access, when one turns one's understanding toward oneself, one can recognize when one has achieved success, and so success is not due to luck. Again, understanding is what removes luck; there is no requirement that it itself not be due to luck.

Even if the skill of understanding is not essentially an achievement, it is still an epistemically valuable state. This explains both why the acquisition of understanding through effort is considered an achievement, and why it is not problematic that it is not essentially an achievement. Bringing about a good state through one's own effort is an achievement. As shown above, in most cases that interest us, understanding does come about through effort and is thus an achievement. The fact that some instances are not strictly speaking achievements, or at least it is possible for such to occur, does not undermine the fact that understanding is an epistemically good state.

5.7 Where luck ends

Though understanding is compatible with luck in terms of origins, there is one important sense in which understanding is incompatible with luck. A conclusion that one reaches via the use of understanding cannot be due to process luck; understanding is the ability to get things right, it is the right process. Understanding allows us to form conclusions that are true by design, not by luck. Understanding is one way in which we can avoid luck and guessing. One who understands has the ability to answer questions within a topic. Thus, when one employs understanding to answer a question, and gets it right, the fact that one

got it right is not due to luck. It is due to one's cognitive abilities. Understanding is one way to remove a kind of luck from one's belief formation (it does not remove Gettier luck, however, for in Gettier cases the reasoning used is good, luck enters later).

Understanding allows one to see the significant patterns and make the right inferences, one does not go astray. This really does appear to preclude understanding being completely due to luck: once you come to have the capacity, perhaps by luck, then you can use it correctly to identify the truth of the matter. Thus, there is no anti-luck condition for understanding, but understanding itself rules out the possibility of luck. At some level all of our cognitive capacities are not due to the function of properly functioning cognitive system recognizing the value of what is going on, unless we were designed by God (in which case we could say something similar about god).

This accounts for the intuition that understanding cannot be due to luck. Our familiarity with understanding leaves us with the strong sense that one who understands is not merely groping in the dark, randomly happening across useful information. This much is entirely correct: one who understands has the ability to get things right by skill. One who understands is familiar with how to find valuable information. However, this may tempt one to attribute the absence of luck to the origins of understanding itself. As I have argued above, this is a mistaken attribution.

5.8 Debunking arguments, defeating evidence

In response to the above considerations, one may claim that understanding is nevertheless incompatible with luck because a reasoner who has gained dispositions characteristic of understanding through luck has a more tenuous disposition. One who was lucky is subject to a kind of defeat via finding out about the lucky origins of the dispositions. This could undermine the ability. The lucky origin makes one subject to debunking arguments. For example, if Sally from the example above learns that the book about Henry VIII was entirely made up, she will cease to believe what it said, and without those beliefs, she will

no longer be able to answer questions about Henry VIII. This could be taken as another objection to understanding by luck: understanding should be more robust than this.

However, the potential for debunking arguments only makes a difference in virtue of the beliefs that agents have and inferences that they make, not whether the beliefs are due to luck. Suppose we plant misleading evidence suggesting that Sally's beliefs about Henry VIII are ultimately due to careful scholarship, rather than lucky guesses. Then suppose we find someone whose grasp of Henry VIII is actually due to careful scholarship, and we plant misleading evidence suggesting her beliefs are based on guesses. In this case, the person whose beliefs are due to luck has greater stability than the person whose beliefs are not due to luck. The actual origins do not protect the one exposed to true scholarship. The lucky origin itself is not the source of instability and thus does not undermine understanding.

Furthermore, the miraculous origins of an ability have less of an undermining affect once discovered in cases where the ability is evident. If one discovered that one had been taught to play the violin by an by unreliable teacher, but one was making a living as a concert violinist, then the unreliable origin would have little effect on what it was rational to believe about one's violin playing ability. In the same way, if one has independent confirmation that one's understanding allows one to successfully solve problems, the fact that the dispositions were acquired by luck should have little effect. In some cases, one can have independent confirmation of one's understanding. For example, one can successfully pass tests in the relevant subject. Experts in the subject can tell one that one's statements make sense. One may be able to solve problems with easily recognized success states (e.g. designing a rocket that flies as intended). One can engage in research that ends up accepted and published. Understanding can give rise to evidence of its own presence. This makes understanding unlike a belief: the justification of a belief depends strongly on the evidence bearing on the belief.[12] Evidence about an ability plays a lesser role in making the ability possible. One does not lose an ability simply by learning information that (falsely) leads

[12]Some, such as (Conee and Feldman 1998) argue that the justification of a belief depends entirely on the evidence

one to think that one lacks the ability.

Further, even if an ability is tenuous, that does not mean it is not an ability. One can understand in the face of evidence that one's understanding has questionable aspects. The fact that one can lose an ability such as understanding does not mean that one does not have it. At one point I understood several methods of solving differential equations that, thanks to the magic of forgetting, I no longer understand. This does not prove that I never understood at all. Vulnerability to defeat does not show that understanding is incompatible with luck.

Having the ability to answer questions means that one has the ability to acquire many true beliefs about the world through the operation of one's ability. One may have already used the ability to form many true beliefs. Though initially understanding may have been a matter of luck, once one has understanding one is able to achieve cognitive success. Just as finding out one's tightrope teacher was a scam artist could not overcome the manifest evidence that one can in fact walk a tight rope, so one who has the ability characteristic of understanding has been able to use this ability and is now in a position where it is no longer a matter of luck, and it is no longer tenuous.

You would not like a method of teaching that only by chance gives the skill. You want something that does it reliably. But do we have separate concepts for genuine understanding, and sham understanding that only came about through luck? "You don't understand it, you merely happen to be able to use information to reliably draw correct conclusions using reliable methods." This sounds ridiculous, and there is no reason to draw this distinction.

If we want to know whether someone has an ability to do something, it does not matter how they came to be able to do it. Understanding is compatible with luck.

5.9 Anti-luck intuitions

What should we make of the remaining intuitions that understanding cannot be due to luck, then? There are alternative explanations for the intuition. First, the intuition is partly that

one does not deserve understanding by luck. The method by which understanding was acquired is not the kind of method one would want to employ to get understanding. This does not challenge the compatibility of understanding and luck; the fact that one objects to the method of acquiring X does not imply that X is not X. One would not try to win a ping pong tournament by plying blindfolded, but if one did win blindfolded, it would still be a win. Thus, intuitions about luck and understanding can be attributed to our reaction to the unreliable method, not to the resulting state. We do not need to change the account to deal with the intuitions.

We must ask exactly why anyone should think luck is a problem. There is a clear sense in which getting a correct answer by luck is a problem. We form so many beliefs each day—not just about science and philosophy, but about the location of our shoes, what we did, who we saw, what we scheduled, etc.—that a bad belief forming method, even if successful once, will likely bring untold trouble. However, abilities and dispositions, and thus understanding, are not in the same position.

An ability is a stable trait, not a one-off performance. While there can still be cause for concern regarding how one gains abilities, the concern is not as acute. The manner in which one acquires abilities is not as rapid and as simple as the acquisition of beliefs. One can acquire a belief from a single use of a capacity, but abilities tend to be gained over longer periods of training. This training is meant to produce a stable set of dispositions that lead to reliable success. There are moments we sometimes describe as flashes of insight where it seems that all of a sudden one understands what one could not understand before. There are also some topics that one understands immediately upon one's first encounter with them. It can seem that understanding just happens, however, it likely comes about through a process that is anything but random. Once one has a skill, it is a stable trait that allows for success by design rather than luck.

Evaluating methods of training dispositions is a higher-level evaluation than evaluating beliefs. At the bottom level, we ask whether a performance was a success (e.g. whether a

belief is true). At a higher level, we ask whether the process that led to the belief is reliable (e.g. assessing justification). At a higher level still, we ask whether the reliable dispositions are formed in a way that reliably leads to reliable dispositions (e.g. whether understanding is due to luck). At some point, this evaluation will cease to make sense. At some point, there will be no identifiable process that led to the process that led to the process that led to the process... whose reliability we can assess. As we go higher, the processes in question are deployed less frequently, and are harder to describe or identify.

I suggest that beyond assessing whether one has understanding, higher level evaluations are of vanishing importance. Luck is a problem when forming beliefs and finding answers. We are concerned whether someone has come to a correct conclusion only through luck. One alternative to luck is skill. Thus, our concern for luck is primarily about conclusions, one kind of luck is actually removed by the presence of this ability. Are we concerned about luck at the higher level, whether the presence of the skill is only a matter of luck?

At some level, our possession of any cognitive skills is a matter of luck. It is a matter of luck that there is a universe capable of supporting life. A matter of luck that there was a universe in which life arose. A matter luck that life persisted long enough for intelligent beings to evolve. A matter of luck that our parents met. None of this undermines understanding. It just turns out that luck can go one step further than some have thought, and understanding can be acquired through luck.

Conclusion

The relationship between understanding and luck is now clarified. Understanding can be acquired due to luck. However, understanding in many ways stands opposed to luck. It is the force that keeps luck at bay. Understanding is what allows us to find the right answer due to our own achievement, rather than by luck.

There is a close connection between understanding and explanation, but what is that connection? In this chapter, I argue that explanation depends on understanding. An explanation is an act that produces understanding, given the relevant cognitive background. The relevant cognitive background depends on the context and audience. Explanation is thus relative to context and psychology, contrary to widely accepted views explanation.

5.10 Introduction

Our study of understanding would not be complete without an examination of the relationship between understanding and explanation. Traditionally (at least in 20th century analytic philosophy), explanation was taken to be the more tractable of the two, and so took center stage.[13] In part this was due to the fact that being a good explanation was taken to be an objective property, unlike understanding which is a state of an agent's psychology.[14] The thought behind this is that an explanation is good or bad independent of our interests or beliefs. Understanding is just a psychological relation to an explanation.[15]

In this chapter, I start from two basic assumptions. First I assume that, other things being equal, a unified account of explanation is preferable to a pluralistic account. If it is possible to give a satisfactory account of explanation that lays bare what unites explanations in physics, biology, mathematics, psychology and everyday life, this is a benefit of the theory. Second I assume that, other things being equal, a theory of explanation that lays bare the connection between understanding and explanation is preferable to one that does not. It is beneficial if our account of explanation demonstrates how explanations increase understanding, as is commonly claimed. This gives two criteria for an ideal account of explanation: first it should be unified, and second it should show the connection to under-

[13]This is argued most forcefully in (Trout 2002), (Trout 2007) and (Khalifa 2012), but is also found in (Hempel 1965), (Kitcher 1981), (Salmon 1998), (Strevens 2009) and elsewhere.

[14]There are philosophers who argue against this, including (Achinstein 1983) and (van Frassen 1980), but the majority who write about explanation accept that understanding is objective. Many examples will be given in the remainder of the chapter.

[15]In (Kitcher 1981), p. 508 "A theory of explanation should show us how scientific explanation advances our understanding." Similar statement occur in (Woodward 2004), (Strevens 2009), (Lipton 2009) and (Grimm 2010), who further say that understanding is the goal of explanation.

standing.

As I will argue in this chapter, previous accounts of explanation do not do well in satisfying these two criteria. The accounts either fail to unify explanation, or fail to illuminate the connection between understanding and explanation, or both. In this chapter I survey the most prominent non-psychological approaches to explanation: the deductive-nomological account, the unification account, causal accounts, and contrastive accounts. These approaches come from a long-standing research programme in the philosophy of science, the goal of which is to find a theory of explanation according to which *being a good explanation* is an interest-independent, mind-independent property of an argument or an act. As we will see below, according to these mind-independent theories the value of a good explanation is that the characteristics that make it a good explanation tend to be useful to us in many ways. Whether something is an explanation or not is independent of any practical value or psychological effect it might have.

