UNDERSTANDING AS AN EPISTEMIC GOAL

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Stephen R. Grimm, B.A., M.A.

________________________
Michael DePaul, Co-Director

________________________
Ted A. Warfield, Co-Director

Graduate Program in Philosophy
Notre Dame, Indiana

July 2005
Among epistemologists and philosophers of science, one often hears that someone with understanding is able to “see” or “grasp” how the elements of a subject “cohere” or “fit together”—but just what is involved in the seeing or the grasping is usually left to the imagination. I argue that the most productive way to make progress on this issue is by first identifying the kind of explanation-seeking why-questions that drive the search for understanding in the first place. In particular, I suggest that if we can get a good grip on why a situation stands in need of explanation for us in the first place, then we will in turn be in a good position to determine how we might satisfy that need.

In this respect, I argue that a situation stands in need of explanation for us, and therefore inspires our why-questions, in virtue of our sense that there are various ways that we think the situation might have been. If a situation stands in need of explanation for this reason, moreover, then in order to satisfy the need for explanation our objective will be to identify what the difference between these alternatives depends on.
To say only this much still leaves a great deal underdetermined, however, for we might “identify” what the difference depends on in a number of different ways: for example, by having a true belief about what the difference depends on, by knowing this on the basis of testimony, and so on. What is distinctive about the state of understanding, I suggest, is the particular way in which we identify what the difference depends on: namely, by grasping what the difference depends on. I explore at length what exactly this grasping amounts to.
DEDICATION

For Rachael, my joy
CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... iv

CHAPTER 1: INTRODUCTION ................................................................................................. 1

CHAPTER 2: THE STATE OF THE QUESTION .......................................................................... 8

CHAPTER 3: THE NEED OF EXPLANATION .......................................................................... 43

CHAPTER 4: THE GOAL OF EXPLANATION .......................................................................... 63

CHAPTER 5: RELIABILITY AND UNDERSTANDING ............................................................... 97

CHAPTER 6: IS UNDERSTANDING A SPECIES OF KNOWLEDGE? .................................... 118

CHAPTER 7: EPISTEMIC GOALS AND EPISTEMIC VALUES ............................................... 150

CHAPTER 8: CONCLUSION .................................................................................................... 169

BIBLIOGRAPHY .................................................................................................................... 173
ACKNOWLEDGMENTS

If this dissertation has any merits, it is largely due not just to the penetrating feedback and helpful advice of my advisors, Mike DePaul and Fritz Warfield, but also—perhaps more importantly—because in thinking about a particular argument or way of putting things I would inevitably hear their voices at the back of my mind saying, “That’s not quite right....” Asking Mike and Fritz to be my co-directors was the best choice I made as a graduate student, and I will sorely miss their friendship and their presence down the hall when all this is over.

More generally as a philosopher, although for several years before I met him I was engaged in something that I would have called “philosophy,” John Greco taught me how to actually do philosophy. More than anyone, in attempting to look at arguments in as clear and honest a way as possible, I have tried to model my work after his own. (I’m pretty sure he modeled his haircut after me, however.)

During my 2003-04 academic year at Brown, I also had the distinct pleasure of learning from Greco’s own mentor, Ernest Sosa. If there is a more gracious person than Ernie, I have yet to meet him; certainly, no one I know of comes close to matching Ernie in terms of combining graciousness with extraordinary philosophical ability. In terms of the broad approach to epistemic
evaluation I adopt in this dissertation, moreover, Ernie’s influence is stamped on every page.

With respect to the details of this project in particular, beside my advisors I have learned the most from Jaegwon Kim, Peter Lipton, and James Woodward. Jaegwon’s notion of “dependence” emerged as more and more fundamental to the theory as things progressed, and Lipton’s and Woodward’s attempts to put the causal account on better footing helped me to see things that would have likely taken me years to see otherwise.

For helpful feedback and support along the way I would also like to thank Robert Audi, Jason Baldwin, Marian David, Catherine Elgin, LinDa Grams, Allan Hazlett, Montey Holloway, Tom Kelly, Matt Kennedy, Peter Kwasniewski, Alasdair MacIntyre, David Manley, Christian Miller, John O’Callaghan, Michael Pace, Alvin Plantinga, Catherine Robertson, Donald Smith, Douglas Sylva, Leopold Stubenberg, and John Turri.

Finally, I would like to thank my family. My father and mother, George and Eileen, along with my sisters and brother, Laura, Sharon, and Chris, have provided me with unstinting love, support, and encouragement throughout my time in graduate school.

Words suddenly begin to ring hollow when I think about the importance of my beautiful wife, Rachael, my sweet daughter, Abigail, and my little son, Theo. I dedicate this book to them, but especially to Rachael, my joy.
CHAPTER 1
INTRODUCTION

It is a commonplace among ethicists to claim that understanding is one of the great goods that makes life worth living,¹ and among philosophers of science to note that understanding is one of the driving goals of science.² Even epistemologists, who in modern times at least have been preoccupied not with the notion of understanding but rather with notions such as knowledge and justification, have increasingly begun to regard understanding as a valuable epistemic good worth theorizing about in its own right.³

Despite the recognized status of understanding in both science and everyday life, philosophical accounts of the nature of understanding have been conspicuously lacking. Among epistemologists, one often hears that someone with understanding is able to “see” or “grasp” how the elements of a subject “cohere” or “fit together”—but just what is involved in the seeing or the grasping,

¹ See, for example, Griffin (1986, ch. 4), Crisp (1997, pp. 60-61), and Raz (2003, pp. 48-49).

² According to Hempel (1965), for instance, “Our main concern has been to examine the ways in which science answers why-questions of the [explanatory] type and to characterize the kind of understanding it thereby affords” (p. 488). See also Salmon (1998a). Unsurprisingly, it is likewise a commonplace among scientists themselves to cite understanding as one of the primary goals of science. See, for example, Weinberg (1994, 2004) and Atkins (2003, pp. 357-63).

or what it takes for the elements in a subject to cohere with one another, is usually left to the imagination. After recently proposing that to understand a subject one must “somehow see the way things ‘fit together,’” for example, Wayne Riggs (2003) continues, “This is, of course, fairly uninformative. Unfortunately, there is very little literature on this notion of understanding, as distinct from the kind of understanding one has of a language” (p. 218).

Among philosophers of science, who in principle at least have long been interested in spelling out the kind of epistemic good that explanations are thought to provide, the lack of in-depth work on the nature of understanding is perhaps even more surprising. Michael Friedman (1974) noted some thirty years ago that philosophers writing on explanation,

have relatively precise proposals as to the nature of the explanation relation, but relatively little to say about the connection between their proposals and scientific understanding, i.e., about what it is about the relation they propose that gives us understanding of the world. (p. 6)

Writing a decade ago, Jaegwon Kim (1994) observed that the situation had hardly improved:

The actual theories of explanation that we have accumulated to date, such as Hempel’s covering-law theory, the causal theories of Salmon, Lewis, and Humphreys, the pragmatic theories of Bromberger, Achinstein, and Van Fraassen—don’t look much like theories of understanding or accounts of a type of knowledge. Although explanatory understanding is sometimes mentioned...it seems quickly to be lost sight of when serious theory construction begins, and terms like ‘understanding,’ ‘intelligibility,’ and ‘explanatory knowledge’ seldom make an appearance once the initial stage-setting is over. (p. 52)  

Kim later puts this point even more dramatically: “We see in retrospect that somehow the concepts that ought to have been center stage in the theory of explanation had been shunted aside. Think of a theory of justice that never mentions justice, or a theory of numbers that never mentions numbers. The anomaly would be exactly the same. In a nutshell, then, the trouble is that we have models, but not theories, of explanation” (p. 53).
Finally, bringing the picture up to date, Nancy Cartwright (2004) has recently insisted that the need for a philosophical account of understanding is urgent:

The need for a philosophical treatment [of understanding] is pressing. When does a false model provide understanding? What kind of understanding is involved? And what practically are we able to do with the understanding that we acquire from a blatantly false model? These are crucial questions about an important and widespread scientific practice—and a practice that frequently informs policy—for which we currently have no good answers. (p. 240)

After noting this substantial gap in the literature, philosophers such as Friedman and Kim (among others) have at least attempted to address the gap, if only by gesturing in the sort of direction where they think a satisfying account lies. But none claim to have done this to their satisfaction. As Kim points out, it is one thing to identify a pressing need for a theory of understanding, quite another to provide one: “Producing a usable account of understanding is the hard part; one scarcely knows where to begin” (p. 69).

1.1 The Scope of the Project

In this dissertation I will try to remedy this gap, at least in part, by offering an account of understanding, but I should begin by noting that the account is modest in several respects.

First, although the notion of understanding covers a vast amount of ground, I intend to address only a relatively narrow part of it. We understand (or fail to understand) many different kinds of things. As Catherine Elgin (1996) aptly notes,

We understand rules and reasons, actions and passions, objectives and obstacles, techniques and tools, forms, functions, and fictions, as well as facts. We also understand pictures, words, equations, and patterns.
Ordinarily these are not isolated accomplishments; they coalesce into an understanding of a subject, discipline, or field of study. (p. 123)

In what follows, however, I will basically limit my attention to our understanding of natural phenomena (broadly understood), and I will not pursue the question of how—if at all—the account I offer here can be made to apply, for example, to the kind of linguistic understanding we have of concepts or meanings.\(^5\)

Second, I will nowhere offer a formal analysis of understanding in terms of necessary and sufficient conditions. If, as some still think, providing such an analysis of a concept is the pinnacle of philosophical achievement, then with respect to the concept of understanding this dissertation can be thought of as a kind of bivouac partway up the mountain.

What I provide here is still genuinely an account of understanding, however, because it situates understanding on the philosophical map. Thus, for example, I ask and try to answer the following questions: In what sense can we have understanding in the absence of truth? Is understanding a species of knowledge? Is it the goal of explanation? Is it, more broadly, the epistemic goal? And what role do phenomenological elements such as the celebrated Aha! experience play in understanding?

### 1.2 The Basic Plan

The basic plan for the dissertation is as follows. After reviewing the state of the debate in Chapter 2, in Chapter 3 I argue that the most productive way to approach the sort of questions mentioned at the end of the previous section is by

\(^5\) This is not meant to suggest that concepts and meanings are not, in some sense, natural phenomena (for example, they’re hardly unnatural or supernatural), but I trust that the distinction I am trying to draw is clear enough.
first identifying the kind of explanation-seeking why-questions that drive the search for understanding in the first place. In particular, I suggest that if we can get a good grip on why a situation stands in need of explanation for us in the first place, then we will in turn be in a good position to determine how we might satisfy that need.

In this respect, I argue (roughly) that a situation stands in need of explanation for us, and therefore inspires our why-questions, in virtue of our sense that there are various ways that we think the situation might have been. If a situation stands in need of explanation for this reason, moreover, then in order to satisfy the need for explanation our objective will be to identify what the difference between these alternatives depends on.

In Chapter 4 I claim that to say only this much still leaves a great deal underdetermined, for we might “identify” what the difference depends on in a number of different ways: for example, by having a true belief about what the difference depends on, by knowing this on the basis of testimony, and so on. What is distinctive about the state of understanding, I suggest, is the particular way in which we identify what the difference depends on: namely, by grasping what the difference depends on. Indeed, I argue that dependencies of this sort are by their very nature made to be grasped. Because when two things depend on one another they are modally connected in some way, what grasping the modal connection involves is the ability to (as it were) manipulate the connection, thus “seeing” or being able to anticipate how a change in the status of one element will (ceteris paribus) lead to a change in the status of another element. I likewise propose (along with the results of current cognitive science) that we have a
special capacity to grasp dependencies in this way, a capacity that I call our “grasping faculty,” and that the exercise of this grasping faculty is among other things accompanied by its own distinctive, and distinctively desirable, phenomenology.

In Chapter 5, I ask how we should make sense of the fact that our grasping faculty seems to be prone to spectacular mistakes. After all, conspiracy theorists, astrologers, alchemists, and so on all seem to have “grasped” or “seen” some incredible things. The way to make sense of these failures is not to claim that our grasping faculty is unreliable, however, but rather that it is, at best, conditionally reliable. More exactly, whether our graspings latch onto genuine dependencies or not will depend in large part on the quality of our background beliefs.

In Chapter 6 I take up the question of whether understanding is a species of knowledge. Against the current tide in epistemology, I argue that it is. Just as knowledge in general requires that one be securely related to the object of one’s knowledge, so too, though in a more specific way, understanding requires that one be securely related to the dependency grasped.

Finally, in Chapters 7 and 8 I apply these results to contemporary debates about the nature of epistemic value and epistemic appraisal. It is often said, for example, that the truth per se has an intrinsic value for us just because, as naturally curious beings, we have a standing interest in finding out how things stand in the world. In Chapter 7 I argue that this is a mistake. Our curiosity is elicited not by an interest in the truth per se but rather by truths—or, better, by subjects—that have certain additional features. In particular, and appealing to a point made earlier in Chapter 3, I argue that our curiosity is elicited when we
think that certain situations might have been otherwise, and we want to figure out what the difference between these alternatives depends on.

In Chapter 8 I then point out that if these claims about epistemic value are correct then one popular view of the nature of epistemic appraisal—the “teleological approach” associated with Alvin Goldman and William Alston among others—loses a great deal of its appeal. I close, finally, by asking how we should think about the nature of epistemic appraisal in light of these results.
CHAPTER 2
THE STATE OF THE QUESTION

Knowing that something is the case and understanding why it is the case are clearly different cognitive achievements, achievements that can and characteristically do come apart.¹ I can know that the sky is blue, for instance, without understanding why it is. Or that the garbage disposal in my sink isn’t working, without understanding why it isn’t. Or, to consider a more explicitly scientific question, that the tiny particles suspended in liquid are moving without understanding anything about the ins and outs of Brownian motion. Are they moving because of the pressure of external forces? Or are they moving because of dynamic properties internal to the particles themselves? However we settle why-questions of this sort, the fact remains that I do know that the particles are moving.²

¹ At least, the difference seems clear in the modern world. As Michael Williams (2000) points out, in the ancient world things were rather different. Williams writes: “The classical conception [of knowledge] is rich as well as demanding. In coming to see how a proposition follows from simple, self-evident necessities, we do not just acquire evidence that it is true, we see why it is true: indeed, we see why it could not have been otherwise. In associating knowledge with demonstration, the classical conception connects knowledge with explanation and understanding” (p. 42).

² It would be possible to recast this point in non-factive terms, either by supposing that my belief that p isn’t an item of knowledge (because the belief is false or just not hooked up to the truth in the right way) or by supposing that the epistemic gain that I experience is something less than understanding because it appeals to a false explanans. Just how we should characterize such non-factive epistemic gains is a topic I will take up later.
When we move from knowing that p to understanding why p we therefore experience an epistemic gain of some kind. So how should we think about this gain?

2.1 Understanding as Explanation?

A natural first thought is that the something more that I grasp is an explanation. Specifically, that to understand why p is to grasp the explanation of why p. It is worth beginning with this idea, not because it is seriously defended in the literature, but rather because it will help us to get a better sense of how to think about the relationship between explanation and understanding.

The difficulty with this way of characterizing the epistemic gain we experience when we understand is not that it is false—in fact, I think it’s basically true—but rather that it is uninformative. Suppose we grant that understanding involves grasping an explanation—or, better, grasping what we might call an explanatory story of some kind. The problem is now that there are several varieties of explanatory stories on offer, ranging from dormative virtue stories to stories that claim to describe the deep structure of the world. So why do some of these stories seem explanatory while others do not? Why do some of these stories seem to offer an epistemic gain while others do not?

A plausible answer is that the goodness of explanatory stories is tied to their ability to generate understanding. In other words, that one of the tests of

---

3 That is, the kind of stories, made famous by Molière, that attempt to explain (for instance) someone’s current drowsiness by reference to the “dormative virtues” or “soporific powers” of something they recently drank.
the goodness of an explanatory story is whether it generates understanding. As James Woodward (2003) writes:

> It is a plausible constraint on what an explanation is that it must be something that provides understanding. To say that certain information is ‘part’ of an explanation or contributes to its explanatory import is to say that this information contributes to the understanding provided by the explanation. (p. 179)

This suggestion helps to put the normative relationship between explanation and understanding into clearer focus. If understanding is the good at which explanatory stories aim, then an explanatory story will be good only insofar as it is capable of leading to understanding. But what this suggests is that, just as the utilitarian needs a prior account of the good (or the valuable) in order to account for the notion of right action, so too we need a prior account of understanding in order to account for the notion of a satisfying explanatory story.4

Evidence of the normative priority of understanding can be found throughout the literature on explanation. Consider, for example, how Wesley Salmon (1998b) objects to Bas van Fraassen’s (1980) claim that a good explanation must “favor” the thing to be explained. According to van Fraassen, the goodness of an explanatory story comes from the fact that it shows that the thing we want to explain is more probable, given the explanation, than the relevant alternatives. But Salmon objects to this account on the grounds that it ignores the fact that we seem to understand improbable events (e.g., brown-eyed

---

4 As Paul Humphreys (2000) likewise notes: “[S]cientific understanding provides a far richer terrain than does scientific explanation and that the latter is best viewed as a vehicle to understanding, rather than an end in itself” (p. 267). According to John Haugeland (1998), the two notions border on the equivalent, but understanding is clearly the goal: “Philosophers of science speak more often of explanation than of understanding, but these come to the same: to explain is to render intelligible—to show that and how something can be understood” (p. 260).
parents giving birth to a blue-eyed child) just as well as we understand probable events (e.g., brown-eyed parents giving birth to a brown-eyed child). And if this is so, Salmon argues, then the goodness of an explanation cannot come from the fact that it shows how the thing we want to explain is more probable than the relevant alternatives. He writes:

The requirement of favoring has the consequences that we can explain the more probable outcome when it occurs, but not the less probable outcome when it occurs. Since, as noted before, it seems that we understand each alternative equally well, or equally poorly, the favoring requirement leads to an unsatisfactory asymmetry. (p. 322, my emphasis)

Whether or not Salmon is correct in thinking that we understand the brown-eye to blue-eye chain as well as we understand the brown-eye to brown-eye connection is a question we can put to one side at this point. What is worth emphasizing instead is his way of evaluating van Fraassen’s proposal: to test of whether a certain account of explanatory goodness is adequate or not we look and see whether an explanatory story with the supposed good-making features described by the account is capable of generating understanding or not.

Some of the most famous counterexamples to Hempel’s Deductive-Nomological account of explanation take the same approach. So, for example, if we are able to show that on Hempel’s theory a flagpole’s shadow can explain the height of the flagpole (rather than vice versa), or that a man’s failure to get pregnant can be explained by the fact that he faithfully took his birth control pills (rather than because of the rather more conspicuous fact that he is a man!),\(^5\) then by common consensus we’ve shown that there is something wrong with the

\(^5\) See Salmon (1989) for a thorough survey of objections to Hempel’s view.
proposed account of explanatory goodness. Such accounts of explanatory
goodness can’t be right, we conclude, because they allow for cases where the
good-making features identified by the theory are in place but the particular
explanatory story fails to generate understanding. As Margaret Morrison (2000)
notes, “There have been many criticisms of the D-N model, not because of the
way it handles unification but because of its failure to produce an ‘understanding’
of the processes to be explained” (p. 3).

I will have more to say about Hempel’s Deductive-Nomological model in
the following section, but for now we can safely conclude the following: any
attempt to give an account of understanding’s epistemic gain in terms grasping
an explanation will not take us very far for the simple reason that we evaluate
explanatory stories in terms of the contribution they make to understanding.

In the remainder of this chapter I will turn to two more substantial
attempts to account for the epistemic gain we associate with understanding.
Taking the accounts in order, I will first look at Hempel’s claim that
understanding a situation amounts to having a rational expectation that it will
occur, and second at causal accounts—especially David Lewis’s and Wesley
Salmon’s—that claim that understanding a situation amounts to knowing its
cause.⁶ I consider the causal account last not because it is the most novel (if

---

⁶ I use the language of “situations” here in order to bring the various accounts into line. In reality, however, there are significant disagreements about exactly what is explained: Lewis prefers “events,” for example, and Hempel “aspects of events.” I defend a preference for
“situations” briefly at the beginning of Chapter 3.
anything its roots go back the furthest\textsuperscript{7}), but rather because it comes closest to the view that I will ultimately attempt to defend here. Following tradition, however, I begin with Hempel.\textsuperscript{8}

### 2.2 Hempel: Rational Expectation

Hempel is well-known for his view that good explanations invoke laws. What is less well-known, however, is why exactly he thought that invoking laws was \textit{explanatory}; that is, why he thought that invoking laws could account for the epistemic gain we associate with understanding.

According to Hempel (1965), good explanations have the form of arguments in which the conclusion of the argument is a sentence describing the thing to be explained (the explanandum). Explanations, moreover, come in two basic types: Deductive-Nomological (or D-N explanations) and Inductive-Statistical (or I-S explanations).

In Deductive-Nomological explanations, the thing to be explained is deduced from the relevant set of antecedent conditions and universal laws. So, for example, a good Deductive-Nomological explanation might explain why the pavement buckled during a hot summer day by citing (i) the heat of the

\textsuperscript{7} Thus we find Aristotle arguing in the \textit{Prior Analytics}, “We suppose ourselves to have unqualified scientific knowledge of a thing, as opposed to knowing it in an accidental way in which the sophist knows, when we think we know the cause on which it depends.”

\textsuperscript{8} One prominent proposal that I will not discuss in this chapter is Friedman’s (1974) and Kitcher’s (1989) view that the epistemic gain that comes from understanding why P has to do with unification—specifically, with the ability to deduce P from larger and more general patterns that one accepts. Although it would take some work to show this, for one thing I am hopeful that the account of understanding I offer here at least partly accommodates the intuitive idea that understanding involves seeing how things fit into larger patterns. For another, Friedman’s and Kitcher’s proposals have encountered basic technical difficulties that make them difficult to salvage. As Peter Lipton (2004a), perhaps understating the problem, notes: “One salient difficulty with the unification model is that the notion of unification turns out to be surprisingly difficult to analyze” (p. 28).
pavement, together with (ii) the relevant laws relating heat to expansion. An Inductive-Statistical explanation substitutes probabilistic or statistical laws for universal laws in one of its premises.\(^9\) Not just any statistical law will do, however: according to Hempel the statistical law has to confer a highly probability or likelihood on the explanandum. So, for example, we can explain the fact that John Jones recovered from streptococcus by citing (i) the fact that Jones was treated with penicillin, together with (ii) a statistical law to the effect that streptococcus patients tend to recover when they are treated with penicillin.\(^{10}\)

Why are both types of explanation good or desirable from an epistemic point of view? That is, why are both forms of explanation *explanatory*? Hempel’s answer is that in both cases the argument allows us to *anticipate* or *expect* that the explanandum will occur. In other words, Hempel’s answer to the epistemic gain question is that working through a D-N or I-S explanation puts us in a position to rationally expect that the thing we want to explain will occur.\(^{11}\) As he writes,

A D-N explanation answers the question “Why did the phenomenon occur?” by showing that the phenomenon resulted from particular circumstances, specified in \(C_1, C_2, \ldots, C_k\), in accordance with the laws \(L_1, L_2, \ldots, L_r\). By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred. (1965, p. 337)

---

\(^9\) How exactly probabilistic or statistical laws differ from universal laws on Hempel’s view is an interesting question, though one I will not consider here.

\(^{10}\) The last example is from Hempel (1965, pp. 381-82).

\(^{11}\) Or, looking backwards, the knowledge that the explanandum was going to occur.
Naturally, Hempel grants that we rarely meet these conditions in the strictest sense; i.e., we rarely if ever have a complete D-N or I-S argument in our pocket. But to the extent that we understand the occurrence of some event, he thinks, we must at least have an implicit grasp of such an argument. For, lacking such a grasp, we would no longer be in the position to rationally expect the event described by the explanandum, and hence would fail to understand why it occurred.

This account has obvious appeal. The fact that someone is able to anticipate that a situation will follow from certain antecedent conditions is often a good indication that the person understands why the situation obtains. For example, if a mechanic is able to anticipate that, given a certain adjustment, the car will do such and such, this is generally a good sign that he understands why the car did such and such. Likewise, the chemist who is able to anticipate that the presence of a certain element will corrupt the experiment is likely someone who understands why the experiment was corrupted.

Nevertheless, the rational expectation model faces several well-known problems. Suppose you see the barometer falling and accordingly predict the onset of the storm.\textsuperscript{12} Here you certainly have a rational expectation that the storm is about to occur, but this expectation seems unrelated to any genuine grasp of why the storm is about to occur. Related problems crop up when we consider the prevalence of testimony in our lives. If you are an experienced vulcanologist and you tell me that Mount Rainier is about to erupt, then I

\textsuperscript{12} See Friedman (1974, p. 8).
certainly will form a rational expectation that Mount Rainier is about to erupt even though I might not even remotely understand why Mount Rainier is about to erupt.

In addition to these fairly standard objections, there are also revealing problems of detail in the neighborhood, problems that help to shed light on the role that rational expectation does have to play in understanding. Suppose that Paul eats five pounds of arsenic at a single sitting. I can combine this fact about Paul with the relevant laws of biochemistry and deduce the fact that Paul will shortly die. But now suppose that my expectation that Paul will die is based solely on the fact that I have competently deduced this conclusion from the relevant premises. Can performing such a deduction by itself really generate understanding?

To see why this question should give us pause, imagine that after deducing the conclusion I am asked whether Paul would still have died if he had eaten four pounds of arsenic instead of five, or six instead of five. From the mere fact that I was able to deduce Paul’s death from the relevant premises, nothing seems to follow about my ability to answer these nearby questions. And yet, if I am entirely at a loss with respect to nearby questions of this sort, if the only basis on which I take myself to understand why Paul will die is that I have deduced it from the relevant premises and I fail to grasp that four pounds of arsenic would have been plenty enough to kill him, then the idea that I understand why Paul died begins to seem odd.13

13 Of course, it is unrealistic to suppose that I can answer all of these “what if things had been different?” questions. For example, I might know enough to know that 4 pounds would be
We might therefore agree with Hempel that understanding involves an element of rational expectation, but the rational expectation must come, as it were, from the right source: not just from testimony, and not simply from deriving the thing to be explained from a law, but rather from a grasp of the relevant law (or rule, or law-like generalization). In grasping a law what we appreciate is how the status of the thing we want to explain depends on the status of other features of the environment.

Since applying this suggestion to the relevant laws of biochemistry is beyond me, we can make the point instead in terms of a much simpler law: Newton’s second law of motion, \( f = ma \). Suppose then that I know the law, and I know that the mass of a certain object is 8 kilograms and that it is accelerating at \( 5 \text{ m/s}^2 \). I am then asked whether the force produced by the object would be less than 40 newtons if the acceleration of the object were reduced by \( 1 \text{ m/s}^2 \) while keeping the mass constant. If I have genuinely grasped the law—if I am able to apply it to particular instances and “see” how a change in the value of one of the variables will lead (ceteris paribus) to a change in the value of another variable—I will say Yes. If I don’t know how to answer questions of this sort, however, then not only have I not genuinely grasped the law but arguably the explanatory power of the law will be lost on me.\(^{14}\)

---

\(^{14}\) Arguably one reason why laws and principles have proven to be so conducive to the progress of science is because they encode information about dependencies in a remarkably easy-to-grasp form. As Fred Dretske (1977) notes: “Laws figure in the explanation of their instance because they are not mere summaries of their instances.... I can explain why the current increased upon an increase in the voltage by appealing to the relationship that exists between the flow of the charge (current intensity) and the voltage (notice the definite articles). The period of a pendulum decreases when you shorten the length of the bob, not because all pendulums do that,
I will have more to say about the sense in which grasping how things depend on one another is tied up with understanding in the following section on causal explanation, to which we now turn.

### 2.3 Understanding as Identifying the Cause

As things stand, the common problem with the leading causal theories of explanation is that it remains obscure exactly why learning about causes should have the sort of epistemic payoff we associate with understanding. As Jaegwon Kim (1994) puts the point:

Take the causal theorist like Lewis and Salmon; we still need him to answer the following question: what does causal knowledge have to do with understanding? That is, why, and in what way, does the knowledge that some event, \( g \), caused \( e \) produce, or enhance, our understanding of \( e \), or help make \( e \) intelligible? In general, then, the explanatory realist [i.e., someone who claims that understanding is grounded in our knowledge of metaphysical relationships in the world] has the task of explaining how knowledge that the grounding relation \( R \) is instantiated in a given situation promotes our understanding of that situation. (p. 60)

As Kim notes, although causal accounts claim that the move from knowing that \( p \) to understanding why \( p \) involves acquiring knowledge about the cause of \( p \), what they conspicuously fail to explain is why exactly this represents an epistemic gain. What exactly is the epistemic reward (or payoff or benefit) that comes from acquiring knowledge about the cause?