Each mind-independent account has problems. While each problem is not necessarily fatal, together they constitute prima facie reason to consider an alternative approach. While existing mind-independent theories all do identify features that are interesting and often valuable, none of these accounts is adequate to capture the nature of explanation. First of all, they are subject to counterexamples. Second, the accounts are not general accounts of explanation, even in ambition; they only attempt to capture scientific explanation, leaving everyday explanations to be covered by different accounts. This complaint has already been noted in (Khalifa 2012), (Salmon 1990a). When taken as accounts of explanation in general, objective accounts deliver unacceptable results; the accounts each merely identify features of arguments that often come in handy. Noticing this, many, such as (Khalifa 2012), have given up the hope that a unified account of explanation exists, preferring to hold that there are several kinds of argument that are classified as explanations, but no illuminating account of what makes them all part of the same genus.

I argue that there is a unified account of explanation that covers scientific and everyday

cases. I argue that once we accept that the proper account of explanation is mind-dependent, we can give a unified account of everyday, personal, and scientific explanations. Understanding is the uniting factor at the core of explanation. An argument counts as a good explanation when it provides understanding of a subject matter related to the fact to be explained. This account works for scientific, mundane and personal explanations. Once we move past mind-independent accounts of explanation, the dependence between understanding and explanation is the reverse of conventional wisdom: the essence of explanation depends on understanding. Nothing is an explanation in the absence of understanding.

I argue that understanding is more fundamental: explanation is defined by its ability to produce understanding in the relevant psychology. As this suggests, being a good explanation is relative to psychology. It depends on the relevant contextually salient psychology (explained in section 5.15). While I agree with the generally accepted claim that explanations provide understanding, I disagree about the reason. It is not because understanding is dependent on explanation in some way. It is not because understanding just is bearing the right psychological relation to an explanation. It is the reverse: explanations produce understanding because, to a close approximation, that is what it is to be an explanation.

Below I present and defend this view with the following: In section 6.2, I describe in general terms what I am trying to provide an account of. Then in section 5.12, I survey the major objective accounts of explanation, arguing both that they exhibit useful features but also fail to capture explanation. In section 5.13, I argue that in addition to the inadequacy of existing objective accounts, we have reason to believe that no objective account of explanation can be adequate. In section 5.14, I briefly recap my view of understanding and how it is relevant. In section 5.15, I present and defend an account of explanation in terms of understanding. In section 6.7, I explain how my account deals with the problems faced by other accounts. Finally, in section 5.17 I argue that understanding is the independent phenomenon and explanation is dependent.

5.11 Explanations

I will approach explanation via the role it plays in inquiry. This section clarifies the target of the various accounts discussed, including my own. Explanations are sought in circumstances when there is a sense that the fact to be explained is a guidepost to further discovery. Some facts cry out for explanation, raising why-questions such as 'Why is this true?' A good explanation does more than merely describe a phenomenon or fact. A good explanation satisfactorily answers the why-question and fulfills the demand for explanation. The term 'explanation' can refer either to the act of explaining something or to that which is offered in an act of explaining. For example, a why-question can be answered by presenting an argument. We can refer to the act of providing the argument as an explanation, or alternatively we can refer to the argument itself as an explanation. The first is the act of answering a why-question, and the second is that which can serve as a satisfactory answer to a why-question, or can serve as a satisfactory response to a request for an explanation. I will focus primarily on the act of presenting an explanation, since this act includes that which is presented as object, while the object itself does not include the act. Focusing on the act allows us to include more potentially relevant details.

We now must consider which kinds of acts can be acts of explanation. The act that has received the most attention from philosophers is the act of providing an argument. Some acts of providing an argument are acts of explanation, but not all are. For example, presenting the obligatory *Socrates is mortal* argument in a logic class is not an act of explanation, while presenting the same argument when asked why Socrates is mortal can be an act of explanation. The argument is then the explanation as an object. Several of the most prominent theories of explanation identify arguments as the primary or only form of explanation, including (Hempel 1965), (Kitcher 1989), (Salmon 1990a), and (Strevens 2009).

However, explanations encompass more than just arguments. Presenting a model is another mode of explanation common in sciences. Models can be presented as diagrams, as drawings, as mathematical structures, or as three dimensional mechanical objects that

exhibit the relevant explanatory features. In such cases, the act of explanation can be to present the model and perhaps draw attention to certain aspects of the model. The import of this act, and the reason it satisfies the demand for explanation, cannot always be captured in a single argument, as argued in detail in (Bechtel and Abrahamsen 2005). For example, one could explain retrograde motion to someone using an orrery (a mechanical representation of the solar system that models the relative motion of the planets), and this explanation could be successful, allowing that person to move forward with her inquiries about the planets. However, this does not mean she can provide an argument for the conclusion that retrograde motion occurs. The ability to use the model can allow an explanation to be successful even in the absence of explicit arguments. There might be aspects of the model that can be encapsulated in argument form, but the entirety of the information contained therein, and the reason it satisfies the relevant demand for explanation, cannot in many cases be so encapsulated.[16]

Finally, if we consider everyday cases of explanation, we discover further examples of satisfactory explanations that do not take the form of an argument. For example, John asks why my desk is in the hallway. In response, I nod in the direction of my office. He looks and discovers that painters are at work covering the walls. I have not uttered a word or made any argument, but John is satisfied that I have explained why my desk is in the hallway. This expands on the previous point about models: an *act* of showing, demonstrating or calling attention to something can be a good explanation. There might be ways to try to reconstruct what is conveyed by my head nod in the form of an argument (as I implicitly did in describing the scene to which I nodded), but the act of explanation is not the act of presenting an argument.

The class of acts and entities that I have identified as potential explanations is broad, including arguments, statements, models, and gestures. This may raise concerns that my

[16](Lipton 2009) presents the use of models without arguments as cases of understanding without explanation. While I agree that models can allow for understanding, I reject the interpretation that they cannot be explanations: these models can be used in satisfactory responses to why-questions.

target is too broad, that explanations are inherently linguistic or propositional. However, the broad class of acts is united by their ability to serve as satisfactory responses to the demand for explanation. They all play the role in inquiry that explanation plays. Thus, they make a united and philosophically interesting category in their own right. One could respond to any of the acts noted above by saying, "That explains it".

An act of explanation is goal-directed: it aims to provide an answer to a why-questions. Explanation is *successful* when it answers the questions in the manner intended. An act cannot be a successful act of explanation if it was not intended to explain. If you program a computer to generate random phonemes and it emits sequence of phonemes that happens to sound like an argument that fulfills your friend's need for explanation, no one has performed a successful act of explanation. If you absent-mindedly swing your arm in a way that looks like you are pointing in a direction, and that happens to lead someone to find the explanation they were looking for, your act was not a successful act of explanation. While someone has their explanatory demand satisfied—and in each case, satisfied by something that could be presented as a successful act of explanation if it was intended as such—it is not an act of explanation. If we focus on the act, we can see that intentions matter when it comes to describing something as a good explanation.

To summarize, acts of explanation are acts that can adequately answer explanatory why-questions and satisfy demands for explanation, and an explanation as an object is what is offered in the act of explanation. Such acts can include presenting an argument, presenting a model, making a demonstration, making a brief statement or making a gesture. Henceforth, I will take as given that such acts can be explanations. With this in mind, I will survey existing mind-independent accounts of explanation. I will show that each has some difficulties, and none provides a general account encompassing scientific and non-scientific explanation. I take this to provide *prima facie* reason to seriously consider a psychological account of explanation.[17]

[17]None of the objections is decisive. Proponents of the various theories have offered responses to many. However, the fact that none emerges as clearly correct supports the search for a better theory elsewhere.

5.12 Mind-independent accounts

The dominant view of explanation is that it is mind-independent, that features of arguments or models objectively determine whether they are good explanations, independent of any agent's psychology.[18] In this way, the project of providing an account of explanation mirrors accounts of logical validity. Whether an argument is valid is an objective fact about the argument, and the practical or psychological impact derives from the fact that arguments that are valid tend to be useful. Explanation is thought to be the same.[19]

Hempel summarizes the project to provide an objective account of explanation thus: "What scientific explanation, especially theoretical explanation, aims at is not [an] intuitive and highly subjective kind of understanding, but an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of common, underlying structures and processes that conform to specific, testable, basic principles." (Hempel 1966, p. 83)[20] Kitcher is likewise clear in his intentions regarding his account of explanation: "I shall try to . . . [show] that there are certain context-independent features of arguments which distinguish them for application in response to explanation-seeking why-questions, and that we can assess theories (including embryonic theories) by their ability to provide us with such arguments," (Kitcher 1981, pp. 511-512). The motivating idea is that the practical utility of explanations derives from the fact that their objective features are often useful, in much the same way that validity is objective yet it is often a useful feature of an argument. [21] Even some who allow more than just arguments to count as explanations claim that the features of the models/diagrams/gestures that make them good explanations are meant to be mind-independent, as in (Bechtel and Abrahamsen 2005).

The purpose of the present section is to demonstrate that the project of searching for a

[18] By this I mean that no agent's psychological relation to the argument makes it an explanation. An explanation about psychological matters relies on psychological facts, but that does not make it agent-dependent in the relevant sense.

[19] Though some have challenged this, such as (van Frassen 1980, ch. 5), (Achinstein 1983).

[20] See also (Hempel 1965, pp. 345, 444).

[21] See also (Trout 2002) for related claims.

mind-independent theory of explanation has so far failed to produce a unified account of explanation. There are multiple mind-independent accounts that all conflict with each other, none of which has anything like a consensus. Each faces counterexamples. While each theory does identify objective features of arguments that we often find useful in explanatory contexts, no theory successfully unifies scientific and non-scientific explanations. Some who endorse the claim that explanations are mind-independent have resigned themselves to the fact that there is no unified theory of explanation. Instead, these philosophers endorse the claim that there are several kinds of explanations (causal explanations, structural explanations, etc.) but no account that captures the essence of explanation in general. For example, this is argued in (Khalifa 2012) and (Salmon 1998) and hinted at in the conclusion of Woodward's *Stanford Encyclopedia of Philosophy* entry on scientific explanation.[22] It would be a benefit for a theory of explanation to be unified.

To show why the mind-independent project is not up to the task of providing a theory of explanation, we will first consider the most prominent mind-independent accounts of explanation and the various ways in which they fail. We will consider the DN account, the unification account, and the causal/interventionist account. Showing the defects of each will serve as prima facie reason to consider mind-dependent accounts.

5.12.1 D-N account

First we consider Hempel's deductive-nomological account of explanation. He focuses on explanations as objects, identifying them with arguments. In our terms, he takes explanations to be acts of presenting arguments. According to this account, explanations of events are derivations of events from general laws plus initial conditions. An argument is an explanation of p at time t just in case it has the following features:

1. It contains a true statement of conditions before t;

2. It contains true statements of scientific law;

[22]Woodward's entry is found at http://plato.stanford.edu/entries/scientific-explanation.

3. The laws and initial conditions together imply that p at t;

4. There are no irrelevant premises.

Explanations of this variety provide understanding, it is claimed, because an explanation demonstrates that an event was to be expected. According to this account, understanding why an event occurred is identical to seeing why it was to be expected, which in turn is just being able to show that it followed from the laws and initial conditions. Having an explanation for a past event prepares the mind to understand the future by giving the template for future predictions.

Derivation from universal natural law is often considered the gold-standard of explanation in science. Primarily, this is due to two features of such explanations: they are often useful, and they often contribute to understanding. D-N explanations are often useful because with only minimal competence in deductive logic, one can use a D-N explanation to derive the occurrence of the target phenomenon. One only needs to grasp the meanings of the claims and make the right logical inferences. This often allows one, with small changes to the parameters, to derive related phenomena. For example, if one has mastered the Newtonian derivation of orbital motion around a central body with mass M, a simple substitution is generally enough to derive the orbital motion around a central body of mass M'. The explanation serves as a useful template for deriving related facts.

Second, D-N explanations often aid understanding. Understanding means having the ability to extract and exploit the relevant information to answer questions within a subject matter. A deductive argument from natural laws tends to highlight facts that are often relevant for answering questions in the relevant field. The psychological state of grasping an explanation and seeing that the conclusion follows is a relatively trivial addendum to explanation. D-N explanations can contribute to understanding, which is a central goal of science.

The D-N model of explanation suffers from difficulties, the main being that many derivations from laws and initial conditions are intuitively not explanatory. For example,

consider the familiar fact that from the laws of optics and geometry, one can deduce the length of a flagpole from the length of its shadow, along with the positions of the sun and the ground. However, derivation from the length of the shadow does not seem to be any kind of explanation for why the flag pole is as tall as it is.[23]

Second, not all explanations are deductive. We can explain why John has PTSD by reference to his experience as a soldier in Iraq. The fact that he was a soldier in Iraq, even if we include details of the situations he found himself in, does not entail that he gets PTSD. Not everyone who experiences traumatic events gets it, but his experience still explains his condition. To attempt to escape this, we can refine the D-N account to include deductions of statistical claims. One can explain an event/fact if one has a deductive argument to the conclusion that the event will very likely occur, or is more likely to occur than otherwise.

However, the refinement fails as well. One can explain the fact that a weed is alive by the fact that it was healthy and then sprayed by a cheap weed killer that kills 99% of healthy weeds, sparing 1%, rather than a more expensive brand that kills all sprayed weeds. Thus, we can explain its survival in terms of something that implies it almost certainly will not survive. We can derive that the survival of the weed is possible (there is a 1% chance), but this deduction does not explain why the weed survived.[24]

Finally, the account is excessively restrictive. It cannot accommodate explanations that do not take the form of arguments. Even among explanations that are arguments, the explanations offered in everyday life rarely if ever appeal to laws of any kind, statistical or not. While explanations in science often refer to laws, when explaining why a store is closed or why the door is stuck, appeal is made to various factors but not to laws.

One might respond that everyday explanations and non-verbal explanations are mere indicators that point us to genuine explanations.[25] They are not explanations but *explana-*

[23]This objection to Hempel was originally introduced by (Bromberger 1966).

[24]See (Railton 1978) for a discussion of statistical explanations for unlikely events. His model of D-N statistical explanation is more refined, but I claim it still falls short. I lack space to defend this in detail, but in short, his account defines explanations of how something is possible, but not why it happened.

[25]For example, one could offer a partial explanation and then rely on the audience to construct the complete explanation. A partial explanation of p is an explanation of some q such that p implies q. See (Lipton 1991).

tion sketches, to use the term from (Hempel 1965). One can consistently claim both that being a complete explanation is a mind-independent property and also that our interests determine whether a partial explanation is acceptable in a context. Partial explanations are often acceptable in everyday situations because once added to background knowledge they form complete explanations. Similarly, one can argue that non-verbal explanations draw attention to facts that make up an objectively good explanation, so that the audience is able to piece the explanation together as a result of the non-verbal act.

However, this response is problematic. It incorrectly assumes that the shared background that allows for the success of brief or non-verbal explanations is a background of beliefs (or propositional attitudes such that the propositions form an explanatory argument). However, a background of beliefs is not sufficient to make up the difference. If the explanation offered is a model, for example, the audience needs the capacity to use the model, a capacity which can by no means be taken for granted. One needs to know how to use an orrery in order for it to serve as an explanation of retrograde motion.[26] If the explanation is a gesture, the audience needs to have the right tendencies to interpret gestures and to notice features of the environment that are gestured at. If they do not attend to the relevant features of the situation, then pointing in a certain direction will not satisfy any need for explanation. The audience must be prepared to extract the relevant information, which cannot be assumed. If the explanation is something short and cryptic, the person must be able to discover how it is relevant to the situation at hand, which also cannot be taken for granted. In every case, in order for the explanation to be successful, we must assume a complex and robust set of not just background beliefs, but background capacities for recognition, attention, and determining relevance.

It is not just that these capacities need to be in place for the audience to grasp the explanation: in some cases explanations are good (or bad) in virtue of these capacities. Suppose I have some need for an explanation. Johnny presents me with a model for the

[26] An orrery is a mechanical model of the solar system that exhibits approximately correct ratios of the relative revolution rates of the planets.

phenomenon in question, and having this model to hand allows me to alleviate the demand for explanation. Having the model in mind gives him the ability to answer further questions related to to the explanandum. This need not be accompanied by my having worked out any answers yet, so it is not the case that it is a good explanation because it allows me to gain further beliefs that constitute a D-N explanation. Simply observing the model will allow me to answer questions, if it is genuinely a successful explanation. Given the same background beliefs but different capacities, the same model would not allow me to be successful in answering questions. It is in virtue of the fact that I am able to use the model that it satisfies the demand for explanation and answers the relevant why-question. Thus, it is not solely in virtue of any argument that the explanation is good.

It is not always beliefs that do the work. One need not explicitly believe that if there is fire there is smoke; one can simply have the tendency to infer smoke from fire. The tendencies to make such non-deductive inferences, or "material inference" following (Sellars 1953), is a crucial aspect of our reasoning. Our ability to go from an explanation to a useful conclusion depends on inferential tendencies. No matter how much background information we have, a D-N argument is useless if we lack the ability to use it, to make inferences from it in a fruitful way. The presence of the rest of the argument in the background and environment is not necessarily enough, for that by itself does not guarantee any usefulness.

Normal usage of the term 'explanation' allows for the existence of short, cryptic, or nonverbal explanations. The sometimes only distantly related facts presented in answer to why questions are what we most often refer to as explanations. Not only that, but we also never give the more complete explanation. We never list the entirety of the causal antecedents of an event, we never have a clear sense of patterns that unify background knowledge, we lack laws to cover most of what we explain, and we rarely have a clear sense of contrast when giving everyday explanations. The flaw in each case, which there is good reason to suspect would affect any objective account, is that no matter how much explanation we give, we always rely on the background to make it work. We do not refer to

implicit, complicated arguments that we cannot even entertain as 'explanations.' This puts the burden of proof squarely on the shoulders of those defending the D-N account.

As we will see below, we can give an account of explanation that does not force us to make such revisions. The arguments we describe as explanations appear to be heavily dependent on pragmatic context, so a mind-independent account must either match this dependence, or show convincingly that we are usually mistaken in what we call explanations in everyday life. I am not optimistic that this will ever be done, and viable pragmatic accounts exist, as I will show.

In the end, despite the usefulness of many D-N type explanations, the D-N model is inadequate as a general account of explanation.

5.12.2 Unification accounts

Unification accounts of explanation claim that we have an explanation of a phenomenon when we can unify that phenomenon with others. If we can give an argument for a proposition that the phenomenon occurs using widely applicable argument patterns starting from few assumptions, then we have unified the phenomenon with others that can be deduced with the same general pattern of argument. While (Friedman 1974) provides the first detailed statement of the view, Kitcher has given the most fleshed out exposition of unification in the relevant sense, as in (Kitcher 1981) and (Kitcher 1989). In his system, unification takes place when we are able to derive all of our knowledge in a domain from a small set of reasoning patterns and basic assumptions. Let K be a state of background knowledge. Kitcher defines E(K) to be what he calls *the explanatory store over K*, which is a set of arguments that count as explanations given background knowledge K. What arguments are in E(K)? According to Kitcher, E(K) is the set that best unifies K. In other words, it is the set of arguments that (1) we accept as valid (in a broad sense to include inductive arguments), (2) establishes the truth of propositions in K, and (3) has the fewest argument patterns and unproven assumptions. E(K) is said to be more unified when the arguments in it are all in-

stances of a small number of general argument patterns from a small number of premises. E(K) is more unified if it contains fewer brute facts (facts that themselves do not have explanations in E(K)). E(K) is more unified when it contains fewer argument patterns from fewer brute facts, and good explanations use these argument patterns to derive as many and as varied phenomena as possible. Thus, what sets apart explanations from other arguments is that they are part of the system of arguments that best unifies background knowledge.

Explanations that match the unification account have the desirable feature of flowing from a well-organized system of argument patterns of high generality. Consider the ability to use Newtonian Mechanics to deal with the motions of planets, the motions of projectiles and the motion of billiard balls rather than requiring different systems for each. Organization of memory is easier if one only needs to remember one argument pattern that applies with little or no variation across several phenomena rather than separate arguments in each area. Ease of recall and ease of use through habituation can be increased by reducing the number of facts and reasoning patterns one must be able to bring to mind on demand. If one is trained to recognize a pattern in disparate contexts, and then successfully use it to derive needed conclusions, then that will help one to more effectively solve problems. Likewise, when attempting to teach someone a subject, reducing the number of patterns required will make it easier to learn.

Kitcher claims that unification produces understanding because, "By using a few patterns of argument in the derivation of many beliefs we minimize the number of types of premises we must take as underived," (Kitcher 1981, p. 259). This is essentially all that has been offered in support of the thesis that unifying explanation provides understanding.

There are objections to unificationism. One objection comes from (Gijsbers 2013): Suppose that the laws of physics vary wildly from place to place, such that the laws that govern physics in my basement are different from the laws governing everything else in the universe. It seems that one could perfectly well explain the goings-on in my basement using the laws that operate there, though such explanations would not be unified with other

phenomena via operating under the same laws. This is explanation without unification. However, perhaps a pattern of reasoning that works for a different phenomenon happens to work in the basement. Is this unification in the relevant sense? In this case, it is not clear why using the same pattern of reasoning in my basement that works elsewhere is by itself explanatory, as unificationism suggests.

Unificationism also rules out Dennett's view regarding psychological explanations. (Dennett 1991) argues that psychology provides genuine explanations of events by identifying patterns, but that this does not imply a higher level of facts than the microphysical level (he calls the higher level "quasireal" and adopts "mild realism" based on the fact that psychological reasoning works). If this view is correct, fields like psychology add a system of patterns of reasoning that are genuinely explanatory, but are not reducible to the lower level explanations, while dealing with the same basic reality. There is less unification, but better explanation. This is ruled out by unificationism. One might be concerned that a view about explanation should be neutral regarding this view of psychology.

Further, the unification account allows that in many cases, explanations do not contribute to understanding, against one of our desiderata of a good account of explanation. Suppose we start out only knowing general relativity, which is very difficult to apply in specific situations. Then, someone derives Newtonian Mechanics as an approximation that works in many conditions. The addition of Newtonian Mechanics to the physicists' arsenal allows them a greater understanding: once one is used to using NM, one is able to find answers one could not when one was forced to rely on general relativity. However, the addition of NM does not unify the field more than it already was. It does not contribute to the explanatory store of arguments. Nonetheless, it improves understanding. It seems like NM gives us new explanations that increase understanding, but it is not clear how unificationism can accommodate this.