---

**but because the period and the length are related in the fashion** \( T = 2\pi\sqrt{L/G} \). The principles of thermodynamics tell us about the relationships that exist between such quantities as energy, entropy, temperature, and pressure, and it is for this reason that we can use these principles to explain the increase in temperature of a rapidly compressed gas, explain why perpetual motion machines cannot be built, and why balloons do not spontaneously collapse without puncture” (p. 22).

---

15 Woodward (2003) is a recent exception to this tendency.
Consider, for comparison, how much more explicit Hempel’s account is on this very point. As we saw, on Hempel’s account the difference between merely knowing that p and understanding why p comes from being able to subsume p under laws. But he also has a story to tell about why this difference translates into a genuine epistemic reward or payoff: namely, that subsuming events under laws allows us to predict or anticipate the occurrence of events. However vulnerable to counterexamples Hempel’s account might be, he has at least not left us guessing when it comes to specifying the epistemic payoff we associate with a satisfying explanation.

I agree with Kim that it is this second question, the epistemic gain or reward question, which is critical. Moreover, I think that if we hold this question at the front of our minds, Salmon’s and Lewis’s specific ways of developing the causal account begin to look seriously incomplete. I will begin with Salmon’s account.

2.4 Causation as Visualizability

According to Salmon, the “fundamental philosophical question about the nature of scientific explanation” (1984, p. 259) is this: “What sort of knowledge is explanatory knowledge, and on what basis can we say that it constitutes or contributes to our understanding?” (p. 259). One point that stands out from the beginning, then, is not only that Salmon is sensitive to Kim’s epistemic gain question, but that in theory at least he takes answering this question to be the driving concern behind his whole project.
On one level at least, his answer to the question is straightforward. The defining feature of a good explanation is that it in some sense reveals causes—more exactly, a good explanation reveals causal mechanisms. Thus he claims that, “underlying causal mechanisms hold the key to our understanding of the world” (1984, p. 260). To understand what exactly he means by this, we need to take a brief tour through Salmon’s views on causality.

According to Salmon, the intersection of two processes is a causal interaction if both processes are modified in ways that persists beyond the point of intersection, even in the absence of further intersections (1998c, p. 71). The collision of two billiard balls therefore counts as a causal interaction on Salmon’s view because both balls are modified as a result of the collision: their structure is slightly different after being hit, their momentum shifts, and so on. Moreover, for Salmon a process counts as causal if it is capable of producing such a modification—in Salmon’s terms (adopted from Reichenbach), if it is capable of leaving a mark. Among other things what this implies is that in order for two processes to be causally relevant to one another they must be spatiotemporally connected in some way. It is only possible to transmit a mark, according to Salmon, or for one process to change the structure of another process, through direct physical interaction.

---

16 In later work, and in response to critics such as Philip Kitcher, Salmon softens this stance and claims that understanding can be achieved through “theoretical unification” as well as through identifying causal mechanisms (Salmon 1998a, 1998c). Taking that development into account, the complaint here would then simply be that Salmon’s account of the epistemic gain that comes from knowing causes is not sufficient for understanding.

17 While the details have changed over the years, the underlying picture—the one described in this sketch—has remained basically the same. See Salmon (2002) for his account of the progression of his view.
Bearing this in mind, we can now rephrase our original question as follows: What kind of epistemic reward or payoff is supposed to come from learning about spatiotemporally continuous mechanisms? Even though Salmon identifies this question as his special concern, his answer is surprisingly indirect and sketchy.\(^\text{18}\)

Consider, for example, the following claim from his 1984 *Scientific Explanation and the Causal Structure of the World*:

[C]ausal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand why certain things happen, *we need to see how* they are produced by these mechanisms. (1984, p. 132; my emphasis)

As a first pass, if we focus on Salmon’s vocabulary of “seeing,” the epistemic gain that comes from identifying causal mechanisms seems to be in some way visual. A good explanation of some phenomenon allows us, in some sense, to *see* the mechanism that produced it. Shortly thereafter he expands on this idea. Good explanations apparently work by revealing for inspection mechanisms that were once hidden or obscure. As he writes,

To understand the world and what goes in on it, *we must expose its inner workings*. To the extent that causal mechanisms operate, they explain how the world works. (1984, p. 133; my emphasis)

In later work, Salmon continues to describe the payoff that comes from learning about spatiotemporally continuous mechanisms along these same lines:

\(^{18}\) As Jaegwon Kim (1994), in criticizing Salmon’s failure to tell us how revealing causes is supposed to yield understanding, notes: “the present issue is never faced, at least not directly” (p. 61).
The aim of explanations of this sort is to exhibit the ways in which nature operates; it is an effort to lay bare the mechanisms that underlie the phenomena we observe and wish to explain. (1998c, p. 71; my emphasis)

Here again, the epistemic gain seems to have something to do with vision—or, more broadly, with visualizability or picturability. According to Salmon, scientists explore nature as though it were a “black box” whose contents are revealed (laid bare, made observable) through scientific investigation:

There is a different fundamental notion of scientific understanding that is essentially mechanical in nature. It involves achieving a knowledge of how things work. One can look at the world, and the things in it, as black boxes whose internal workings we cannot directly observe. What we want to do is open the black box and expose its inner mechanisms. (1998c, p. 77; my emphasis)

So what exactly is the epistemic payoff that comes from identifying causal mechanisms? Developing his black box image, a macro example of his story might go as follows. Suppose what you want to understand is why the hands on a particular pocket watch move as they do. In this case, a significant epistemic gain will evidently be made when you pry open the back of the pocket watch and reveal the inner movement of the watch: the various springs and levers and

---

19 Salmon likewise uses the “laying bare” language in another essay (1998b) where we contrasts his so-called “ontic” account of explanation with Hempel “epistemic” account. He writes: “To shift from the epistemic to the ontic conception involves a radical gestalt shift. It involves relinquishing rational expectability as a hallmark of successful scientific explanation. Instead of asking whether we have found reasons to have expected the event to be explained if the explanatory information had been available in advance, we focus on the question of physical mechanisms. Scientific understanding, according to this conception, involves laying bare the mechanisms—etiological or constitutive, causal or noncausal—that bring about the event to be explained” (p. 328; my emphasis).

20 Essentially the same claim is made, still emphasizing the visual or quasi-visual gain that comes from identifying causal mechanisms, in his earlier (1984): “The ontic [i.e., Salmon’s] conception looks upon the world, to a large extent at least, as a black box whose workings we want to understand. Explanation involves laying bare the underlying mechanisms that connect the observable inputs to the observable outputs” (p. 276).

21 Salmon himself uses a pocketwatch as his example of a “black box” in his (1998a, p. 87).
sprockets that were previously hidden from view. More exactly, after opening the watch you will be in a position to see how the various parts of the watch are spatiotemporally connected. What was once hidden to you is now revealed. And, as far as Salmon’s explicit discussion suggests, it seems to be this epistemic gain that we were after when we began searching for an explanation of the moving hands.

Being able to picture or visualize—how else are we to understand the metaphor of “laying bare”?—the spatiotemporal connections among various processes therefore seems not only necessary for understanding on Salmon’s view but also (apparently) sufficient. It is difficult to say all this about Salmon’s view with complete confidence, of course. As I have suggested, Salmon does not explicitly argue that the epistemic gain associated with revealing or laying bare mechanisms lies along the lines of picturability or visualizability. Since he characterizes things this way at every opportunity, however, it is at least worth considering how well the suggestion fares.

22 Like Salmon, James Cushing (1991, 1998) argues that our ability to “picture” the causal mechanisms that produce phenomena is necessary—and apparently even sufficient—for understanding them. He writes, “The argument presented here actually begins from the intuition, based on experience and on (some) history of physics, that understanding of physical processes must involve picturable physical mechanisms and processes that can be pictured” (1991, p. 341). Later he reiterates this idea, “My basic position in this paper is that, not an a priori, but rather an historicist argument shows that picturable models or explanations are necessary for producing in us a sense of understanding. This necessity is rooted in the way we think about and understand physical phenomena” (1991, p. 351). The “historicist” argument that Cushing offers in favor of his view is basically comparative. Which theories don’t seem to yield understanding? According to Cushing, the ones that fail to reveal a picturable mechanism. Thus for generations Newton’s theory of gravitation was regarded with suspicion because it seemed to depend on spooky “action at a distance”—in other words, action without intervening mechanisms. It was only after Einstein’s theory of general relativity gave us a picture of how gravity worked—with the mass of bodies bending the physical shape of space—that gravity began to seem like an intelligible, and therefore potentially explanatory, force.
2.5 Problems

The short answer is: Not well. Considered as an account of the epistemic gain we experience when we understand, the ability to visualize (picture, take in) spatiotemporally continuous mechanisms is neither necessary nor sufficient.

To begin with the claim about sufficiency, one obvious problem is that laying bare mechanisms—thus allowing us to visualize or take in the mechanisms—often manages to produce more confusion, more perplexity, than before. Consider the pocket watch again. In the typical sort of watch, the movement—and therefore the causal picture story—will be exceedingly complex, so complex that it is not at all obvious how merely taking in or seeing the movement will generate understanding.

The more fundamental problem, however, is that it remains unclear how visualizing (taking in, etc.) even the simplest of mechanisms is correlated with understanding. Leo Tolstoy ([1869] 1957) makes this point as elegantly as anyone in War and Peace:

No matter how long or carefully I observe the hands of my watch, the valves and wheels of a locomotive, or the buds of an oak, I know that I will not discover what makes bells chime, locomotives move, or the spring breezes blow. To learn these things I must shift my viewpoint and study the laws which govern these things. (Bk. 3, Prt. 3, Ch. 1)

Just how exactly Tolstoy thinks that a “study of the laws” sheds light on things is not clear, but his basic claim is surely right. Merely observing the movement of even the simplest of mechanisms does not magically produce understanding. At the very least, some other cognitive ability, one that Salmon never attempts to specify, seems needed to mine these observations in the appropriate way.
The ability to visualize or picture spatiotemporally continuous mechanisms is also not necessary for understanding. The best way to see this is by taking stock of the countless little things that we take ourselves to understand over the course of the day. For example: Why the volume on the radio just increased (because I turned the knob to the right). Why my knee is throbbing (because I just played basketball). Why the soda has lost its fizz (because it was left uncovered all night). Why my eyes are watering (because I am cutting onions). Why the TV just sprang to life (because my brother just clicked the Power button on the remote). And so on.

What do these examples of apparent understanding have in common? For now it is enough to offer a negative, but still significant, answer: Few if any of these examples involve me visualizing the spatiotemporally continuous mechanisms that connect causes to effects. Naturally, this is not to say that there are not such mechanisms in these cases, or even that I don’t take it for granted that there are such mechanisms. The point is rather that visualizing or picturing such a mechanism has no obvious role to play in a great deal of the understanding that we seem to enjoy throughout our lives.

What would Salmon make of such cases? Perhaps he would want to introduce a distinction along the following lines. Whatever other epistemic achievement someone in such a position might lay claim to, they are nonetheless not cases of scientific understanding. After all, how much of a scientist does one have to be to attribute my sudden tears to the fact that I am chopping onions? Call the onion-chopping, TV-remote control kinds of examples cases of lay understanding, and call the kind of understanding that scientists are after
scientific understanding. The idea would then be that although the ability to reveal and picture spatiotemporally continuous mechanisms is not necessary for lay understanding, it is necessary for scientific understanding. Moreover, since scientific understanding is the achievement we are really interested in insofar as we are curious beings who want to know how the world works, then the ability to picture mechanisms in fact turns out to be fundamental in the way that Salmon suggests.

The basic problem with this distinction is that it immediately makes an enormous amount of what looks like impeccably conducted science into non-science. Countless examples would bear this out, but the following interview with Hugh Sampson, a leading allergy expert from the Mount Sinai School of Medicine in New York, illustrates the general idea. Here Sampson is being interviewed by Norman Swan about the link between food allergies and eczema.

Norman Swan: So tell me about the work you’ve done relating food allergy to eczema, and just how it plays out.

Hugh Sampson: What we did was take children who had moderate to severe atopic dermatitis and bring them into a clinical research centre where we could really isolate all the potential factors that would trigger their eczema, and one of the problems you get into is there are things such as heat or humidity or irritants, potentially airborne allergens that all can play into the breakdown of the skin in these patients. So we would bring them into a unit where we could control all that, and then the single factor that would be altered would be their diet. And we would bring them in, put them on very limited diets, and then put them through a series of challenges where they would get a certain food in a camouflaged way so that nobody really knew what they were getting, and we would observe them for symptoms. And by doing that we were able to show that a large percentage of children would develop skin symptoms following ingestion of certain foods. One of the problems we’ve always had in this area is that when allergists do skin tests or get blood tests for food allergy, you oftentimes will get large numbers of positive responses, but when you actually challenge them, you find that only about a third of those really translate into clinical symptoms due to that particular food.
**Norman Swan:** ... [So] there’s no substitute for trying, you just have to try and eliminate and see whether they make a difference?

**Hugh Sampson:** Yes, really at this point in time there’s no way other than to do elimination and challenge to really know what the cause and effect relationship is.  

For concreteness, suppose that as a result of the elimination and challenge method Sampson identifies a link—better, a dependency—between eating eggs and eczema flare ups in a particular patient, call her Sharon. Sharon then eats some eggs and her eczema flares up right on cue, all under Sampson’s nose. What we want to say here, naturally, is not only that Sampson understands why the patient’s eczema flared up—because she just ate some eggs—but moreover that this understanding clearly qualifies as a form of scientific understanding, the conclusion of a long process of controlling for alternative influences and so on.

For our purposes, moreover, the important thing to bear in mind is that throughout all of this no spatio-temporally continuous mechanism was—in the requisite sense—open to Sampson’s view. Granted, he observed the eggs pass through Sharon’s lips, so there was a physically continuous process of some kind to be pictured. But this could not be a mechanism of the sort Salmon has in mind. After all, presumably Sampson could observe the same egg-passing-through-lips sequence in scores of other patients who would fail to develop eczema. But then, Sampson’s ability to grasp the existence of such a dependency without being able to visualize a spatiotemporally continuous mechanism shows

---

23 The full transcript of the interview, see Sampson (2002).
that the ability to visualize mechanisms of this sort is not required for genuine understanding, scientific or otherwise.