The problem comes from the fact that understanding is a cognitive ability, the ability to answer questions. While we could improve this ability through ever more general patterns

of argument, it is also possible to increase the ability to answer questions by internalizing a wider set of easily deployed, but very specific forms of reasoning. It is in principle possible for one to have a method of answering questions about monkeys and a method of answering questions about humans that does not rely on the commonalities among them. If this occurs, the connection between unification and understanding is broken, at least for the case at hand. Unification can be seen as something that often leads to understanding, but does not necessarily.[27]

Further, the fact that one can use related reasoning to derive two things does not seem to increase understanding necessarily. If it is an accident that the two things can be explained in similar ways, then unifying them under one banner does not provide understanding. If purely by accident, the same set of equations can be used to find the chance of rain in a location and the chance of a drop in a stock market, this is not explanatory. One does not understand either phenomenon better by noticing that they happen to have the same formula. If explanations are supposed to provide understanding, as is commonly supposed, then unification does not capture the core of explanation.

5.12.3 Causal accounts

Next, we consider causal accounts of explanation. Proponents of the various causal accounts also claim that their explanations (as identified by their theory) provide understanding.[28] They support this claim by appeal to the fact that answers to why questions are almost always causal histories, which best provide understanding. In the words of one proponent, "To seek to understand almost any empirical F is just to seek the knowledge of its causal basis," (Barnes 1992, p. 8). This intuition is strengthened by the idea that explanations seek to open black boxes and uncover the mechanisms that make the world work. Mechanisms are causal by nature, and knowing the mechanisms is knowing their causal

[27](Barnes 1992, p. 6) presents an argument against unification accounts of explanation that shares some similarities with the one presented here.

[28]For example, (Barnes 1992), (Salmon 1990a), (Woodward 2004), (Strevens 2009), (Strevens 2013).

powers and histories.[29]

A variant on this strategy is the *interventionist account of explanation*, which says that an explanation is an argument that cites facts that could be manipulated to change the outcome that is being explained. In other words, an explanations consists of knowing, or being in a position to know, a series of counterfactuals regarding what would happen to the system under various interventions. In brief, the insight behind the account is that explanations show us why things are one way rather than another by pointing out what could have been manipulated to lead to other outcomes. We seek explanations so we can figure out what to do.

Knowing the causal structure behind an event is useful because it gives information about possible interventions. Knowing what would result if parameters were changed is valuable in making decisions regarding such changes. It opens one's mind to the opportunities for change that are available. It might even tell one that there is no way to intervene to get desired results. Furthermore, determining the cause of something, when the cause turns out to be a person, can be crucial in assigning praise and blame. In addition, knowing the causal history or causal dependencies among objects should allow an agent to answer questions about what would happen if things were different, or if something was manipulated.

Unfortunately, these accounts have problems as well. Causal accounts cannot deal with explanations in fields such as pure mathematics where the facts explained and cited in explanations are not causal facts. They cannot account for the difference between explanatory and non-explanatory mathematical proofs, since neither cite causes.

Further, it is difficult for causal accounts to handle explanations of physical phenomena that rely almost entirely on mathematics, abstracting away from physical aspects. For example, an explanation for why a person failed to untie a knot might appeal only to the topology of the string, rather than citing the specific forces applied to the string during

[29]I also include the mechanistic account of explanation under the heading of causal account. For details of this account in particular, see (Bechtel and Abrahamsen 2005).

attempts to untie the knot.[30] In general, some fields of science appear to eschew causal notions but provide explanations just fine, quantum mechanics being a prominent example. Robert Clifton argues that, "[. . .] if we are to understand quantum theory . . . we will have to take seriously the idea that locating phenomena within a coherent and unified mathematical model is explanatory in itself" (Clifton 1998, p. 6). As one illustration, our best explanation of the general uncertainty principle in QM is in terms of the formal features of Fourier transforms of wave functions. This explanation relies only on the formal features of the mathematical representation of the phenomenon, rather than causal aspects. It turns out that given the way position and momentum are related in Hilbert space, the uncertainty of one increases when the other decreases. There is no field or causal force that does this, it is a formal feature.[31]

Similarly, explanations in general relativity often proceed in terms of the structure of spacetime without assigning any clear causal relations. These explanations do not rely on deciding whether matter causally influences space or space influences matter. Closed time-like curves are an example of this. In a space with a closed time-like curve, there are world-lines (i.e. paths that travel forward in time from the point of view of someone traveling on that path) that loop back on themselves, returning to the same point in time and space. In such spacetimes, there is no clear way to separate past from future, and it is difficult to parcel out causation. However, the formal features of the spacetime allow for explanations of what occurs in a closed time-like curve. For example, there are constraints on consistency: one can explain the fact that one cannot travel a closed timelike curve and kill one's past self by appeal to the fact that it would imply inconsistent states at some point in one's worldline. However, no specific causal process can enforce this; it is a global consistency constraint that explains it, as described in (Earman 1995).

Another problem case is equilibrium explanation. One can sometimes explain why p

[30] This example comes from (Kitcher 1989).

[31] For more detail about the explanation of the uncertainty principle and its formal nature, as well as related examples, see (Dorato and Felline 2011, p. 169).

obtains by appeal to the fact that the system will somehow ensure (via laws or some other features of the situation) that in the end p will obtain. The fact that p is true is an equilibrium state for the system. An equilibrium explanation does not provide the causal history that actually led to p, but it does explain it.[32] One might argue that equilibrium explanations are genuinely explanatory because they outline a class of causal histories, each of which leads to the explanandum, and one of which must have happened given the boundary conditions. However, this fails to account for the explanatory power of such explanations. Equilibrium explanations are more than simply disjunctions of causal histories, all of which lead to the same state. Equilibrium explanations appeal to generalizations that are not purely captured by causal histories.

Strevens attempts to incorporate structural and equilibrium explanations within a causal framework by appealing to abstraction, as in (Strevens 2004), (Strevens 2009) and (Strevens 2013). He claims that an explanation is an argument whose premises form what he calls an *explanatory kernel*. An explanatory kernel is a set of causal factors that (i) jointly entail that the explanandum obtains, and (ii) are as abstract as possible while maintaining cohesion (see his paper for more detail about exactly what he takes these to mean, the details will not concern us here). A good explanation applies to many possible models that are all cohesively united by relevant causal features.

Strevens argues that this can account for structural explanations: structures are cohesive abstractions of specific causal situations, and thus they fit within his causal explanatory framework. While not necessarily specifying the exact causes of the event, it locates the event within a class of models, all of which metaphysically imply the explanandum. Arguably, equilibrium explanations can also be accommodated in this manner: such expla-

[32] One expression of this worry: "[...] equilibrium explanations are not causal, because they do not identify the causal trajectory that a system takes from its initial state to the state that is to be explained. An evolutionary biologist, for example, might explain the roughly one-to-one sex ratio in a certain species by showing that any population with a different ratio will be pulled by natural selection toward the even ratio, thus that the even ratio is a unique evolutionarily stable equilibrium. In explaining the population's even sex ratio, then, they do not specify the ratio at any earlier time, or the path by which the earlier ratio (if non-even) became an even ratio. Apparently, they describe no causal process at all," (Strevens 1999).

nations identify—in an abstract way—a cohesive class of causal histories that consistently remain within an equilibrium state.

One problem is that, as Strevens admits (Strevens 2004, p. 161), his account is deterministic and thus unable to handle probabilistic explanation. If there are genuine chance phenomena in quantum mechanics, then even this improved version of the causal account cannot handle quantum mechanics.

Furthermore, this theory cannot deal with explanations in pure mathematics. It fails to imply that the brief or non-verbal acts of explanation discussed above are explanations at all. Strevens manages to include some of the more formal structural explanations, and while it provides some of the benefits of both the causal and unification approaches, it still fails to deliver a complete theory of explanation. It is still not clear what class of causes are identified by Fourier-series-based explanations of the uncertainty principle. It is the formal nature of the wave function (in Hilbert space) to have greater variance in position when it has lesser variance in momentum. Even the best versions of the causal theory of explanation cannot account for all of scientific explanation, let alone mathematical, ethical, and everyday explanation. The best versions of the causal theory of explanation fail to deliver.

One might follow (Lewis 1986) and claim that something close to the full causal story is the true explanation, while brief or nonverbal explanation are merely pragmatically acceptable pointers toward the true explanation. However, this faces the same problems raised for the analogous response to the D-N account. Explanations are good or bad in virtue of our capacities, and normal language usage favors treating these as genuine explanations. This gives us reason not to be satisfied with causal accounts, and to be open to psychological accounts.

5.12.4 Contrastive

Finally, we consider the contrastive account of explanation, defended for example in (Lipton 1991). According to this account, an explanation of p either implicitly or explicitly specifies a contrast q for p. Explaining why p, according to this account, is always a matter of explaining why p rather than some contrast. The meaning of the contrasts can be captured with the idea that explaining why p rather than q amounts to demonstrating that p given the presupposition that $p \vee q$. The explanation rules out every contrast to the explanandum. Contrasts will either be explicitly stated or will be implicitly determined by context. They represent the contextually salient alternatives to the explanandum. This makes explanation a three-place relation among the explanandum, explanans and contrast.

The value of a contrastive explanation is that we often find ourselves in a position where one of a defined set of alternatives is salient, and we must establish why that particular alternative, rather than the others, obtains. Possession of an argument that rules all but the truth out is valuable. The alternatives that form the contrast, since they are relevant to this question (why this rather than those?), will likely be relevant to other questions as well. Having them ruled out already is valuable.

For those instances in which a contrast for p is not explicitly stated, features of context—potentially including agents' psychology—determine what the relevant contrast is. Features of context can include conversationally and practically relevant facts or facts about what the agents happen to find compelling. However, I still count this as a mind-independent theory because we can interpret context as influencing *what is explained* rather than *whether a purported explanation is good*. The contrastive account implies that an explanation is always an explanation of p rather than q. Once we know which p and q, the explanation is no longer relative to psychology.

This account faces difficulties as well. The first difficulty regards the meaning of 'ruling out a contrast.' The simplest account would be that one rules out a contrast by adducing facts that imply its falsity, without begging the question. The problem is that this would

allow for the negation of the contrast to be an explanation. One could explain why the capitol of the USA is in Washington D. C. rather than New York by appealing to the fact that it is not in New York. This is unacceptable, though perhaps such explanations are question-begging and so do not pose a problem. However, that does not help the fact that one can explain the height of a flagpole by citing the length of its shadow and the position of the sun. The length of the shadow, plus the relevant laws, rules out other heights for the flagpole.

Perhaps, then, the relation of ruling out contrasts is not strictly logical. In that case, we find ourselves left with the task of defining which methods of arguing against a contrast are explanatory and which are not. The contrastive structure alone does not provide an account. Further, the best candidates for separating explanatory versus non-explanatory ways of ruling out a contrast are the features appealed to in the other mind-independent accounts of explanation. One could appeal to laws, causal mechanisms, counterfactuals of intervention, or unified systems of argument. The problems facing the other accounts will recur once we try to describe the difference between explanatory ruling out and non-explanatory ruling out. In short, the contrastive account is primarily a hypothesis about the structure of explanation (claiming that it is a three-place relation), while relying on further theories to fill in the details. The filling in recreates the problems of other theories.