It might be pressed, however, that the reason why none of this amounts to scientific understanding is because of Sampson’s ignorance of the relevant biochemical story behind the flare up; in other words, the story that pinpoints the chemical properties of the egg that were in fact poorly received by Sharon’s immune system.

But while becoming familiar with the relevant biochemical story no doubt represents a gain of some kind, it is a mistake to suppose that learning such a story should disqualify Sampson’s earlier state as an instance of scientific understanding. After all, once you claim that scientific understanding requires us to think in terms of the very small (as it were), then the biochemical story too begins to look far too gross. For the really small story, we need to appeal to the language of physics, not the language of biochemistry. But I take it that this is an unwelcome result: Just as we want to say that biochemists might enjoy a properly scientific understanding of why Sharon’s immune system reacted poorly to the chemical properties of the eggs, so too we want to say that Sampson enjoys a properly scientific understanding of why Sharon’s eczema flared up.

More importantly in the context of our present discussion, it is hard to see how any of this helps to support Salmon’s own account of the epistemic gain that we experience when we understand. For assuming that learning the relevant biochemical story—for that matter, even the relevant physical story—represent an epistemic gain of some kind, in what sense does it represent an epistemic gain? On Salmon’s way of looking at things, the biochemical story represents an
epistemic gain because it allows us to visualize how—down to the cellular or molecular level—Sharon’s immune system reacted poorly with certain chemical properties of the egg.

But if this is all the epistemic gain amounts to then Tolstoy’s objection pops up all over again! Even though what we are observing now is much smaller—biochemical reactions—we are still, for all Salmon has to tell us, simply observing it. For the biochemical story to represent an epistemic gain, some other cognitive ability is clearly needed; but again, not only does Salmon not attempt to specify what that ability might be, from what we have seen so far it is doubtful that the exercise of this other ability will be deeply connected with the ability to visually spatiotemporally continuous mechanisms in the way Salmon suggests.

All in all, Salmon’s account seems to take much too literally the commonsense idea that the goal of explanation is to “see” how things are connected—or, to use Salmon’s preferred image, to unlock the black box of nature in order to lay bare its contents for inspection. If we think of “seeing” in a more metaphorical way, however, then recognizing “seeing” as an important and legitimate kind of epistemic gain begins to look much more plausible.

Specifically, on one way of unpacking the “seeing” metaphor, to say that someone “sees” how things are connected is really to say something along the lines of that the person grasps how things are connected; alternatively, that they grasp how things depend on one another. If that is the case, however, then, as a first approximation, the epistemic gain that would come from knowing the cause of a thing would involve “seeing”-cum-“grasping” what that thing depends on.
I will have a good deal more to say about this proposal in the following chapters. For now, however, on to David Lewis.

2.6 Lewis

David Lewis’s (1986) account of causal explanation is straightforward: “Here is my main thesis: to explain an event is to provide some information about its causal history” (p. 217). Clearly, this is a very natural idea. A good explanation of why I’m late to pick up my wife, for example, will point to an event from the recent past: the fact I got into an accident along the way. Likewise a good explanation of why the soda lost its fizz will point to the fact that the can was left open all night. In attempting to explain some event, in short, we characteristically appeal to some part of the event’s causal history, and a good or successful explanation is one that tells us something true about that part of the event’s history.

A great deal hangs of course on how Lewis characterizes the key terms: causal history, cause, and information. Starting with causal history, for Lewis a causal history is a relational structure whose relata are events, and one event E belongs to the causal history of another event E* just in case E* depends for its existence on E. Here is how he describes the sort of dependence he has in mind: “I take such dependence to obtain when the objective chances of some events depend counterfactually upon other events: if the cause had not been, the effect would have been very much less probable than it actually was” (p. 217).\(^{24}\) Lewis

\(^{24}\) Lewis has since modified his account of causation, rejecting the “probability raising” element, among other things. See Lewis (2000). One can only speculate about how Lewis might have altered his account of explanation accordingly, but since even his modified account of
does not go into detail about how we should think about the “very much less probable” provision. But the point seems clear enough. We think of the lit match as a cause of the fire because, were it not for the match, the fire would not have occurred. The fire in this sense counterfactually depends on the existence of the lit match, and the match can therefore be thought of as having a role in the causal history of the fire.

According to Lewis, every event therefore trails a causal history—a continually branching string of dependence relations—that extends back a vast distance in time. Indeed, the question for Lewis is not whether the branching history for an event is large or small, but rather whether it is “infinite” or merely “enormous.” Moreover, strictly speaking the information collectively contained in this history is the only set of information that deserves to be called “the” explanation of the event. As Lewis writes:

Among the true propositions about the history of an event, one is maximal in strength. It is the whole truth on the subject—the biggest chunk of explanatory information that is free of error. We might call this the whole explanation of the explanandum event, or simply the explanation. (pp. 218-19)

Every other explanation is therefore, at best, only a partial or selective description of this colossal structure of interconnected events.

Since for Lewis to explain an event is to provide information about its causal history, and since the causal history of every event is so extensive, then granted our many obvious cognitive limitations there is a clear sense in which we can never fully or completely explain an event. But many of our explanatory causation makes essential appeal to what might be thought of as “ancestral” counterfactual relations, the objections offered in what follows, it seems to me, would not be substantially affected.
stories strike us as good or successful. How is this possible? How does an admittedly incomplete snapshot of an event’s causal history manage to be satisfying?

Lewis’s answer is that the extent to which we take an explanation to be good or satisfying depends in large part on the sort of question we want answered in the first place. As he notes, our desire for causal information is usually more focused than accounts of explanation tend to acknowledge: we want to know why New York City blackout occurred, for example, and when we learn that it was due to a power disruption in Ohio our curiosity is generally satisfied. The surest way to ruin an analysis of explanation, he claims, is therefore to insist that there is some “ideal unit of explanation” (such as Hempel’s deductive-nomological model) in virtue of which we acquire a special insight into the event to be explained. Instead, “We should not demand a unit, and that demand has distorted the subject badly. It’s not that explanations are things that we may or may not have one of; rather, explanation is something we may have more or less of” (p. 238).

Even though there is no ideal unit of explanation, Lewis nevertheless claims that not all causal information is equal; in particular, he thinks that larger chunks of information tend to be more helpful than smaller chunks (pp. 226-28). But he stresses that suggestions along these lines are only rough guidelines. Whether or not you happen to find a particular bit of information about an event’s causal history satisfying or not is largely (if not exclusively) a function of the particular why-question you bring to a situation.
2.7 Objection 1: Explanatory Information

Perhaps the fundamental problem with Lewis’s account is that he too easily assumes an equivalence between accurate information (or, as he sometimes puts it, correct information) and explanatory information. That is, he too easily assumes that explanatory stories that provide accurate information about the causal history of an event are thereby good (or satisfying, or desirable)—even if not completely satisfying, at least to some degree. But accurate information and explanatory information are significantly different notions, as examples along the following lines suggest.

Jeb the police inspector takes himself to have a good explanation of why the house fire occurred: because a defective lava lamp ignited a stack of newspapers. He subsequently writes this information on page 17 of his police report and tells his partner, George that, “The cause of the fire is listed on page 17.”

George now presumably possesses accurate information about a central event in the causal history of the fire, at least in a sense that satisfies Lewis’s minimal requirements: were it not for the defective lava lamp, the fire would not have occurred, and its being listed on page 17 of the police report is a piece of accurate information about this event. But is this piece of accurate information about the event really explanatory information? Unless we are straining to

25 For related concerns about conflating accurate and explanatory information, see Ruben (1990, pp. 160-80).

26 All Lewis claims is that a good explanation needs to “specify” (p. 219) some portion of the causal history of the event; he places no limits, however, on how the specification should occur. Indeed, as he elsewhere writes, “Information about what the causal history includes may range from the very specific to the very abstract” (p. 220).
defend a thesis, it seems obvious that the answer is No. George is now no closer, epistemically, to understanding why the fire took place than he was before Jeb told him about the information in the police report. For someone interested in finding out why the fire occurred, this explanatory story, taken neat at least, would not be a satisfying answer.

Of course this much is true: George is now in a better position to acquire information which will further his understanding than he was before. Now, as it were, he knows where to look to acquire information that we will find satisfying; just as I would be in a better position to explain why the icecaps were melting if you handed me a recent article from *Nature* and told me, “The explanation’s in there.” But this kind of information clearly does nothing to facilitate his understanding of why the fire occurred.

Peter Railton (1981), whose account of explanatory information closely parallels Lewis’s,27 acknowledges that this “unilluminating yet accurate information problem” is a difficulty, but he attempts to minimize its importance. Railton writes:

The account of explanatory information just outlined is liable to trivialization. For example, suppose that an alpha-decay occurs, and suppose that someone proffers the following ‘explanation’: ‘the relevant ideal text contains more than $10^2$ words in English.’ For all that has been said here, this remark, if true (as it doubtless is), would count as conveying explanatory information and hence as some sort of explanation.... In order to avoid further complication, I will simply tolerate this kind of degenerate case, relegating it to the very low end of the continuum of explanatoriness. (p. 173)

27 Which is no surprise: Railton was Lewis’s student and Lewis at several points acknowledges his debt to Railton’s analysis.
Clearly something has gone wrong here, however. Any analysis that counts as (even slightly) explanatory a piece of information that is not even remotely explanatory has made a fatal mistake somewhere, just as any analysis of the concept living that counts as alive someone who is stone-cold dead has made a fatal mistake somewhere.

2.8 Objection 2: Depth

Another indication that something has gone wrong involves the notion of explanatory depth. One of the most striking aspects of understanding is that it comes in degrees. The novice mechanic understands a car engine passably, the expert understands it forwards and back. You might partially understand why terrorism is on the rise in the Middle East when you know about the religious conflicts in the region; when you combine this with a grasp of the economic conditions, you seem to understand still more.

A viable theory of explanation must therefore be able to make sense of, if not offer an explicit account of, such differences in explanatory depth. Lewis’s analysis of explanation is again valuable in this respect because even though he thinks that—if understanding is taken in the weak sense he is willing to tolerate— the more accurate information you possess, the deeper your understanding. Thus he claims: “But understanding why an event took place might, I think, just mean possession of explanatory information about it—the more of that you possess, the better you understand” (p. 228). Put in terms of one of the analogies he offers, explanatory information is like pieces of treasure

28 In Chapter 4 I will consider Lewis’s reservations about a “stronger” view of understanding.
scattered on the ocean floor; the more you manage to find, the richer your understanding.\textsuperscript{29}

It hard to see, however, how acknowledging that this account of explanatory depth follows from his view does not ultimately amount to a \textit{reductio} of the view. Suppose you claim to understand why Sam is slurring his words: because he just drank a fifth of peach schnapps. I likewise think Sam is slurring his words because he just drank a fifth of peach schnapps. But I also think something else: namely, that it was the \textit{alcohol} in the peach schnapps that was responsible for Sam’s condition. Not recognizing that it was the alcohol that was responsible for his condition, you seem to understand less about why Sam is slurring than I do, especially if your conclusion about Sam is based on a limited range of examples: perhaps, for example, you have only ever seen someone slur their words after drinking peach schnapps, and as a result you might think that peach schnapps is the only thing that can do the trick. If so, one reason why I will be in a better explanatory position than you, one reason why I will have a better \textit{understanding} of why Sam is slurring, is because I know that if Sam had opted for a bottled of red wine instead of the schnapps, then that too—due to the alcohol content of the wine—would have led to his slurring.

Can Lewis account for this difference in our understanding of Sam’s slurring? Arguably not. Neither of our stories seems to be naturally \textit{deeper} than

\textsuperscript{29} Thus, against Carl Hempel’s (1965) claim that “To the extent that a statement of individual causation leaves the relevant antecedent conditions, and thus also the requisite explanatory laws, indefinite it is like a note saying that there is treasure hidden somewhere” (p. 349), Lewis replies with an analogy of his own: “I say it is not like that. A shipwreck has spread the treasure over the bottom of the sea and you will never find it all. Every doubloon you find is one more doubloon in your pocket, and also it is a clue to where the next doubloons may be. You may or may not want to look for them, depending on how many you have so far, and how much you want to be how rich” (p. 237).
the other, in Lewis’s sense, because neither tells us more about the causal history—this huge branching structure of events—than the other. This result looks even more strained when we imagine how the alternative Lewisian story about explanatory depth might go. Suppose that you, unlike me, have accurate information about many more of the events that led up to Sam’s slurring his words—many more of the events that compose the huge branching structure of events that Lewis thinks of as “the” causal history of the event. For example, you know that Sam’s father-to-be accidentally bumped into his mother-to-be while passing on the train thirty years ago, thereby sparking a relationship that would lead to Sam’s birth. Moreover, you know that the earth cooled to just the right temperature to sustain life, thereby making it possible for creatures like Sam to exist. On Lewis’s view, acquiring information of this sort should help increase, if only incrementally, our understanding of why Sam is slurring his words. But that seems wrong; acquiring accurate information of this kind does not seem to enhance our understanding of why he is slurring at all.30

From this we can draw two conclusions. First, that at some stage, and most likely quite early on, there will be a point of epistemically diminishing returns when it comes to acquiring information about the events that—in Lewis’s sense—constitute the causal history of an event. Second, that when it comes to acquiring deeper understanding there seems to be an important horizontal dimension to understanding that Lewis, with his focus on the vertical (or

30 Of course, we could always make up an exotic story linking the two: perhaps Sam is drunk because he would rather never have existed, and he blames the meeting of his parents for his current misery. But for simplicity let’s imagine that there is no such exotic story in the background.
historical), cannot obviously accommodate. My awareness that the alcohol content of the peach schnapps made a difference to Sam’s slurring his words does not mean that I know more about the events that led up to the slurring than you but rather that (among other things) I am aware that if things had been a bit different—had he chosen the red wine instead—that still would have accounted for the difference between Sam’s slurring and his speaking clearly. To have, in this sense, a better grasp of what made the difference is not simply to have more information about the causal history of the event. What instead seems more relevant is my ability to answer a greater range of what James Woodward (2003) has nicely called “what-if-things-had-been-different?” questions.

2.9 Objection 3: Chanciness

A third and final worry, this time concerning the role of chance in explanation, presses essentially the same point about difference making. According to Lewis, since understanding comes from possessing accurate information about causal histories, then it would seem to follow that chance events—say, the radioactive decay of a particular uranium atom—can be explained and hence understood. After all, chance events seem to be events like any other; they are have own distinctive histories. Taken in a certain way, Lewis gladly accepts this implication (p. 231).

Lewis accepts this implication only “in a certain way,” however, because he thinks a distinction needs to be made. On his view, when certain kinds of questions—what he calls contrastive questions, or questions of the form “why

31 The horizontal/vertical distinction is from Hitchcock and Woodward (2003a, 2003b).
this *rather than that*?—are posed, providing accurate information about the causal history of an event fails to be explanatory. And the reason is that there is simply nothing about the causal history of a genuinely chance event that tells us why this rather than that outcome came about. As he writes:

A contrastive why-question with ‘rather’ requests information about the features that differentiate the actual causal history from its counterfactual alternative. There are no such features, so the question has no positive answer. Thus we are right to call chance events inexplicable, if it is contrastive explanation we have in mind. (p. 231)

In addition to contrastive questions, however, Lewis argues that there are also *plain* questions. Instead of asking (as contrastives do) “Why this rather that that?” plain questions simply ask: “Why this?” Moreover, if our plain question is something of the form “Why did chance event C occur?” then according to Lewis we can provide a plain story that tells us why it occurred. He writes,

Take away the “rather”... and explanation [of chance events] becomes possible. Even a chance event has a causal history. There is information about that causal history to be provided in answer to a plain why-question. And thus we are right to proceed as we do in explaining what we take to be chance events. (p. 231)

Despite its creativity, this compromise is hard to sustain. Suppose a photon is fired towards a sheet of thin gold leaf, and that there is a 90% chance that the photon will bounce back and a 10% chance that it will tunnel through. The photon bounces back, let’s say, and I—knowing that there was a 10% chance that it would tunnel through—ask you why it bounced back. You, in keeping with Lewis’s recommendation, start describing the causal history of the event: that the photon was fired from the photon gun, that it flew towards gold leaf, and so on.

---

32 Taking chances here to be objective—part of the furniture of the world.
Then you stop. Or perhaps you describe the flight of the particle right up to the very last moment and then say something to the effect that: “The photon at this point had a 90% chance of bouncing back and a 10% chance of tunneling through and, by chance, it bounced back.”33

A story of this kind would not strike me satisfying at all, and I suspect the same is true of you. More exactly, a story of this kind would fail to satisfy my original question: “Why did the photon bounce back?” When I am aware that there is a 10% chance that it might have tunneled through, what I want—as far as I can tell, more or less irresistibly—is something that accounts for the difference between its bouncing back and its tunneling through. What is misleading about Lewis’s distinction between plain versus contrastive why-questions, therefore, is his suggestion that a “rather than” (i.e., contrastive) question of this kind is optional, and it is misleading because it suggests that there is some measure of understanding to be gained from learning information about the plain history of the event that fails to tell us why one of these alternatives obtained rather than the other. It is not just that “plain histories” of the sort described by Lewis give us less of what we were looking for when we asked the question; they don’t seem to give us what we want at all.

33 Peter Railton (1981) applies a causal story along these lines to the following chancy example: “Consider the following example. Suppose that we are applying a herbicide to a patch of healthy milkweed, and suppose that a dose of this herbicide alters the biochemical state of milkweeds from a normal, healthy state S, in which plants have probability .9999 of surviving 24 hours, to a state S’, in which there is but probability .05 of lasting that long. When we return to the milkweed patch 24 hours after spraying and find to our consternation that a particular plant which received a full dose of herbicide is still standing, how are we to explain this? Presumably, we should point out that the plant was in state S’ after the spraying, that in this state it had probability .05 of surviving 24 hours, and that, by chance, it did. The spraying, then, is part of the explanation of survival even though it lowered the probability of survival for this plant from what it would otherwise have been” (p. 240). Railton, somewhat mystifyingly, takes the story to be explanatory.
One way to emphasize what is wrong with this “plain history” approach is by connecting it with the point about Sam’s slurring of his words from the previous section. Suppose that I ask you why Sam is slurring his words. You then include as part of your explanatory story the fact that the earth cooled to a certain temperature X billion years ago, that his parents accidentally met on the train, and so on. Here again, it is difficult to see how this information helps to account for his current slurriness at all, even a tiny bit. Why? Because intuitively whether the earth cooled to a certain temperature rather than not does not make a difference to his slurring—more exactly, it does not help to account for why he is slurring his words rather than speaking them clearly.\textsuperscript{34} Of course, presumably it does make a difference to whether he exists or not, and without existing he could neither speak clearly nor slurrily. But why Sam exists rather than not is a different question, and not the one that I was asking.\textsuperscript{35}

\textbf{2.10 Summing Up}

Assuming we want to follow common sense and allow that learning about causes generates understanding, a basic conclusion to draw from this discussion is that information about causal histories is explanatory just to the extent that it reveals information about \textit{difference makers}.\textsuperscript{36} Although in some sense I acquire

\textsuperscript{34} Again, one could make up an exotic story here: perhaps, that he is drunk \textit{because} he is depressed about his existence. But let’s keep things simple.

\textsuperscript{35} Or perhaps a better way to put the point is to say that when we asked our question we were \textit{presupposing} that he exists, so learning that he exists will not be satisfying—it will not be what we were looking for when we posed a question with this presupposition—at all.

\textsuperscript{36} Alternatively, the information only \textit{belongs} in the causal history it if reveals a difference maker. As Jonathan Schaffer (2004) argues: “So what is causation?... The moral I want to draw is that causation involves at least some aspect of \textit{difference making}. In both positive and negative
accurate information about the causal history of the fire (say) when I learn that it
was written about on p. 17 of the police report, I do not learn difference making
information: that the fire was written about on p. 17 or 18 or 19 or what have you
does not seem to make a difference to whether or not the fire occurred.

Provisionally, then, the epistemic gain we experience when we understand
seems to have something to do with grasping difference-making information. In
the following chapters I will try to develop this idea.

causations, whether or not the cause occurs makes the difference as to whether or not the effect
will occur. For instance, the pulling of the trigger makes the difference as to whether or not the
gun will fire, and the absence of the flow of blood makes the difference as to whether or not the
victim will live. Misconnection is not causation, because it makes no difference as to whether or
not the effect will occur. For instance, Red’s spraying [of the rock that is about to shatter the
window] makes no difference as to whether or not the window will shatter, and Tim’s gaping [at
the rock in mid-flight] makes no difference there either” (p. 214).
CHAPTER 3
THE NEED FOR EXPLANATION

In this chapter I will try to set the stage for what follows by addressing two main questions. First, what is it in virtue of which a situation stands in need of explanation for someone? If your refrigerator breaks down, your non-functioning, room-temperature refrigerator will be something that stands in need of explanation for you. But why? In virtue of what? Second, given that a situation stands in need of explanation for you, what is involved in the satisfaction of that need? In other words, what is it that you grasp that takes you from having a need for explanation to having satisfied that need?

In a sense that will become clearer as we go, one way to think of this chapter is as an attempt to make progress on both of these questions, but especially the first, by elaborating an insight proposed by Robert Nozick (1981) in his *Philosophical Explanations* some twenty-five years ago. According to Nozick,

I am tempted to say that explanation locates something in actuality, showing its actual connections with other things, while understanding

---

1 Following Peter Menzies (1989, 2004), I mean the notion of a “situation” to be broad enough to encompass both states of affairs and events. As Menzies puts it: the notion of a situation is “meant to include anything that common sense dignifies as causes and effects—events, states of affairs, absences, omission, and other nonoccurrences” (2004, p. 151). If a state of affairs consists of a substance or system bearing a property (Uncle Albert’s having lung cancer), and an event consists in a substance undergoing a change in properties (Uncle Albert’s dying), then I intend a situation to be broad enough to include both of these categories, for there is no doubt that we are often interested in explaining instances of both. It is worth pointing out that on this model situations are quite fine grained. The thing that explains Uncle Albert’s dying for me—in other words, why he died rather than lived—will not necessarily be the thing that explains, say, why he died in his sleep rather than awake.
locates it in a network of possibility, showing the connections it would have to other nonactual things or processes. (Explanation increases understanding too, since the actual connections it exhibits are also possible.) Recall how illuminating it can be to place something, something actual even, in a typology or a two-by-two matrix, how salient is the insight gained through locating it in that network of alternative possibilities. (p. 12)

Building on the connection that Nozick draws between the search for understanding and one’s sense of possibility, in response to the first question what I will argue is that (roughly) a situation stands in need of explanation for someone in virtue of the person’s sense that there are various ways the situation might have been.² The answer to the second question then naturally suggests itself: satisfying the need requires grasping what the difference between these alternatives depends on.

In addition, I will argue that what Nozick characterizes as “a network of possibility” is in fact essential to the project of understanding. In order to defend this last claim properly, more will need to be said concerning how to think about the network of possibility, and I will try to address this issue in Section 3.3.

3.1 The need for explanation

Our first goal is to get a better sense of what lies behind the need for explanation in the first place. Given your background beliefs, some things seem to stand in need of explanation and others do not. So what accounts for the difference? This question is crucial because if we can get a good grip on why a situation stands in need of explanation, we will then be in a better position to

² Brain Skyrms (1980) makes a similar proposal: “Let \( \{S_i\} \) be an exhaustive set of mutually exclusive possible states of the system of which \( S^* \) is a member…. Degrees of explanatory power are relevant to the force with which the explanation resolves the puzzlement: why \( S^* \) rather than something else?” (pp. 140-41).
determine how we might try to satisfy that need. Moreover, on the assumption that explanation is oriented towards understanding (an assumption we will look at more closely in Chapter 4), there is a natural sense in which figuring out what it takes to satisfy the need for explanation and figuring out what it takes to understand go hand in hand.

Our question—“In virtue of what does a situation stand in need of explanation?”—is complicated because there are two quite different senses in which a situation might stand in need (or fail to stand in need) of explanation for someone, one pragmatic and the other epistemic.

From a pragmatic point of view, whether or not a situation stands in need of explanation for you will depend on whether you deem it worth the bother, in some all-things-considered sense. In this sense, my leaky roof will stand in need of explanation for me but not for you in virtue of the fact that although the well-being of my roof is prudentially valuable to me, this is (let’s suppose) far from true for you.

There is another sense in which situations stand in need of explanation, however, which considers things from a purely epistemic or intellectual point of view. As Carl Hempel (1965) notes in the opening paragraph of his “Aspects of Scientific Explanation,” in addition to the obvious practical incentives we have for wanting to explain our environment, we also have a more purely intellectual desire. As Hempel puts it, it is a desire rooted in man’s

*sheer intellectual curiosity*, in his deep and persistent desire to know and to understand himself and his world. So strong, indeed, is this urge that in
the absence of more reliable knowledge, myths are often invoked to fill the gap” (p. 333; my emphasis).3

Taking a cue from Hempel, we might therefore say that the sign that a situation stands in need of explanation, \textit{from a purely epistemic point of view}, is that it is capable of eliciting our curiosity. To say that a situation’s ability to elicit our curiosity is a “sign” that it stands in need of explanation for us, of course, does not yet tell us what it is, in virtue of which, a situation stands in need of explanation. What it does do, however, is offer a new and potentially more tractable way of posing our original question: Namely, in virtue of what do certain situations elicit our curiosity and others not? Although in what follows I will for the most part continue to pose our question in its original “stand-in-need-of-explanation” form, in order to emphasize the fact that this is an epistemic question we are asking, I will from time to time use the alternative “elicit-our-curiosity” form. It will, at any rate, be important to keep this curiosity-driven need for explanation at the back of our minds throughout.

As an illustration of what is at issue here, Elliot Sober (1986) offers the following example of the kind of situation that does \textit{not} elicit our curiosity. As he writes, the question

- Why is Kodaly a Hungarian rather than a vegetarian?

fails to grab us, even if we accept that Kodaly is both a Hungarian and a vegetarian, simply because there seems to be nothing to be curious \textit{about}

---

3 More recently, Roger White (2005) likewise connects our interest in why-questions with our natural curiosity: “In asking a why-question we are seeking to satisfy a peculiar kind of curiosity; we are seeking \textit{understanding} and trying to \textit{make sense of things}” (p. 2).
Kodaly’s being both a Hungarian and a vegetarian (p. 145). Why can’t he be both? Why, in order to explain the one, do we have to rule out the other?

But if not just any situation stands in need of explanation for us from a purely intellectual point of view, then what is it about those that do? To the extent that philosophers of science have addressed this question, two answers have been particularly prominent.

The first suggestion, usually associated with P. W. Bridgman (1968), is that a situation stands in need of explanation for someone in virtue of the fact that, relative to the person’s background beliefs, it is unfamiliar. The behavior of gases when excited, for instance, or the way electricity passes through a copper wire might be unfamiliar to someone, and because unfamiliar, in need of explanation.

The second suggestion is that a situation stands in need of explanation for someone in virtue of the fact that it is surprising or unexpected. As C. S. Pierce ([1908] 1968) argued:

4 It might be said that although relative to our usual beliefs the question “Why is Kodaly a Hungarian rather than a vegetarian?” is a poor one, relative to a different set of beliefs the question might seem entirely appropriate—for instance, relative to a set of beliefs according to which Kodaly was offered a choice at some point between being a Hungarian and being a vegetarian. But in a way this just makes our original question more pressing: What is it about someone’s epistemic background in virtue of which some explanatory requests have purchase and others do not?

5 Thus P. W. Bridgman (1968) claims that, “I believe that examination will show that the essence of explanation consists in reducing a situation to elements with which we are so familiar that we accept them as a matter of course, so our curiosity rests” (p. 37). Ernst Mach ([1911] 1968) applies this suggestion to understanding in particular: “Understanding consists in analysis alone; and people usually reduce uncommon unintelligibles to common ones” (p. 32). More recently, Peter Godfrey-Smith (2003) has made a similar suggestion: “Indeed, it seems clear to me that what produces an ‘Aha!’ reaction is often the realization that some odd-looking phenomenon is really a case of something more general” (p. 196).