Thus, the main contenders for mind-independent account of explanation do not work in the end. Any hope for an objective account must lie in hope that a new account will come along that will unite the insights of the others without the difficulties. In the next section, I argue that there is reason to doubt that such an account will ever be forthcoming.

5.13 Against objective accounts in general

There was never any a priori reason to think explanation should be mind-independent. Nothing about the use of explanation, or the nature of explanation demands that explanation is mind-independent. The problems facing all proposed mind-independent theories of

explanation provide reason for suspicion. It is time to reconsider whether the search for an objective account of explanation is taking us in the right direction. The best philosophical work in explanation has generated a multiplicity of mutually incompatible theories, each based on genuine insight, but each beset with difficulties. None of these accounts capture a sufficiently inclusive set of good explanations, while ruling out a sufficient number of intuitively bad explanations or non-explanations. These theories clearly form a plurality rather than a unity. This situation has long been lamented in the philosophy of science. Several philosophers sympathetic to the mind-independent project, including (Khalifa 2012), (Diéz, Khalifa, and Leuridan 2013) and (Overton 2013), have given up on a unified theory of explanation. These authors have adopted pluralism about explanation, but we could instead see the multiplicity of theories as a symptom of looking for mind-independent pegs to fill mind-dependent holes. The state of this literature thus provides prima facie reason to stop insisting on mind-independent accounts of explanation. More consideration reveals further reasons.

Consider the following claim that almost all mind-independent accounts agree is true: explanations provide understanding. The management theory of understanding, defended in previous chapters, says that understanding why p is the ability to find relevant cognitive resources to answer questions related to why p. As argued in Chapters 4 and 5, there are sometimes surprising ways in which understanding can be improved, including by adopting false beliefs, changing inferential dispositions, and acquiring tendencies to attend to certain types of phenomena. The only requirement is that after acquiring the belief (or whatever it is) one is better able to answer questions in the relevant subject matter. There is no reason to expect that everything that contributes to this ability is limited to a specific objective pattern. If explanations do in fact provide understanding, then the management account of understanding gives us reason to reject the mind-independent project. The ability to provide understanding is more deeply psychological.

The role that explanation plays in inquiry is fundamentally pragmatic. The issue of

explanation only arises when inquiry reaches a point such that a fact or phenomenon "cries out for an explanation." Explaining that fact or phenomenon then becomes important for the success of inquiry. The broader nature of the inquiry determines what kind of explanation is needed. Separating explanations from the context in which they are demanded already begs the question against mind-dependent accounts by suggesting that it makes sense to ask what counts as a good explanation for p without a context specified. The kinds of interests that can spark a search for explanation vary widely, depending on the situation, as do the kinds of explanations that satisfy the demand. We seek an explanation for a purpose, but this purpose could be to find information that will help future work, find an interesting fact, assign praise or blame, and more. This diversity goes so far that disparate explanations can satisfactorily apply to the very same explanandum, given different goals.

For example, suppose a man is found dead with no obvious wounds and no obvious cause of death. For a detective, the death cries out for an explanation as part of a larger project of making decisions about whether a crime has been committed and, if so, by whom. However, suppose the dead man was also a test subject in a clinical trial of a drug. The death would cry out for an explanation to a doctor working on the trial as part of a larger project of evaluating the safety of the drug. On the other hand, a family member might primarily be concerned with an explanation for the death as part of determining whether life can matter given that someone so important could die. The explanations that satisfy each of these agents are different. The explanation sufficient to satisfy the doctor need not settle anything about possible murder prosecutions or how to deal with senseless loss. When we, as philosophers without any demand for an explanation of why he died, theorize about potential explanations, asking whether something is a good explanation of why he died leaves open which of these explanations (or possibly others) is relevant.

In general explanation commands our attention because it plays a role in inquiry; inquiry is guided by the search for answers, some of which are explanations. Once we have explanations, we use them to help us navigate subjects related to the explanation. What

appears to set explanations apart as a philosophically significant category is that explanations serve to assist our inquiries. This is reason to suspect that explanation is a functional category based on practical role.[33]

To summarize the above points in an argument:

1. Acts of explanations are different from other acts because they serve cognitive goals in inquiry.

2. The various properties that serve cognitive goals in inquiry are not unified in any way in themselves.

3. Therefore, the unifying feature of explanations is the role they play in inquiry (rather than an objective, non-psychological feature).

Explanations in each instance are sought because they serve our cognitive goals. An explanation is an act that, given our cognitive background, ties the explanandum into our inquiry. Depending on our situation and our cognitive goals, vastly different arguments can serve to assist. There is no reason to expect that these all must fit into an objectively natural category.

In addition, the insights behind each of the objective theories of explanation provide evidence in favor of taking a mind-dependent approach. Each mind-independent theory of explanation derives its plausibility from the fact that explanations conforming to it often give us the ability to move forward in inquiry. An explanation of the D-N variety is a derivation from general laws, which can be used as a template to deploy again in the future. Unification explanations by definition unite several phenomena under a small number of patterns. These patterns can then be deployed for all of those phenomena after being learned once. An explanation that fits within the causal/intervention account provides one

[33]One might respond that the role of explanation in inquiry is given by the importance of inference to the best explanation in our reasoning, as in (Lipton 2004). While there is no space for a full response, I suggest that seeking to make an inference to the best explanation is not even close to the only reason we seek explanations. The role of explanations is broader.

with information about causally relevant factors that can give one the ability to manipulate the situation or predict outcomes of different scenarios. Contrastive explanations allow one to tell the difference between salient alternatives, discriminating among them. Each account picks out features of an argument that tend to make it useful for future inquiry. To the extent that the various objective theories of explanation are unified, it is by the fact that they are often useful in future inquiry. I contend that we have a multiplicity of objective accounts because there is not one objective feature that is coextensive with usefulness. The accounts are united in that possession of explanations conforming to them puts one in a position to do more than one could do with just knowledge of the thing explained. This suggests that explanation's role in inquiry is what is essential, and objective accounts merely identify different ways of filling the explanation-role.

The fact that their role is what unites explanations is especially apparent in everyday explanations. Everyday explanations often consist of no more than one or two sentences. The facts expressed can be arbitrarily distant from the fact to be explained. For example, someone may ask, "Why is the car upside down?" A good explanation might be: "The wind blew the flowerpot over," (suppose everyone had been talking about how the pot was perched just above the controls for the car lift in a mechanic's garage, and how the car was half on the lift, so would be tipped if the lift was activated). "Why are you nervous?" "It's Thusay," (when it is known that the second person has an interview that day). It is prima facie unlikely that there is any objective property that these explanations share that distinguishes them from non-explanations, but in each case they clear up confusion and allow those who grasp them to move forward in the inquiry that set the context. One class of everyday explanation that highlights this further are personal explanations of intentional actions. Personal explanations often appeal to factors that do or would make the action to be explained reasonable. The explanation can be detailed and complex, or one brief sentence.

This is not amenable to mind-independent accounts because these explanations require

an extensive psychological background before they can serve as good explanations. The best hope for non-psychological accounts is to claim that these are explanation sketches, mere pointers to more complete explanations. However, as argued in section 6.3.1, this response is problematic. In the end, the kinds of facts that can serve as good explanations have no discernible, in-principle limit; they simply must fill in a missing piece to allow inquiry to go forward. This suggests we ought to seek a mind-dependent account of explanation.

Similarly, the use of gestures or models as explanations does not fit well with objective accounts. You ask why my desk is in the hall. I point to my office. You look and notice the painters in the office and say, "That explains it." By pointing, I have provided the needed explanation. The act of pointing can be an act of explanation. Pointing allowed you to gain the ability to continue in your inquiry, and not always by pointing to a more complete explanation as posited by non-psychological theories.

We have seen that there is reason to be open to a mind-dependent account of explanation. If we want a unified account of explanation that covers scientific and non-scientific contexts while maintaining a close connection to understanding and inquiry, mind-independent accounts seem dissatisfying. A list of various kinds of explanation with no unifying nature does not have the same impact. It would be good to have a unified, understanding-connected, inquiry-driven account of explanation if possible. Such an account could help explain why seeking explanations is such an important part of what we do. Such an account could shed light shed light on inquiry in general. We can give such an account.

Below I will present a theory of explanation and argue that it avoids all of these problems. It provides a unified account of explanation that illuminates the role and nature of explanation.

5.14 Management account of understanding

My account of explanation depends on my account of understanding, which is presented in Chapter 3.

Management Account: to understand a subject matter is to have a sufficient degree of ability to manage cognitive resources to successfully answer questions within that subject.

Understanding is the ability to solve the following problem: there are far more unhelpful and irrelevant pieces of information and inferences available to us than there are helpful ones. We need to be able to find the useful, and the ability to do so within a subject is understanding that subject.

Based on this account, I argued that understanding why p is the same as understanding the subject matter defined by the question why p, and questions whose answers partially answer why p. Thus, understanding why p means having the ability to manage cognitive resources to reliably answer questions relating to why p. This is the form of understanding that is most at issue when dealing with explanation, since explanations are the answers to why-questions.

Note that according to this account, understanding a *subject matter* is primary. Understanding why a proposition is true falls out of the above account in the following way: we can treat the proposition as a subject area. This can only be accomplished in a context dependent manner, but briefly, the proposal is that in any context, the truth of a proposition relates to other propositions, and raises questions. These related propositions and questions form the subject matter for p. Understanding why p, then, just is understanding the p-specific subject matter. There are questions related to it so that the proposition itself defines a subject area.

What about questions that one has already answered? What can we say about understanding and its relation to those? First, one method of answering a question is to recall

the answer from memory when it is needed. The ability to do this is at least a degree of understanding. If an answer is stored in memory, but can only be accessed under special circumstances that do not obtain, one is worse off than someone who does not have the answer but can easily derive it when prompted. Thus, the ability to recall answers to previously answered questions by itself confers a degree of understanding.

There can be differences in the ability to answer questions to which one already knows the answer, if one is better at finding and recalling relevant information. Beyond recall, if one's cognitive abilities allow one to find that answer even if one does not already have it, one has the ability to answer that question. One's cognitive abilities are more adept at managing the resources available, and thus they constitute a greater degree of understanding.

For example, someone who understands engineering and train design can watch a train malfunction and immediately notice the features that diagnose what is wrong (in an ideal case). They are then able to call to mind exactly what to do to fix it. Out of all the things they could notice, all the facts they could call to mind, and all of the solutions they could think of, they have the disposition, probably a hard-won disposition, to go straight to the solution. The degree to which one can do this is one's degree of understanding.

Thus, in order for explanation to provide understanding, it must put one in a position to answer questions in a relevant subject. Knowing the explanation should help one to immediately come to conclusions that are useful. We will see how this relates to explanation.

5.15 The Understanding Account of Explanation

While mind-independent accounts have problems, advocates of such accounts nevertheless agree that explanations are important because of their ability to provide understanding. Explanations play an important role in inquiry, much as understanding does. A good explanation must be able to fit into one's reasoning to allow one to discover useful information. The production of understanding is the reason we separate explanations from non-explanations. This feature of an explanation is central to its use and its value, and I will argue that it is

in fact the essential defining feature of explanations. I propose that an explanation is an act such that, given a cognitive background, the act induces the ability to understand the explanandum. In other words, an explanation is an act that, when consumed by the intended audience, gives understanding in areas relevantly related to why p, barring performance failures.

Understanding Account : For audience A, a good explanation for why p is an act that, given the relevant cognitive background, provides understanding of why p in the manner intended.