6 Likewise, Hempel (1965) writes, “Explanation-seeking questions of the standard type ‘Why is it the case that \( p \)?’ are often, though by no means invariably, prompted by the belief that \( p \).
Every inquiry whatsoever takes its rise in the observation... of some surprising phenomenon, some experience which either disappoints an expectation, or breaks in upon some habit of expectation of the inquisiturus; and each apparent exception to this rule only confirms it. (p. 143)

There are problems with both of these suggestions, however. For one thing, it is not necessary for a situation to be unfamiliar or unexpected in order for it to stand in need of explanation. Indeed, many situations that stand in need of explanation are the very opposite of unfamiliar and unexpected. The squeaking wheel on my bicycle, for example, might be painfully familiar to me as well as entirely expected, yet for all that it might still be something that stands in need of explanation for me.7

Moreover, these accounts imply that the need for explanation will be satisfied when by our lights a situation comes to seem familiar or expected. But it is hard to see how this will give us what we are looking for. If we want to understand why a gas expands at rate \( r \) when heated rather than at some other rate \( r^* \), it is not immediately obvious how thinking of the gas on the model of little billiard balls—in other words, how thinking of the gas along lines that make it seem in a certain sense more familiar or picturable—will help me to understand

---

7 This is a modified version of an example used by Peter Lipton (2004a, p. 26). With respect to the familiarity proposal in particular, one might object that the squeaking wheel on my bicycle is not familiar in the requisite sense. The difficulty with assessing this objection stems from the vagueness of the familiarity proposal in general; it's just not clear what it takes for a situation to count as familiar or unfamiliar, or what—from an epistemic point of view—it would take to regard a situation as familiar.
this question. Moreover, as we learned from Chapter 2, being in a position to expect a situation is not the same as understanding it. After learning how to interpret the barometer outside my window, I will be in a position to expect the onset of the tomorrow’s storm, but this expectation is compatible with a failure on my part to understand why the storm is due to occur.

Part of the reason why familiarity and expectability theories have appealed to people over the years, I suggest, is because of a failure to distinguish between the epistemic and pragmatic aspects of the need for explanation. Consider the expectability account again. It is plausible to think that there is a good pragmatic reason why, among all the things that are capable of eliciting our curiosity from an epistemic point of view, unexpected things should stand out. After all, it is decidedly not in our pragmatic interests to have the world constantly unfold before us in surprising ways—and this will inspire us to want to get to the root of the surprisingness. What’s more, if our interest in forecasting tomorrow’s weather is pragmatic, then our rational expectation that a storm will pass by, based on our reading of the barometer, will serve us just fine. But none of this seems to get to the heart of the epistemic need for explanation, and so it should not be surprising that none of this seems close to specifying how that need might be satisfied.

If these suggestions are inadequate, however, then in virtue of what does a situation stand in need of explanation, from a purely epistemic point of view?

---

8 As I suggested in the previous footnote, this is more a complaint about the obscurity of the familiarity proposal than an outright objection.

9 For further critical discussion see Friedman (1974, pp. 9-10) and Lipton (2004a, pp. 24-25).
Taking another look at the squeaking wheel of my bicycle suggests the outline of an answer. As a first pass, the reason why the squeak seems to stand in need of explanation for me is because of my awareness that there are other ways my wheel might have been—namely, non-squeaky. In other words, it is because I believe that the wheel is capable of being non-squeaky that there is something that elicits my curiosity about it squeakiness. Indeed, if, for whatever reason, I believed that the wheel was not capable of being non-squeaky, then it is plausible to think that the situation would not elicit my curiosity at all. From my point of view, there would simply be nothing about the squeakiness of the wheel to be curious about.

In his recent *Epistemic Justification*, William Alston (2005) nicely elaborates the basic idea:

> Whenever any claim is made about something other than an explanation, one can seek an explanation of its being so rather than otherwise. We can ask, “Why does this object have this property [rather than some other]?,” “Why did this happen where and when it did [rather than elsewhere and at another time]?,” “Why does this proposition entail that proposition [rather than some other]?,” “Why does this substance weigh more than that one?,” and so on. So explanation is one sort of thing we can seek with respect to any subject matter whatever. (p. 165)

One might worry that, if we take this view seriously, far too many things will count as standing in need of explanation for us. But no one is saying that the

---

10 But doesn’t that mean that its inflatedness will call for explanation because it is capable of being uninflated, that its blackness will call explanation because it is capable of being brown, and so on? On the one hand, Yes, and on the other hand, Maybe. Yes in the epistemic sense: if we think that the wheel is capable of these alternative states, then by our lights there is something there to be explained (and hence understood) about its actual state. But Maybe in the pragmatic sense: it may be that you have already satisfied the call, or it my be that you lack an all-things-considered interest in satisfying the call.

11 In adding the brackets, I have tried to draw out the contrastive (“why this rather than that?”) structure to these questions that Alston calls attention to in the first sentence.
need for explanation brings with it any kind of intellectual requirement. It is up to individuals to decide how many of these questions they want to pursue, and which are most important. The claim at issue is much more general: simply that, from an epistemic point of view, it is in virtue of our sense that things could have been otherwise that situations elicit our curiosity in the first place.

In the next section I will try to take this suggestion a step farther, in two respects. First, I will consider two examples which should help to further illustrate the way in which the need for explanation is tied to our sense of possibility. Second, I will consider what these examples have to teach us about how one might go about trying to satisfy that need.

3.2 Greene and Tye

The first example comes from the popular literature on string theory and the second from recent debates in the philosophy of mind. It is worth noting at the outset that I offer these examples as merely representative, rather than as special paradigms or ideals.

Consider first then Brian Greene’s (1999) defense of string theory as a potential “theory of everything.” As Greene notes, throughout the second half of the 20th century researchers seemed to uncover new elementary particles on a daily basis: no longer just electrons and protons, but suddenly quarks, muons, neutrinos, taus, and so on. What amazed particle physicists was not simply the fact that these particles existed, however, but that their properties were so oddly diverse. As Greene writes, the questions quickly piled up.

Why are there so many fundamental particles, especially when it seems that the great majority of things in the world around us need only
electrons, up-quarks, and down-quarks? Why are there three families? Why not one family or two families or any other number? Why do the particles have a seemingly random spread of masses—why, for instance, does the tau weigh about 3,520 times as much as an electron? Why does the top quark weigh about 40,200 times as much as the up quark? These are strange, seemingly random numbers. Did they occur by chance, by some divine choice, or is there a comprehensible scientific explanation for these fundamental features of the universe? (pp. 9-10)

According to Greene, there is thus something about these particles, over and above the mere fact that they exist, that eludes understanding and stands in need of explanation. But what is it, exactly? What is so puzzling or hard to grasp about the fact that the tau particle (say) weighs 3,520 times as much as the electron?

Despite what Greene himself suggests, the problem does not seem to be the “seeming randomness” of the properties. Suppose that the mass ratios among the elementary particles were much simpler (to our minds at least) than the actual values Greene cites: for example, suppose the particles stood in mass ratios of 1:2 or 1:3 rather than in ratios of 1:3,520 or 1:40,200. Although these alternative ratios seem (again, to our minds) much less random than the ratios that actually hold, it should be clear that these too can be considered brute and inexplicable in their own way. Suppose the tau weighed three times the mass of the electron—that is, that the two stood in a 1:3 ratio. The fact that this simple ratio held rather than some other simple ratio would still presumably be something that stood in need of explanation. For that matter, the fact that this simple ratio rather than some other, more complex ratio held would likewise still seem to stand in need of explanation.
If Greene’s argument succeeds in convincing us that there is something that stands in need of explanation about the properties of these fundamental particles, the problem must therefore be more basic than that the properties bear certain less-than-simple relationships to one another. Instead, the real worry seems to be this: The reason why the mass of the tau stands in need of explanation for us is because of our awareness that other things of this kind—specifically, other elementary particles—have different masses, so in order to understand why the tau has the particular mass it does rather than another mass that—as an instance of this kind—it might have had, we need to know what it is that differentiates the tau from other elementary particles. More briefly, we want to know the difference between the tau’s mass and the mass of the other elementary particles depends on.

Consider a comparison. If you believe that the corn stalks in a certain field can range in mature height between 6 feet and 8 feet, then presumably the differences among these heights will be something that is capable of eliciting your curiosity and therefore of standing in need of explanation for you, from a purely intellectual point of view. In order to satisfy this need for explanation what you will therefore need to do is identify what the difference between a given stalk’s being 6 feet and its being 7 feet or 8 feet high depends on. Perhaps what makes the difference is that the fertilizer was distributed unevenly throughout the field, or perhaps pests inhibited the growth of some stalks but not others. Whatever it might be, to satisfy your curiosity you will need to identify the factor (or factors) that make a difference here.
I will have more to say (though not much more) about the significant notion of a “kind” in the following section, but for now the important thing to notice is that as one of its prime virtues Greene claims that string theory is able to point to such a difference maker. In particular, he argues that elementary particles have the particular masses and charges they happen to have because they are constituted by vanishingly tiny strings whose resonance patterns determine why the particles have masses and charges with certain values rather than others. As he puts it: “Here’s the central fact: Just as the different vibrational patterns of a violin string give rise to different musical notes, the different vibrational patterns of a fundamental string give rise to different masses and force charges” (p. 143). He continues:

So we see that, according to string theory, the observed properties of each elementary particle arise because its internal string undergoes a particular resonant vibrational pattern. This perspective differs sharply from that espoused by physicists before the discovery of string theory; in the earlier perspective the difference among the fundamental particles were explained by saying that, in effect, each particle species was ‘cut from a different fabric.’ Although each particle was viewed as elementary, the kind of ‘stuff’ each embodied was thought to be different. Electron ‘stuff,’ for example, had negative electric charge, while neutrino ‘stuff’ had no electric charge. String theory alters this picture radically by declaring that the ‘stuff’ of all matter and all forces is the same. Each elementary particle is composed of a single string—that is, each particle is a single string—and all strings are absolutely identical. Differences between the particles arise because their respective strings undergo different resonant vibrational patterns. What appear to be different elementary particles are actually different ‘notes’ on a fundamental string. (pp. 145-46)

One might reasonably ask, in passing, how this qualifies as a step towards a theory of everything. Even if string theory is correct and all matter is made of the same stuff—tiny oscillating strings—then there still seems to be plenty of room to wonder why strings of a certain sort, the ones that constitute tau particles for
instance, vibrate according to pattern X while others vibrate according to pattern Y. But the important point here is that if Greene is right that the difference between the masses and charges of the elementary particles depends on the varying oscillation patterns of the strings that constitute the particles, then he has at least helped us to understand this much: namely, why the elementary particles have these mass and charge values rather than the masses and charges that, in virtue of the kind of thing they are, they in some sense might have had.

Consider next the example from philosophy: Michael Tye’s (2002) case against qualia in the philosophy of mind. Qualia are typically thought to be the properties of conscious experience that characterize what it is like to have the experience: e.g., what it is like to be visually conscious of the redness of a red balloon, or tactually conscious of the balloon’s smoothness. Considered in one way, Tye argues, it is undeniable that there are such phenomenal aspects of conscious experience. Of course there is something that it is distinctively like to experience the world; and hence qualia, in a certain respectable sense, undoubtedly exist. What Tye denies, however, is that qualia can be thought of as irreducible or brute nonrepresentational aspects of conscious experience. Tye calls such allegedly brute aspects of experience “Qualia” with a capital “Q” and insists that, understood in this way, there are no Qualia.¹² They are a philosophical fiction. On his view:

Any adequate account of the subjective or phenomenal aspects of our visual states ought to yield an understanding of why those states have those aspects. Why, for example, does having a visual experience of blue ‘feel’ the way it does and not some other way? It is hard to see how any

¹² Below I will follow Tye’s lead in thus distinguishing between “qualia” and “Qualia.”
satisfying answer can be given to this question if the phenomenal aspects of such experiences derive from visual Qualia.

Suppose, for example, that there are visual Qualia and that such Qualia are nonphysical and irreducible. Then the ‘felt’ aspect of the visual experience of blue is a matter of its having a special, nonphysical property. It is the presence of this property that gives the visual experience its distinctive ‘feel.’ Does this really offer us any enlightenment? Apart from the usual concerns about the emergence and causal role of such properties we may still wonder why the visual experience that has the content blue is associated with this irreducible felt quality rather than some other—why, for example, it does not have the felt experiences that represent red. This surely is an impenetrable mystery. (p. 455)

Here again, since our concern here is not with the status of Qualia per se (the capital “Q,” controversial kind), what we should be asking is: Why is this a good reason to worry about Qualia, if it is a good reason to worry? What is it about Tye’s complaint that registers with us, and makes us think that there is something here that stands in need of explanation?

To the extent that Tye succeeds in making Qualia seem in need of explanation, I suggest that it is because he reminds that we can always ask, of any particular Qualia—blue Qualia, for example—why it is associated with one felt experience rather than another. After all, according to Qualia advocates other things of this kind—red Qualia and green Qualia, for example—are said to be associated with other felt experiences. So what accounts for the difference between these things? Why, given the various ways Qualia are apparently capable of being in virtue of the kind of thing they are, are they one way rather than the other?

Naturally, Tye claims that his view has the advantage in terms of being able to answer a question along these lines: by his lights, the thing that makes the difference to why a felt experience is one way rather than another is its
representational content. The reason why you enjoy an experience of red on some occasion rather than an experience of green, for instance, is because the experience is an experience of something red rather than an experience of something green. Tye can therefore point to something that plausibly accounts for why your felt experiences are one way rather another because he can point to a \textit{difference maker}—the representational contents of these states—whereas Qualia advocates can point to nothing that makes a difference to the various ways felt experience seems capable of being.

Again, it is worth thinking about these passages from Greene and Tye for two reasons. First, to see in the concrete how the call for explanation works, which helps to clarify what it is in virtue of which a situation stands in need of explanation. Second, in order to clarify what it takes to satisfy that need.

In both cases, the pattern is remarkably similar. The reason why the mass of the tau seemed to stand in need of explanation for us was because of our sense that other things of this kind were capable of having a different mass, and thus we wanted to know what it was that this difference depended on. Likewise, the reason why the felt experience we associate with red Qualia stood in need of explanation was because of our sense that other things of this kind were capable of having a different phenomenal character, and we wanted to know what it was that \textit{this} difference depended on.\textsuperscript{13}

\textsuperscript{13} Or rather, if we want to remain neutral about whether or not there are Qualia, we should put it this way: it is because of our conviction that other felt experiences have a difference phenomenal character that we want to know why this instance of the kind—a red experience, say—is one way rather than another.
In both cases the need for explanation arose along with our belief that there were different ways that something might have been, given the kind of thing it was. It is now time to try to fill in the details of this account by further elaborating the connection between kinds, capacities, and the network of possibility.

3.3 Proto-Understanding

Although I have no grand theory of kinds to offer, as I am understanding the notion of a kind, to be an instance of a kind involves having a certain capacity, a capacity something has in virtue of being an instance of the kind.\(^{14}\)

To make better sense of the connections I want to draw between kinds, capacities, and possibility, it will help to introduce a new piece of terminology: the notion of proto-understanding. By an agent’s proto-understanding, I mean an agent’s convictions about kinds and capacities.\(^ {15}\) The notion of proto-understanding can therefore be thought of as an elaboration of Nozick’s notion of a network of possibility.

We can try to clarify the notion of proto-understanding by means of the following example. Suppose that two observers watch a brick collide with a

\(^{14}\) This proposal therefore bears strong similarities to Nancy Cartwright’s (1989, 1999) view that our attempt to explain the natural world is grounded in our knowledge of natures and their capacities. As she writes: “Our basic knowledge—knowledge of capacities—is typically about natures and what they produce” (1999, p. 80)

\(^{15}\) Although he argues for the point in a different way, Michael Scriven (1962) makes a similar proposal: “[T]he request for an explanation presupposes that something is understood, and a complete answer is one that relates the object of inquiry to the realm of understanding in some comprehensible and appropriate way. What this way is varies from subject matter to subject matter...; but the logical function of explanation... is the same in each field” (p. 202). As we’ll see, although strictly speaking only individuals possess proto-understanding, we can also speak in an analogical way of the proto-understanding of communities, and this of many sorts: scientific communities, primitive communities, etc.
window, shattering it. The first observer, Stan, has fairly conventional 21st
century beliefs about the world, while the second observer, Harry, believes (as a
result of being raised by a community of delusional J. K. Rowling fans) that the
world is full of objects with magical powers and capacities. Of both, moreover,
suppose we ask a range of *prima facie* unusual questions. For instance: “*Why
did the window shatter rather than turn into a frog?*” And: “*Why did the
window shatter rather than disappear in a puff of smoke?*”

Let’s focus now on how Stan would answer these questions, specifically,
the “*Why did the window shatter rather than turn into a frog?*” question.
Presumably his answer will *not* be: “because it was struck by a brick.” Rather, his
answer will be something like: “because windows can’t turn into frogs. The world
doesn’t work that way.” In other words, rather than answering the question Stan
will reject one of the question’s presuppositions: that the window is capable of
turning into a frog.

What’s more, notice that in making this move Stan has made no direct
progress on figuring out why the window shattered; in pointing out that it was
not possible for the window to turn into a frog, it is not as if he took a step closer
towards identifying what made the difference between the window’s shattering
and remaining intact—alternatively, what the difference between the two
alternatives depended on. Instead, what he has done is mark out a boundary
within which the difference making question can even be asked. If a difference
maker is supposed to tell us why one possibility obtained rather than another,
then it would seem that in order for the notion of a difference maker even to have
application we need to have some sense of which possibilities are relevant and which are not.

Harry by contrast might answer the question in a number of ways, depending on the particulars of his beliefs about the window and its capacities. For instance, he might reply that because the window is capable of turning into any number of things (including a frog), what happens after the window is struck by the brick depends on the enchantment that has been placed on the brick. In particular, suppose he believes (based on his community’s testimony) that if the brick has been sprinkled with newt’s oil, the window will shatter rather than turn into a frog. Suppose too that in this particular case he believes the brick has been sprinkled with newt’s oil. If so, he will identify the sprinkling with newt’s oil as the thing that makes the difference between the shattering and the turning into a frog, and he will thereby take himself to understand why the window shattered rather than turned into a frog on this basis.

The point that bears emphasizing, in any case, is whether someone finds a particular explanation-seeking why-question worth asking or not will depend in the first instance on the person’s sense of which alternatives are relevant and which are not. If it is possible for the window to shatter or to turn into a frog upon being struck by a brick, then there is reason to be curious about why it shattered rather than turned into a frog. If it isn’t possible, however, then the reason for one’s curiosity—the apparent feature of the situation in virtue of which it elicited one’s curiosity—fades away.

There are thus always two ways to approach a “Why this rather than that?” question: either (a) by claiming that at least one of the alternatives is impossible
(i.e., by rejecting one of the question’s presuppositions), or (b) by trying to identify what the difference between these alternatives depends on. But whether one answers with response (a) or response (b) will plainly depend on one’s sense of what a given thing is capable of, given the kind of thing it is.

Either way of responding to the question therefore seems to presuppose a fairly rich sense of possibility; in particular, it seems to presuppose a sense of what things are capable of, given the kinds of things they are. Of course, as we’ve seen, many of these convictions about what things are capable of might be poorly founded: in virtue of being brought up in a community of would-be wizards, one might believe that many things are possible which in fact are not. But it is one thing to say that these convictions might be ill-supported or loony or controversial and quite another to say that they are not central to someone’s sense of whether a particular situation stands in need of explanation or not.

3.4 A Worry

It might be objected that claiming that a situation stands in need of explanation for someone in large part in virtue of their convictions about kinds and their capacities (for short, in virtue of their proto-understanding) will lead to an implausibly individualistic account of explanation, and hence of understanding. If my proto-understanding differs from yours (and how could it not, at least in the details?), then the sort of situation that stands in need of explanation for you will not for me, and interpersonal attempts to explain the world will be seriously compromised.
But while this view does imply that the need for explanation is relative to background beliefs, nothing about this view requires an implausibly individualistic account. So long as one of our primary sources of knowledge continues to be the testimony of others, then that testimony will surely inform our beliefs about kinds and capacities, and in normal circumstances it will lead to a kind of shared sense of what is possible and what is not. The charge that proto-understanding is excessively individualistic therefore hits home only if the sources of our knowledge are excessively individualistic. But there is no reason to think that. Moreover, provided they have roughly similar beliefs about the powers and capacities of things, we can therefore properly speak of the members of a community as a whole—including, say, the members of a *scientific* community as a whole—understanding or failing to understand some situation.\(^\text{16}\)

For creatures like us at least, proto-understanding seems to be the stuff that understanding is made of.

\(^{16}\) It is also worth briefly mentioning research that suggests a more hard-wired source of shared proto-understanding. Baillargeon, Kotovsky, and Needham (1995) report that when 4-month-old infants are shown a released block that remains suspended in the air without any visible means of support (in fact, it is supported from behind in a way invisible to the children), the infants look reliably longer at the suspended block than at blocks which fall to the ground when released. What this suggests is that from very early on, possibly from the very beginning, we find ourselves in the world with proto-understandings that are shaped by a sense of kinds and their capacities. Thus heavy looking objects are the kinds of things that should fall when dropped; or, perhaps better, are the kinds of things that are not capable of rising when dropped. When they remain suspended in midair, our sense of how such things should behave is upended.
CHAPTER 4
THE GOAL OF EXPLANATION

What is the goal of explanation? According to one long tradition, understanding is the goal of explanation. Explaining things, or trying to explain them, is a kind of activity we engage in—as scientists, as engineers, as car mechanics, and more generally as people puzzled about why things are one way rather than another—and the goal of the activity is understanding.

Over the last several decades this claim has come under sustained fire. According to philosophers such as Mario Bunge, David Lewis, W. H. Newton-Smith, and J. D. Trout, understanding is not qualified to be the goal of explanation because understanding is fundamentally a psychological, rather than an epistemic, state.

---

1 According to Julius Moravcsik (1979, 1990) and Myles Burnyeat (1980), in the ancient world it was more or less taken for granted that explanation aims at understanding. For representative contemporary advocates, see Peter Lipton (2004b, p. 1) and Roger White (2005, p. 2).


3 Alternatively, sometimes the charge is that the notion of understanding is too individualistic to serve as the goal of explanation—at least, the goal of scientific explanation. As Paul Humphreys (1989) writes, an approach that focuses on the significance of understanding, “immediately leads to a relativization of explanations to an individual” (p. 127). He therefore suggests that it is best “to set aside that whole issue of what constitutes or promotes understanding and focus only on what is required of an explanatory candidate that may, or may not, lead to such understanding” (pp. 127-28).
What this charge amounts to is not always entirely clear, but the basic idea seems to be that understanding is not truth-connected in the appropriate way. Misguided people of all stripes, it is said, from CIA conspiracy theorists to alchemists to astrologers, seem to enjoy a robust form of understanding—they make connections among their beliefs, they are confident that they “see” why things are one way rather than another, and so on—so whether someone enjoys understanding or not must be unrelated to how things stand in the world.

Moreover (the objection continues), since understanding on this view is not responsible to the truth, then it must be the result of other, more purely psychological factors. Perhaps, as David Lewis (1986) suggests, understanding results from a certain sense that an explanation “seems familiar” or “illuminating,” or that it “looks simpler” (p. 228). But, according to Lewis, regarding an explanation as familiar or simple or the like cannot be the goal of explanation because the goal of explanation is—if nothing else—the truth, and these psychological responses can easily be had in the absence of truth.

Let’s call this the “psychological objection.” In this chapter I will evaluate this objection, as well as the overall case for claiming that understanding is the goal of explanation, by first trying to clarify what it is that we want in the goal. In other words, I will try to clarify what it is that we are aiming for when we engage in the activity of explanation. Once we have a better idea of what it is that we are aiming for, in the remainder of the chapter I will then ask how well the goal thus described corresponds to understanding.

---

4 That it has psychological components? If so, then knowledge would not count as an epistemic state either.
In addition to addressing the worry about truth, I will suggest that part of
the reason why understanding has been dismissed as the goal of explanation is
because the psychological dimension of the goal of explanation has itself been
almost entirely neglected. In turn, the psychological dimension of
understanding—the Aha! experience, the sense that a certain explanation “feels
right,” etc.—has been conspicuously overemphasized. One of my goals in this
chapter will therefore be to try to correct for both of these exaggerations. Just as
the goal of explanation includes a richer psychological—including,
phomenological—dimension than is generally acknowledged, so too, I will
claim, understanding has a far stronger truth connection than is generally
acknowledged.

4.1 Senses of Explanation

To begin with we need to be more precise about what the claim that
understanding is the goal of explanation amounts to. One notorious problem
here is that the noun “explanation” is process/product ambiguous, and this in
more than one sense.5

For instance, if you were to claim that Newton’s explanation of the tides is
correct, you would be making a claim about a certain product: namely, the set of
claims or propositions that constitute Newton’s theory. If, however, you were to
claim that Newton’s explanation of the tides is long-winded and tedious, the

5 See, for example, van Fraassen (1980, ch. 5), Achinstein (1983, ch. 1). Several words in
our language share this ambiguity. “Science,” for example, sometimes refers to the act of doing
science, sometimes to the results. Kelly and Glymour (2004) nicely illustrate the active sense of
“science” in the following passage: “Science has many aims, but its most characteristic aim is to
find true answers to one’s questions about nature” (p. 96).
noun “explanation” here would be picking out Newton’s act or performance of explaining his theory, not the theory itself. In order to avoid confusing these two senses, in what follows I will try to refer to an explanation in the first (product) sense as an explanatory story, and an explanation in the second (process) sense as the act of presenting an explanatory story.

The claim we want to consider—namely, that understanding is the goal of explanation—appeals to yet a third sense of explanation. With respect to our claim, the noun “explanation” again picks out a kind of activity, but it is not the activity of presenting an explanatory story; a story, as it were, that we already have safely tucked away. Instead, the idea is that “explanation” in this sense is an activity or process that we might loosely think of as the process of inquiry.

The process of inquiry begins with the why-questions that occur to us constantly throughout our lives. Why are the tides low today? Why did the soda lose its fizz overnight? Why is the flame from the bunson burner turning purple? Pursuing these questions, pursuing these inquiries, is a kind of goal-directed activity, and it is with reference to this goal-directed activity that our claim has its place. Alison Gopnik (1998) nicely appeals to the relevant sense of explanation in the following passage:

Explanation is a goal-directed human activity. It depends on what is relevant or important to the explainer, it satisfies a special-kind of explanation-seeking curiosity, it answers “why?” questions. (p. 108)

6 It is with respect to this sense of explanation that Sylvain Bromberger ([1962] 1992) writes: “An explanation may be something about which it makes sense to ask: How long did it take? Was it interrupted at any point? Who gave it? When? Where? What were the exact words used? For whose benefit was it given?” (p. 50).

7 See the following footnote for a qualification, however.
Steven Weinberg (in a passage we will return to again in Section 4.8) likewise writes that, “Scientific explanation is a mode of behavior that gives us pleasure, like love or art” (1994, p. 24). As Weinberg suggests, explanation (and not just scientific explanation) is therefore something we do, and it is something we do with a goal in mind.

To ask “What is the goal of explanation?” in this sense is therefore to ask “What is it that we are looking for when we pose our explanation seeking why-questions?” Perhaps better, it is to ask: “What is it that we are looking for when we engage in the activity of inquiry?” We ask our why questions, we engage in inquiry, with a goal or end in mind. So what is it?

We can put our general question in a schematic way as follows: What is \([G]\)? where

\[G = \text{The Goal of Explanation}.\]

### 4.2 Two Points About the Goal

Two things about \([G]\) seem clear from the outset. The first is that explanation is a goal-directed activity with a specific kind of goal, an *epistemic* goal. As cognitive beings—as epistemic beings—certain explanation-seeking why-questions occur to us, and identifying the answers to these questions is a uniquely fulfilling epistemic accomplishment. Second, when we pose our

---

8 It would be a mistake to try to rephrase our question from the start as a question about specifying the goal of inquiry, for the simple reason that not all of our inquiries are *explanatory* inquiries. When I inquire into the departure time for my flight, for example, or into the number of students I have in my class next semester, I am not looking for an explanation. What I’m looking for instead is much more easily described as accurate information.

9 As Jaegwon Kim ([1989] 1993) notes: “I take it that *explaining* is an epistemological activity, and that an *explanation*, in the sense of the “product” or “theoretical content” of such an activity, is something about which we can have various cognitive attitudes” (p. 255).
explanation-seeking why-questions we don’t just want any old answer. Rather, we want an answer that is correct.