A good explanation is offered with the goal of providing understanding to an audience through some means, and it succeeds in making understanding available in the manner intended. Given the relevant background (assuming no performance error), the explanation produces understanding. It has the capacity to give the relevant audience the raw materials to make the inferences that one's cognitive goals demand. The remainder of the section expands and defends this account.

First, the account specifies that a good explanation produces understanding in a *relevant cognitive background*, for several reasons. By moving away from a mind-independent account of explanation, we allow psychology to play a role. In particular, explanations are demanded within inquiry, and good explanations are tailored to meet those specific demands.[34] The demand in part is to bring about understanding in a particular audience. The demand and what satisfies the demand depends on the the psychology, including beliefs and cognitive capacities, of the intended audience. Consuming an act of explanation can only translate into the ability constitutive of understanding if there is a system that, when presented with the explanation, undergoes a change in abilities. The background is necessary to evaluate what effect explanations will have.

[34] Another way to look at it: some people lack the cognitive sophistication to understand why p no matter how cleverly one attempts to explain it. For them, there will be no informative explanation. For others, there will be different explanations for why p.

Understanding requires the harmonious functioning of a variety of cognitive capacities and beliefs. The relevant cognitive background is similarly diverse, including background beliefs, inferential capacities, inferential dispositions, attention, and the organization of the mind. Thus, the account says that an act is an explanation just in case when added to all of this, the result is an overall system that can find valuable information relating to why p. This is a general description of what a cognitive background is and why it is needed, but what specific background is relevant for any given explanation?

We can answer this question by considering the goal of an explanation and to whom it is directed under what conditions. When an explanation is offered for purely individual needs of a particular situation (as in most everyday cases, for example), the relevant background is simply the cognitive state of the intended consumer of the explanation. If John is seeking an explanation for why sharks are fish simply for his own sake (not for the sake of science, or the universal progress of humankind), then John's cognitive background is relevant. Thus, with this as the relevant background, an act is a good explanation just in case when an agent accepts the explanation, given their beliefs they are able to answer questions relating to the explanandum. An explanation should be able to fit within the knowledge and cognitive abilities of the agents involved in order to provide understanding. Again, since understanding is an ability, this definition does not require that an act is only an explanation if the agent actually succeeds in using it to find answers. They only need to have the ability, whether they go on to exercise it or not (barring performance errors in the consumption of the explanation).[35]

The relevant cognitive background is not necessarily always the psychological properties of a single agent. It may be the case that the relevant background is the shared background of multiple agents. If Mary asks Elisa for an explanation for why the door is locked as part of a joint project for which the dormancy of a volcano raises questions, then it is the joint background that matters. An explanation is good to the extent that it allows

[35] See (Sophian 1997) for a discussion of the competence/performance distinction.

both agents to answer these questions correctly to their own satisfaction. We need to consider what they share that allows the explanation to work. There are two ways to look at this: first, it could be that the relevant background is just the beliefs and abilities that both Mary and Elisa have, excluding the ways that they differ. If this is correct, then the explanation would need to be able to fit into this restricted system and give it the ability to go on. This does not seem plausible given the nonmonotonocity of reasoning. Beliefs that are not shared could intervene and leave each individually and both jointly unable to answer questions despite the abstract system they share having that ability. This leaves the idea that in this case an explanation must fit each of Mary and Elisa's backgrounds individually. Thus, it should give Mary the ability and Elisa the ability to answer questions.

The relevant cognitive background also might not involve agents who are present at the time, or perhaps might best be seen as the cognitive abilities of a collective agent. This may be the case in scientific explanation. In science, the goal is to provide an explanation for the scientific community. This does not necessarily require any one person to fully understand on their own. No one person needs to gain the capacity to answer questions related to why p. Instead, a good scientific explanation should allow the community to work together to advance and answer questions. The explanation should allow the group to see how to divide cognitive labor so as to use everyone's cognitive abilities to answer questions. For example, a good explanation of global warming should give the climate science community the ability to answer further questions in that field as a collective. The group as a whole can be viewed as one system, comprised of individuals with their own abilities, but taking part in an organization that disseminates information, makes decisions about resource allocation, and performs actions. In this way explanations are acted upon. It does not matter what any particular agent knows or can do, but what the scientific community as a whole can do. A good explanation should allow the community to understand, and thus it is this collective capacity that is relevant.

The understanding account aligns with the observation that explanations are sought because they fulfill our cognitive goals. A phenomenon cries out for explanation because it sparks some kind of interest, whether practical in nature or just a burning curiosity of a mystery to be solved. An explanation fits well within one's cognitive system, filling in gaps of confusion and giving one the ability to answer questions in the subject. Explanations are rarely if ever sought purely for their own sake; they are sought for the effect they will have on one's cognitive system. Even if the effect is just the satisfaction of curiosity, it still requires a change in cognitive state. One understands a subject matter in virtue of being able to fulfill one's cognitive goals within that subject, by being able to manage resources in order to find valuable information. This is exactly what we look for in potential explanations.

To illustrate, consider the Tacoma Narrows Bridge collapse, a surprising event that led many observers to demand an explanation. In 1940, less than a year after its completion, the bridge over Tacoma Narrows began to oscillate. The oscillations steadily increased until the entire bridge was twisting back and forth several feet in each direction. Finally, the bridge broke apart and fell into the river. It is one of the most famous catastrophic engineering failures. Seeing the video of the collapse, one is almost inescapably led to demand an explanation: why did the bridge collapse? Part of the reason is that the fact that the bridge collapsed raises many questions relating to bridge safety and design. Another reason is that most would not expect a bridge to behave that way, and so it raises questions about the behavior of constructions.

Suppose in response to the demands for explanation, an agency presented information about the forces on the parts of the bridge over time, the stresses and strains, and the materials properties, all of which showed that breaks would occur in certain locations, leading to the bridge falling into the river. In almost any natural context, this is not a very good explanation for the collapse. One could respond, "These are the specific processes that led up to the collapse, but that does not satisfactorily explain why the bridge collapsed."

It does not allow us to find our way to important related conclusions very efficiently. We could use the same laws to model and examine other bridge designs, but it would not be efficient. However, if we have an explanation that identifies some particular feature and its relation to wind in such a way that generalizes and is easily applicable to other bridges, then we have something to work with. It allows us to determine what could prevent this kind of failure, whether the materials were to blame, whether the climate played a role, etc. This is a much better explanation, because it fits into our cognitive economy much more nicely. Given the way we think, due to the demand to solve the problem of managing cognitive resources within the domain of inquiry relating to the fact that the bridge collapsed, we can use those facts efficiently.

When we look for an explanation, often we do not take the thing to be explained to be entirely singular. There is often further interest beyond the single event that attaches to it. In those cases, we seek an explanation that allows us to build on what we know about the phenomenon to be explained. We want to explain a bridge collapse so that we can make better bridges, and so we can tell which bridges are at risk. We want to explain the photovoltaic effect so we can see where else it might occur and how we might make use of it, and so we can uncover more features of matter. We find some events to be signposts for future discovery. In a properly conducted inquiry, it is sometimes clear that there are facts related to the fact that p that will tie into one's cognitive background in helpful ways. A good explanation opens that for us.

When we know a fact but investigate into the explanation behind it, we are seeking more than just to gain knowledge; we seek to understand. Explanations are the answers that we seek for that purpose. Explanations are sought because of their ability to provide understanding. This has previously been taken as a defining feature of understanding, but now we have an independent account of understanding that is not identical to having an explanation. Since understanding is not simply defined in terms of explanation, the connection between the two cannot be secured in that way. With understanding not definable

in terms of explanation, but such a close connection between the two, that gives us further evidence for the understanding account.

We often seek explanations to assist future inquiry, as argued above. This implies dependence on context to determine what counts as a good explanation. The detective wants to explain the death to know how to proceed regarding making arrests or finding evidence for a prosecution. The inquiry is finished when the detective is in a position so that future information regarding the crime can be organized in accordance with the broader epistemic goals. In other words, explanations are sought and are good exactly when they give one understanding in a relevant subject. The understanding account predicts exactly the kind of relativity that we in fact find.

Another advantage of the understanding account is that it allows the brief explanations that we give in everyday life to count as genuine explanations. The simple one sentence explanations work precisely because they fit within a shared background of knowledge and recognitional capacities and tendencies to notice certain patterns over others. With all of that shared, often it only takes a sentence to complete the ability to move forward with inquiry. Once you know that the flower pot fell over, you notice the lever that it fell on, you then remember the connection of the lever to explosives, and you have come to expect that kind of reaction from that kind of explosive. Learning of the flower pot, given what you are apt to notice and do with that information, you can answer questions that arise from the fact that the car is upside down. It all fits together.

5.16 Handling the problems of others

At this stage, we shall return to the cases that cause difficulties for mind-independent accounts of explanation. I will argue that the understanding account can deal with each of these cases, and thus it is a contender as a viable account of explanation. The problem cases to consider are:

1. Explanations in math and ethics,

2. Everyday and non-verbal explanations,

3. Adding premises,

4. The flagpole,

5. Man and birth control pills,

6. Explaining survival because effective weed killer, and

7. "Explaining" a universal generalization by enumeration.

Let's check these out one by one.

5.16.1 Math

Because the understanding account of explanation does not identify explanations in terms of ontological categories or metaphysical concepts, it has no difficulty allowing explanations in mathematics. It does not require mathematical entities or properties to enter into causal relations or take part in mechanisms. It does not restrict explanations to the identification of facts that could have been different or could have been manipulated.

All that is required for a mathematical proof to be explanatory is that when paired with the relevant cognitive system, the system has the ability to answer questions relating to the explanandum. In some situations a proof will do so, while in others it will not. As long as the explanation allows one to answer relevant questions, it can be a good explanation. Nothing separates these kinds of explanations out from the others.

5.16.2 Everyday and non-verbal

The understanding account of explanation also accommodates brief or non-verbal explanations. As long as the act produces understanding, it can be a good explanation. The account by its nature places no restriction on the type of act that can serve as an explanation. No exceptions need to be made to show how these acts can count as explanations. We do not

need to explain them away as non-explanations that play a similar role as explanation. If a salute produces understanding by design, it can be a good explanation according to the understanding account.

5.16.3 Adding premises

The understanding account correctly predicts that adding irrelevant premises makes an explanation worse. Take an explanation that involves presenting a valid deductive argument. Add a random extra true premise. Odds are, the explanation is now worse. Why? In general, we assume that the facts included in an explanation are relevant, so when an explanation includes something irrelevant, we get confused and try to figure out how the random premise connects. This hinders the ability to use the good explanation as it is. We have the sense that we must be missing something and so are in danger of going astray if we try to use the explanation. In short, extra premises hinder understanding, and that is what makes explanations that include them worse.

Further, there is the simple point that discarding the irrelevant, if you can do it at all, introduces an unnecessary step that increases the number of possibilities of error. This lowers the degree to which one can answer questions reliably. The chance of failure can be eliminated by sticking to only the relevant. The understanding account does well here, giving the right answer for the right reason.

5.16.4 The flagpole

A problem facing more than one mind-independent account is the implication that we can explain the height of a flagpole by reference to the length of its shadow. Intuitively, while the length of the shadow can figure in a derivation of the height of the flagpole, it does not explain the height, or at best it is a bad explanation.[36] What does the understanding account have to say?

[36]We ignore deviant cases, such as someone who built a flagpole in order to have a shadow of a certain length, and other odd scenarios that might make the shadow length a good explanation.

The answer here depends to some extent on human psychology. We are designed and trained to think about physical objects in causal terms in most cases. While this is not universally true (as argued above), it is still a powerful way of reasoning about the world. In a case like this where there is an obvious causal story, we will be much more comfortable understanding the situation in those terms. The height of the flagpole, given the conditions, causes there to be a shadow of the given length. We can process this easily, and it allows us to answer other questions about the situation.

On the other hand, when we learn about the length of the shadow, what does that allow us to do? The properties of the shadow are not robust and cannot be used efficiently in deriving related facts. Most derivations will involve going back to the height of the flagpole from the length of the shadow, and then reasoning from the height of the flagpole because that is how we prefer to think. This introduces more steps, and thus more opportunities for error. This decreases one's degree of understanding.

If there were another culture or species who found reasoning based on shadows to be preferable and if they could use facts about shadows to reach useful conclusions, would we have to take them to be wrong about the ability to explain the height of a pole in terms of the height of the flagpole? There is no reason in principle to rule this out. Given our psychology, however, the explanation in this direction will tend not to be useful.

5.16.5 Man on birth control

Some mind-independent accounts, such as D-N and unification accounts, imply that we can explain a man's failure to become pregnant by the fact that he took birth control pills. However, if this is an explanation at all, it is a terrible one. Does the understanding account tell us the same thing?

In order to count as a minimally good explanation by the understanding account, this explanation must fit into our background cognitive situation in such a way as to allow the discovery of valuable information relating to the man's failure to become pregnant. What

does the fact that he took birth control allow us to discover? How does it fit within our overall cognitive structure to allow us to draw conclusions? It does not help.

Again, one makes an assumption of relevance when someone tells one something. Thus, if told that the cisman is not pregnant because he took birth control, one might reasonably expect that birth control is relevant to the question of cismen becoming pregnant. This does not aid inquiry regarding pregnancy; in fact it may hinder it by encouraging one to seek out useless information about the birth control status of cismen.

Furthermore, given that explanations can be better or worse, the fact that there is another explanation that is much better, much more useful, and so obvious, shows that the birth control explanation is degenerate and bad, to the point that we are reluctant to call it an explanation at all. Operating on the basis of it makes one's inquiry more prone to failure, and thus it is a bad explanation. The way we are, the way we reason, we are not helped by it.

5.16.6 Weed killer

One can explain the survival of a weed by reference to the fact that it was sprayed by a weed killer that kills 99% of weeds it touches. How can one explain an event in terms of facts that render it almost certain not to happen? One has to consider for what further ends in inquiry one might seek an explanation for the continued survival of a weed. Here the most common type of scenario is trying to understand weed removal from gardens. One might have questions about whether one applied the weed killer incorrectly, or whether one missed a spot. One might wonder if something needs to be done differently next time. Once one knows about the weed killer's 99% kill rate, one knows to expect some weeds to survive and knows not to be concerned when a few weeds survive. The main questions that the survival of the weed raised are resolved by learning the effectiveness of the killer. One is once again able to understand the situation well enough to continue.[37]

[37]There could be other situations in which this is not a good explanation. If the weed killer manufacturer wants to know why a weed survived and they seek this information for the purpose of improving their product,

5.16.7 Enumerative "explanation"

Some accounts of explanation imply that one can explain enumeratively a universal generalization all F's are G's, by saying for each F that it is a G and then adding that there are no more F's. Yet it is hard to envision a situation in which this would count as an explanation, let alone a good one. The understanding account shows us why: in any investigation in which questions arise regarding a universal generalization, having the information that each F is a G will give us nothing. We will be in no better a position with respect to understanding.

In order for something to be an explanation, it must fit into a background way of thinking. It must be effective within the agent's or community's economy of cognitive resource management. An explanation should be a series of facts such that when accepted by someone with understanding, it allows for the solution of problems and the ability to answer related questions. There are many ways for this to happen. This is why each individual account of explanation seems incomplete, why there is a push toward pluralism, and why objective criteria come up short. The key is that given the idiosyncrasies in the way people think, certain arguments and series of claims, when presented and accepted, allow them to solve related problems. There is no a priori restriction on the form these explanations can take, but given our cognitive systems, there are limits on what will actually work. However, given the diversity of cognitive abilities , many different forms of argument can serve as explanations.[38]

The account can accommodate the cases that have traditionally posed problems.

this explanation fails. This is also captured by the understanding account. The explanation here does not allow them to answer the relevant questions, thus does not produce understanding.

[38]For just one aspect of the diversity of cognitive capacities, consider dual-process accounts of reasoning, according to which we sometimes use fast, imperfect heuristics, and other times use slower, more rigorous reasoning. See (Evans 2003) for discussion.

5.17 Priority of understanding

In leaving behind mind-independent accounts of understanding, one already undermines the case for explanation being prior to understanding. There is no account of understanding that includes explanation as an objective component of understanding to which we add a psychological component. Explanation is not objective in the relevant sense. In addition, we have shown above that analyzing explanation in terms of understanding provides a satisfying account of explanation, providing further evidence that understanding is prior to explanation. Completing the case for the priority of understanding, in this section we show that it is possible to have understanding without having an explanation.

I have elsewhere argued that knowing an explanation is not enough to guarantee understanding. This does not by itself mean that understanding is not dependent on explanation, for knowing an explanation could be a component. How do we show priority?

First of all, consider what it takes to have a degree of understanding. One must be able to find valuable information within a given subject matter. This can in principle be achieved purely in terms of the inferential tendencies of the mind without any input from knowledge or beliefs at all. It is possible to understand without any knowledge at all as long as one is disposed to perform the right cognitive operations. Suppose one's visual system processed visual input in a way that automatically produced conclusions about the paths of falling objects. It is possible to do this without relying on prior beliefs. Something along these lines could constitute understanding, yet there is no explanation involved.[39] Understanding can, at least in principle, be entirely based in perceptual and motor reactions: as long as those reactions allowed one to answer questions in the relevant subject matter. One just is able to navigate a subject matter. Whether this actually occurs, it is in principle possible, which is enough to show that understanding without explanation is possible.

However, it remains true that explanations can provide understanding. In fact, that is

[39] Also, (Lipton 2009) argues for the possibility of understanding without explanation, though I would tend to classify his examples as explanations.

the defining feature of explanation. Explanations are special ways of gaining understanding: they are the method of increasing understanding that can be explicitly presented to others. Much of the process of acquiring understanding is implicit; it involves habituation of patterns of reasoning, gaining recognitional capacities that allow one to jump to the relevant over the irrelevant. We cannot explicitly state or demonstrate to others how we do most of these things. However, explanations are the publicly available components of an increase in understanding. The value of understanding and the convenience of communication guarantees the value of explanations. They have this value precisely because they are defined in terms of understanding. Understanding is prior.

5.17.1 Too lax?

One concern is that the understanding account allows too many things to count as explanations, since it counts anything as an explanation as long as it will produce understanding in the relevant cognitive background. Consider The Slap: you ask a friend for the 1,000th time why there are no dinner rolls. Out of annoyance your friend slaps you in the face, and as a result, your head turns in just the right direction at just the right time to notice a squirrel putting the last of the dinner rolls behind a tree. You check, and notice all of the rolls are behind the tree. As a result, you understand why there are no rolls. Is the slap a good explanation? It does not appear to even be a candidate for a bad explanation, yet the understanding account seems to imply it is a good explanation. Or, consider The Conspiracy: while wondering why a rock flies straight when thrown from a spinning sling, someone suggests it is because shape-shifting reptilians control everything. This reminds you of David Icke, which reminds you of the teal jogging suit he used to wear, which reminds you of a race you once ran called the Inertia 5k, which reminds you of inertia in physics, which then inspires you to apply the concept of inertia to the case of the rock in the sling and thereby understand why the rock flies straight. Is the appeal to reptilians a good explanation? The Slap does not appear to be a candidate bad explanation, and yet

it is an act that leads to understanding, and so appears to fit the understanding account of explanation. The Conspiracy, though intended as an explanation, is an atrocious explanation. Yet it leads to understanding, so the understanding account appears to imply that it is a good explanation. This is a problem, if true.

However, there is no problem here. First of all, notice that no matter what else you want to say about the slap or the conspiracy theory, at the end of each case, you have the explanation you were looking for. Your explanatory demand has been fulfilled. The issue is not about whether you end up with a good explanation—because you do—the issue is whether the acts are successful acts of explanation.

While the acts do produce understanding, the understanding account implies that neither is a successful act of explanation. An act cannot be a successful act of explanation if it was not intended to produce understanding in the manner in which it was actually produced. The Slap was not intended to provide understanding at all; it did so purely accidentally. The agent presenting The Conspiracy intended to provide understanding. However, the agent intended to produce understanding by way of convincing you that reptilians control the world. While the act of telling you about the reptilians caused you to understand why rocks fly straight out of slings, it did not do so as intended. The intended effect occurred, but it was not a successful act of explanation. Thus, both cases are correctly classified by the understanding account.

On the other hand, suppose that the person slapping you and the person spouting conspiracy theories are sufficiently clever and sufficiently knowledgeable to realize that the slap or conspiracy theory would lead you to understand in just the way that they do. Suppose that they knew the series of cognitive steps your mind would take, and deliberately slapped you or spouted conspiracy theories to get you to understand via that series of steps. In that case, the person could perform those acts with the intention of providing you with an explanation, and succeed on purpose in doing so. These could be explanations, but in ways which are not overly problematic. Are these good explanations, according to the un-

derstanding account? (The slap is a needlessly painful way to explain something, but let us set that aside for now.)

Since in each modified case the act was meant to provide understanding, and succeeded in the way intended, they do qualify as acts of explanation. However, the intended method in each case is oblique, relying on roundabout paths dependent on fleeting circumstances and psychological idiosyncrasies in order to achieve understanding. This has two effects: (i) it renders the explanations of limited use, and (ii) it renders the explanations fragile. Regarding (i), the slap only worked because you and the squirrel happened to be positioned just right at just the right moment for a slap to work. The conspiracy theory only worked because you happened to have a series of psychological associations that led you to understanding. For almost any other agent in any other circumstances these acts would not work as explanations at all. Thus, other explanations are better. Regarding (ii), the mental processes that lead to understanding in either case are roundabout and subject to several possible failure modes. If the slap was mis-timed or if it knocked your head in the wrong direction, it would have failed to give you your explanation. If anything had distracted you from any step in the chain of associations from the reptilians, you would not have found your explanation. Furthermore, they each require more steps in reasoning than readily available alternatives. They require more processing and are more prone to failure. Thus, while the understanding account does imply that they are acts of explanation, it also implies that there are better explanations available in each case.

This implication does not strike me as problematic. It appears correct in each case to say that you have your explanation as a result of each act. If the results did not occur as intended, then the acts were not successful acts of explanation. If they did occur as intended, they are explanations but they are flawed. This accommodates our sense that something is deviant about these examples, but retains the fact that explanations are imparted.

At this point, one might be concerned that the understanding account of explanation is unfalsifiable, that nothing could count as evidence against it. Is this true? This theory

would face a difficulty if something could be shown to be an explanation, even a good one, but did not allow for the solution of problems in a relevant context/subject area. The theory predicts that if we can identify the kinds of questions that are on people's minds when they ask for an explanation, and we know the patterns of reasoning they naturally use, we can figure out what will be a good explanation for them and what will not. This is a substantive prediction.

The reason that we seek out explanations is because they have the properties that allow us to use them in managing cognitive resources. They do not have independent value beyond the use we have for them. This is what sets them apart, the thing that makes them a significant category in the first place. They can still be a natural kind, but only to the extent that understanding, of the kind I am talking about, is a natural kind. The idea is this: the facts cited in an explanation can efficiently lead one to the event, and can efficiently lead one to related conclusions.

Conclusion

It is time to take stock of what we have. Explanation is a largely psychological notion; it cannot be captured by a mind-independent account. What makes an explanation a good explanation is dependent on the cognitive background of those involved. This allows us to account for the role that explanation plays in inquiry: pointing the way from facts that are known to the secrets they reveal that tie into our cognitive goals. A fact for which we demand an explanation is a fact that we believe is connected to further relevant facts, and a good explanation lays these further facts bare, making them available.

Note that this retains some degree of objectivity. The understanding account does not imply that explanation is just whatever feels right, it is not just an act that relieves the nagging feeling that an explanation is needed. Only an act that actually has the capacity to produce genuine understanding, that can give one the ability to answer the relevant questions, is a good explanation. The phenomenon known as *illusion of understanding* means

that something can seem like a satisfactory explanation, but fail to be one.

The category of understanding-producing acts is of independent interest. Identifying explanation with such acts explains their value for inquiry. We sometimes demand explanations by asking why questions. The class of explanations—intentional acts that succeed by design in fulfilling these demands—is a worthwhile category to study. It is crucially tied to understanding. This is yet another way in which understanding is indispensable in our cognitive lives, both in science and in everyday life, and it is a fitting end to our exploration of understanding.

BIBLIOGRAPHY

Achinstein, P. (1983). *The Nature of Explanation*. Oxford: Oxford University Press.

Alston, W. (1995). How to think about reliability. *Philosophical Topics Spring*, 1–29.

Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy 100*(7), 356–371.

Barnes, E. (1992). Explanation not unification. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1*, 3–12.

Bechtel, W. and A. Abrahamsen (2005). Explanation: A mechanistic alternative. *Studies in History and Philosophy of the Biological and Biomedical Sciences 36*, 421–441.

Boylu, A. (2010). How understanding makes knowledge valuable. *Canadian Journal of Philosophy 40*(4), 591–610.

Bromberger, S. (1966). Why questions. In R. Colodny (Ed.), *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, pp. 86–111. Pittsburgh: University of Pittsburgh Press.

Burnyeat, M. (1984). Aristotle on understanding knowledge. *Proceedings of the Symposium Aristotelicum 8*, 97–139.

Cartwright, N. (1995). False idealization: A philosophical threat to the scientific method. *Philosophical Studies 77*(2), 339–352.

Christensen, D. (1991). Clever bookies and coherent beliefs. *The Philosophical Review 100*(2), 229–247.

Clifton, R. (1998). Structural explanation in quantum theory. Manuscript.

Conee, E. and R. Feldman (1998). The generality problem for reliabilism. *Philosphical Studies 89*, 1–29.

Craig, E. (1990). *Knowledge and the State of Nature*. Oxford: Oxford University Press.

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

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

Dennett, D. (1987). Cognitive wheels: The frame problem in artificial intelligence. In Z. W. Pylyshyn (Ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, pp. 93–105. Earth: Ablex.

Dennett, D. (1991). Real patterns. *The Journal of Philosophy 88*(1), 27–51.

Diéz, J., K. Khalifa, and B. Leuridan (2013). General theories of explanation: Buyer beware. *Synthese 190*, 379–396.

Dorato, M. and L. Felline (2011). Scientific explanation and scientific structuralism. In M. Bokulich and A. Bokulich (Eds.), *Scientific Structuralism*, pp. 161–176. Dordrecht: Springer.

Dreyfus, H. L. and S. E. Dreyfus (1987). How to stop worrying about the frame problem even though its computationally insoluble. In Z. W. Pylyshyn (Ed.), *The robots dilemma*, pp. 95–112. Norwood, NJ: Ablex.

Earman, J. (1995). *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford: Oxford Uinversity Press.

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

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

Elgin, C. (2009). Is understanding factive? In A. Haddock, A. Millar, and D. H. Pritchard (Eds.), *Epistemic Value*, pp. 322–330. Oxford: Oxford University Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science 11*(1), 19–23.

Evans, J. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences 10*, 454–459.

Fernbach, P. M., T. Rogers, C. R. Fox, and S. A. Sloman (2013, June). Political extremism is supported by an illusion of understanding. *Psychological Science 24*(6), 939–946.

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

Gijsbers, V. (2013). Understanding, explanation, and unification. *Studies in History and Philosophy of Science 44*, 516–522.

Glymour, C. (1987). Android epistemology and the frame problem: Comments on dennett's "cognitive wheels". In Z. W. Pylyshyn (Ed.), *The robot's dilemma*, pp. 65–76. Norwood, NJ: Ablex.

Godden, D. R. and A. D. Baddeley (1975). Context dependent memory in two matural environments: On land and under water. *British Journal of Psychology 66*, 325–331.

Gopnik, A. (1998). Explanation as orgasm. *Minds and Machines 8*, 101–118.

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

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

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Hempel, G. (1965). *Aspects of Scientific Explanation*. New York: The Free Press.

Hempel, G. (1966). *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall.

Hintikka, J. (1988). What is the logic of experimental inquiry? *Synthese 74*(2), 173–190.

Hintikka, J., I. Halonen, and A. Mutanen (1999). Interrogative logic as a general theory of reasoning. In J. Hintikka (Ed.), *Inquiry as inquiry: A logic of scientific discovery*, pp. 47–90. Dordrecht: Kluwer.

Hitchcock, C. (2004). Beauty and the bets. *Synthese 139*(3), 405–420.

Kane, M. J. and R. W. Engle (2000). Working memory capacity, proactive interference, divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition 26*, 336–358.

Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research 66*(3), 612–640.

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

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

Kitcher, P. (1981). Expanatory unification. *Philosophy of Science 48*, 507–31.

Kitcher, P. (1989). Explanatary unification and the causal structure of the world. In P. Kitcher and W. Salmon (Eds.), *Scientific Explanation*, pp. 410–505. Minneapolis: University of Minnesota Press.

Klein, P. (2008). Useful false beliefs. In Q. Smith (Ed.), *Epistemology: New Essays*, pp. 25–62. Oxford: Oxford University Press.

Kornblith, H. (2007). Naturalism and intuitions. *Grazer Philosophische Studien 74*(1), 27–49.

Kowalenko, R. (2009). How (not) to think about idealisation and ceteris paribus-laws. *Synthese 167*, 183–201.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.

Lesher, J. H. (2001). On aristotelian episteme as 'understanding'. *Ancient Philosophy 21*, 45–55.

Levi, I. (2003). Counterexamples to recovery and the filtering condition. *Studia Logica 73*(2), 209–218.

Lewis, D. (1986). Causal explanation. In *Philosophical Papers, Volume II*, pp. 214–240. Oxford: Oxford University Press.

Lewis, D. (1988). Relevant implication. *Theoria 54*, 162–174.

Lipton, P. (1991). Contrastive explanation and causal triangulation. *Philosophy of Science 58*(4), 687–697.

Lipton, P. (2004). *Inference to the Best Explanation*. Oxford: Oxford University Press.

Lipton, P. (2009). Understanding without explanation. In H. de Regt, S. Leonelli, and K. Eigner (Eds.), *Scientific understanding: Philosophical perspectives*, pp. 197–233. Pittsburgh: Pittsburgh University Press.

McCarthy, J. and P. Hayes (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie (Eds.), *Machine Intelligence*, Volume 4, pp. 463–502. Edinburgh: Edinburgh University Press.

Michaelian, K. (2011). The epistemology of forgetting. *Erkenntnis 74*, 399–424.

Miyake, A., N. P. Friedman, M. J. Emerson, A. H. Witzki, and A. Howerter (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology 41*, 49–100.

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

Moore, R. C. (1985). A formal theory of knowledge in action. In J. R. Hobbs and R. C. Moore (Eds.), *Formal Theories of the commonsense World*, pp. 319–358. Norwood, NJ: Ablex.

Moravcsik, J. (1979). Understanding and knowledge in plato's dialogues. *Neue Hefte fur Philosophie 15*(16), 53–69.

Morris, K. (2012). A defense of lucky understanding. *British Journal for the Philosophy of Science 63*, 357–371.

Overton, J. (2013). 'explain' in scientific discourse. *Synthese 190*, 1383–1405.

Paik, E. S. and G. Schraw (2012). Learning with animation and illusions of understanding. *Journal of Educational Psychology doi: 10.1037/a0030281*, 1–12.

Pritchard, D. (2005). *Epistemic Luck*. Oxford: Clarendon Press.

Pritchard, D. (2008). Knowing the answer, understanding and epistemic value. *Grazer Philosophische Studien 1*(1), 20–50.

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science 45*, 206–226.

Riggs, W. (2009). Understanding, knowledge and the meno requirement. In A. Haddock, A. Millar, and D. H. Pritchard (Eds.), *Epistemic Value*, pp. 405–428. Oxford: Oxford University Press.

Rozenblit, L. and F. Keil (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science 26*(5), 521–562.

Salmon, W. (1990a). *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Salmon, W. (1990b). Scientific explanation: Causation and unification. *Crítica 22*(66), 3–23.

Salmon, W. (1998). The importance of scientific understanding. In Salmon (Ed.), *Causality and Explanation*, pp. 79–91. New York: Oxford University Press.

Schacter, D., G. Wig, and W. Stevens (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology 17*, 171–176.

Schaffer, J. (2004). From contextualism to contrastivism. *Philosophical Studies 119*, 73–103.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well taught" mathematics courses. *Educational Psychologist 23*(2), 145–166.

Sellars, W. (1953). Inference and meaning. *Mind 62*(247), 313–338.

Skemp, R. (2006). Relational understanding and instrumental understanding. *Mathematics teaching in the middle school 12*(2), 88–95.

Sophian, C. (1997). Beyond competence: The significance of performance for cognitive development. *Cognitive Development 12*, 281–303.

Sosa, E. (2001). Human knowledge, animal and reflective. *Philosophical Studies 106*(3), 193–196.

Sosa, E. (2010). How competence matters in epistemology. *Philosophical Perspectives 24*(1), 465–475.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Strevens, M. (1999). Objective probability as a guide to the world. *Philosophical Studies 95*, 243–275.

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

Strevens, M. (2009). *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

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

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science 69*, 212–233.

Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass 2/3*, 564–591.

van Frassen, B. C. (1980). *The Scientific Image*. Oxford: Clarendon Press.

Way, J. (2010). The normativity of rationality. *Philosophy Compass 5*(12), 1057–1068.

Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies 115*, 1–31.

Wiggins, G. and J. McTighe (2005). *Understanding by Design, Expanded 2nd Edition*. San Fransisco: Pearson.

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

Williams, M. (2004). Knowledge, reflection and sceptical hypotheses. *Erkenntnis 61*, 315–343.

Woodward, J. (2004). *Making Things Happen*. Cambridge: Cambridge University Press.

Xu, Y. and P. Wang (2012). The frame problem, the relevance problem, and a package solution to both. *Synthese 187*, 43–72.

Zagzebski, L. (2001). Recovering understanding. In M. Steup (Ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, pp. 235–251. New York: Oxford University Press.

Zagzebski, L. (2003). The search for the source of epistemic good. *Metaphilosophy 34*(1–2), 12–28.