If I am puzzled about why my sink is leaking, for instance, then what I want is to find out the truth with respect to why my sink is leaking. Suppose that my sink is leaking because the gasket connected to the “Hot” handle has a hairline fracture. If after poking around I fail to notice this and instead come to the conclusion that my sink is leaking because there is a hole in the drain pipe, then in some obvious sense I will not have gotten what I was looking for when I began my inquiry or when I posed my question. What I wanted was a correct answer to my question, and I missed it. I fell for a false answer instead.

We can try to capture this aspect of the goal in either of the following ways, where the property after the “=” sign represents a necessary condition on meeting the goal.

(1) $[G] = \text{truth}$;

or perhaps as

(2) $[G] = \text{a correct answer}$.

However we care to put this, the basic idea is that what we desire at the end of inquiry, at a minimum, is truth (accuracy, correctness). Incorrect answers, no matter how comforting or familiar they might seem, are not what we are after when we pose our why-questions.  

---

10 For convenience, we can suppose that this story is straightforwardly false: there is, in fact, no hole in the drain pipe.

11 Questions of self-deception, as when one really wants the comforting-but-false answer, complicate the picture but I will not pursue them here.
One further point deserves clarification before moving on. Sometimes we say—indeed, sometimes in what follows I too will say for convenience—that what we want is not just a correct answer to our why questions but the correct answer, as if only one answer could satisfy us. As Wesley Salmon (2001) points out, however, there is virtually always more than one correct answer to our why questions.\textsuperscript{12} To appeal to his own helpful example: If you are wondering why a helium balloon in the middle of an airplane cabin lurches forward rather than backwards upon takeoff, there is one correct answer that makes use only of the laws of classical physics\textsuperscript{13} and another that appeals only to a principle of the general theory of relativity.\textsuperscript{14} Either answer, moreover, would constitute a satisfactory answer to our why-question. To demand exclusiveness where none is needed would be a mistake.

\textbf{4.3 Trout’s Objection}

At a minimum, then, what we want is a correct answer. On some ways of telling the story, moreover, this is essentially \textit{all} that we want; or at least, all that we \textit{need}. To suppose in addition that it is important for the correct answer to “feel right” to us, or for it to give us a warm glow of some kind, is to confuse what

\textsuperscript{12} A point on which, as Salmon (2001) suggests, all of the major contributors to the contentious literature on explanation seem to agree: “As one who has looked at a great deal of the literature on the nature of explanation—scientific explanation, at any rate—I cannot think of any major contributor who has proposed that a given fact has a unique correct explanation or a unique best explanation” (p. 67).

\textsuperscript{13} As Salmon (2001) puts it: “When the airplane accelerated, the rear wall of the cabin would exert a force on the air that would create a pressure gradient in the cabin, and that the helium-filled balloon, possessing less inertia (mass) than an equal volume of air, would experience an unbalanced force that would push it forward” (p. 67).

\textsuperscript{14} Salmon (2001) again: “Since an acceleration is equivalent, by relativity theory, to a gravitational field, the balloon would rise toward the front of the cabin just as it rises in air under the gravitational field of Earth” (p. 67).
it is that we are after from an epistemic point of view with whatever it is we might be after from a psychological point of view. In other words, although from a psychological point of view it might be desirable, for whatever reason, to have an answer feel right to us, from an epistemic point of view such a feeling is essentially irrelevant.\textsuperscript{15}

Examples of the following sort—this one adapted from Trout (2002)—are meant to drive the point home. Suppose that we are at a local airport and that we are watching a pilot practice his take offs and landings. After a particularly nice take off, you ask me (no doubt a bit artificially): “Why can this plane fly?” Given that there is a correct answer to this question that in some way incorporates Bernoulli’s principle,\textsuperscript{16} and given that you want a correct answer to your why-question, it follows that in some sense you will want this information about Bernoulli’s principle.

But now we need to ask an unusual question: In what way do you want this information? More exactly, what kind of epistemic attitude do you want to bear towards this information? After all, we have agreed that the goal of explanation—[G]—is an epistemic attitude of some kind. But this still leaves us with several options. To mention just a few, and focusing on the airplane case, 

\textsuperscript{15} David Lewis (1986) puts this point: “It is as if someone thought that a good explanation of any current event had to be one that revealed the sinister doings of the CIA. When the CIA really does play a part in the causal history, we would do well to tell him about it: we thereby provide correct explanatory information about the part of the causal history that interests him most. But in case the CIA had nothing to do with it, we ought not to tell him that it did. Rather we ought to tell him that it didn’t. Telling him what he hopes to hear is not even a merit to be balanced off against the demerit of falsehood. In itself it has no merit at all” (p. 228).

\textsuperscript{16} That is, the principle that as the speed of a moving fluid increases (air being just one example of a moving fluid), the pressure within the fluid decreases. Or, more formally, \( p + qv^2/2 + qgy = \text{constant} \), where \( p \) is the pressure, \( q \) is the density, \( v \) the velocity and \( y \) the height in a gravitational field of strength \( g \), all measured at the same point.
we might say that with respect to this particular question the goal—call it \[G_{\text{why can this plane fly?}}\]—is to

(3) \[G_{\text{why can this plane fly?}}\] = believe the truth about Bernoulli’s principle, as well as the relevant details about the plane (its weight, the area of its wings, etc.); or

(4) \[G_{\text{why can this plane fly?}}\] = (more strongly) to know Bernoulli’s principle, as well as the relevant details about the plane.

According to Trout, the goal of this question is something along the lines of (4)—that is, the goal is to know this information. In the following passage Trout distinguishes this way of thinking about the goal from a way of thinking about the goal that gives pride of place to the phenomenology that might accompany an answer for someone.

‘Understanding’ has many senses. Consider a sense of understanding associated with an explanation of why jets fly. According to one sense of understanding, I understand why planes fly if I know Bernoulli’s principle. In such a case, I must have some knowledge of relevant background conditions, of course—how much the plane weighs, the area of the wings, etc. But in a stronger sense of understanding, I can know Bernoulli’s principle and still not understand why planes fly, because my knowledge that Bernoulli’s principle applies is not associated with a specific kind of subjective state or feeling; it does not have the phenomenology of understanding. (p. 222)

We might characterize this second, phenomenalized way of thinking about the goal along the follows lines:

(5) \[G_{\text{why can this plane fly?}}\] = to know Bernoulli’s principle, as well as the relevant details about the plane, and to have the feeling that this information correctly answers one’s question.

\[^{17}\text{Trout’s view here is a bit confusing because he suggests that knowing Bernoulli’s principle (plus some details about the plane) is a kind of understanding. To his mind, however, what most people have in mind when they claim that understanding is the goal of explanation is something entirely different than this. For them, understanding involves something along the lines of the second alternative he mentions: namely, not just knowing the correct answer but having a certain kind of phenomenological response to the answer.}\]
By Trout’s lights, moreover, (5) is what thinkers who have taken understanding to be the goal of explanation have generally had in mind.

Trout, unsurprisingly, thinks (5) is ridiculous. To suppose that knowing the correct answer with respect to our explanation-seeking why-question is not sufficient, and that this knowledge needs to be embellished by a certain feeling that the correct answer is right is not just silly but potentially a danger to scientific progress.18 As he writes,

The sense of understanding would be epistemically idle phenomenology were it not so poisonous a combination of seduction and unreliability. It actually does harm, sometimes making us squeamish about accepting true claims that we don’t personally understand, and more often operating in the opposite direction, causing us to overconfidently accept false claims because they have a kind of anecdotal or theoretical charm. (p. 230)

Reaching the goal of explanation, especially the goal of anything that deserves to be called scientific explanation, does not depend on waiting for a feeling of this kind.

Suppose that, for the moment, we agree with Trout that (5) can’t be taken seriously. What should we make of Trout’s favored alternative, (4)?

Despite its obvious appeal, the first strike again (4) is that knowing all of these things is compatible with never attempting to apply this knowledge to the question of why this plane can fly. But it would be a mistake to suppose that my explanation-seeking why question could be answered without having to do any

18 Trout, together with Michael Bishop, mocks this suggestion in their co-authored (2002): “These psychological descriptions of understanding focus on their phenomenology. There is “something that it is like” to understand, and we use the precise character of this subjective sense that we understand—a psychological impression of coherence, confidence, etc.—as a cue that we do indeed understand. But the sense of understanding no more means that you have knowledge of the world than caressing your own shoulder means that someone loves you. Just ask Freud” (pp. 203-04).
work of this kind, or simply in virtue of the fact that the knowledge is stored away somewhere in my mind.\textsuperscript{19}

Second, and more revealingly, even if someone attempts to apply his knowledge—even if, as it were, he gives it a try—he might lack the competence to apply his knowledge correctly. In such a case, he might then know Bernoulli’s principle in a very ordinary sense of know—he might be able to report it to you accurately, if asked on a hundred tests to reproduce the principle he might get the answer right every time, and so on—and yet he might nonetheless lack the ability to apply the principle correctly. Indeed, this phenomenon is quite common. As science teachers are well aware, it is one thing to know a certain principle (Maxwell’s equations, Newton’s laws, etc.) and quite another to be able to apply the principle to particular cases—a failure that is often reflected in a student’s ability to do well on the rote portion of a test and terribly on the application portion of the test.\textsuperscript{20}

But if merely knowing Bernoulli’s principle along with certain facts about planes is not what we are after when we pose our “Why can this plane fly?” question, then what else is needed?

\textbf{4.4 Why Questions}

In the following two sections I will suggest that progress can be made in characterizing [G] in two ways. First, by being more precise about the nature of

\textsuperscript{19} Towards the end of the Trout passage just quoted, there is a suggestion that the person doesn’t just know the principle but “knows that it applies” to this case. The second problem would then take precedence: knowing (perhaps, via testimony) that a principle applies to a particular case is not the same as being able to apply it.

\textsuperscript{20} As Thomas Kuhn (1970, pp. 23-51) and Philip Kitcher (1989, p. 437-38) nicely bring out.
the explanation-seeking why-questions that drive inquiry. Second, by taking a closer look at how a typical first-person case of inquiry might unfold.

Building on an insight of Hempel’s (1965, pp. 421-23), and in keeping with the proposal from Chapter 3, the first thing to notice about our “why questions” is that they are typically more focused and specific than their surface form might suggest. To take one of Hempel’s examples: Suppose you tell me that you are curious about the eruption of Mt. Vesuvius in A.D. 79. My natural response will be: Well, which part of the eruption are you curious about? For notice, prima facie at least the event that answers to the name “the eruption of Mt. Vesuvius in A.D. 79” looks to be enormously complex.21 As Hempel points out:

A complete description of the eruption of Vesuvius in A.D. 79 would have to specify the exact time of its occurrence; the path of the lava stream as well as its physical and chemical characteristics—including temperature, pressures, densities, at every point—and their change in the course of time; the most minute details of the destruction wreaked upon Pompeii and Herculaneum; full information about all persons and animals involved in the catastrophe... and so on ad infinitum. (p. 422)

When you say that you are curious about the eruption of Vesuvius, what you therefore really seem to mean is that you are curious about a particular aspect of the eruption of Vesuvius. Perhaps, for example, your question is: Why did it occur in April rather than May? Or perhaps: Why did the lava flow this way rather than that? Moreover, while it may be that you are curious about many of the aspects of the eruption—perhaps even all!—it seems clear that the respective answers to these questions will invoke very different explanatory stories.

Whatever it is in virtue of which the lava flowed one way rather than another, for

21 Given a coarse-grained view of events, at least. More on this point in the following footnote.
example, will presumably not be the same thing in virtue of which the eruption occurred in April rather than May.22

Peter Lipton (2004), drawing on Hempel, expands on this point:

Suppose we started by construing a phenomenon to be explained simply as a concrete event, say a particular eclipse. The number of causal factors is enormous. As Carl Hempel has observed, however, we do not explain events, only aspect of events. We do not explain the eclipse *tout court*, but only why it lasted as long as it did, or why it was partial, or why it was not visible from a certain place.... We may not explain why the leaves turn yellow in November *simpliciter*, but only for example why they turn yellow in November rather than in January, or why they turn yellow in November rather than blue. (p. 33)

Lipton’s refinement of Hempel’s point bears emphasizing. According to Lipton, the essential lesson to be learned from Hempel is not just that our explanatory interests are quite specific. More exactly, the lesson is that our explanatory interests are typically *contrastive*.23 In other words, when we focus on a particular aspect of an event what we want to know, the thing that stimulates our curiosity, is not simply “Why is the event this way?” but rather the question “Why is it this way rather than that?”24

22 As I suggested in the previous note, Hempel’s view assumes a coarse-grained stance on the nature of events that is widely contested. In a way, however, the ongoing metaphysical dispute about the nature of events is beside the point. Whether, as Hempel thinks, what we explain are aspects of events, where the events themselves are coarse-grained, or whether instead one thinks that events are much more fine-grained and that as a result what Hempel refers to as *aspects* of coarse-grained events are in fact simply events full stop—on either view the basic point still stands: the thing we want to explain is quite fine-grained. For two classic contributions to this debate, see Kim ([1976] 1993) and Bennett (1988). For a good contemporary overview, see Simons (2003).

23 Others who have emphasized this idea include Garfinkel (1980) and van Fraassen (1980, esp. ch 5).

24 As we learned in Chapter 2, it is our sense that things could have been otherwise that plausibly invites the explanation-seeking why question in the first place, and what we want to figure out is why the thing we want to explain is one way rather than another way in which, by our lights, it might have been.
4.5 Leopold

Applying this point to a first-person case of inquiry will help to illustrate the idea. Suppose that three people are presented with an unfamiliar machine that lights up and plays music when some objects but not others are placed on it. Despite the fact that they are all observing the same string of happenings, their explanatory interests might vary considerably. Fran might be interested in why the machine plays one song rather than another, for example, and Clark in why the lights flash in this order rather than that. Leopold, finally, might simply be interested in why the machine sometimes lights up and plays music rather than remains silent. Moreover, not only are Fran’s, Clark’s, and Leopold’s why-questions and explanatory interests therefore quite different, in pursuing their questions they will presumably be led to quite different answers: the thing in virtue of which the lights flash in a certain order will most likely be different from the thing in virtue of which the machine lights up rather than remains silent, for example, and so on for the other answers.

Suppose we now focus on one of these questioners, Leopold, and imagine that in searching for an answer to his question Leopold proceeds by testing a group of potentially relevant factors one by one. First he tests to see if the shape of the objects is correlated with the music and lights. What he finds is that it is

---

25 The example that follows is a modified version of a study designed by the cognitive psychologist Alison Gopnik and her collaborators. See Gopnik and Sobel (2000), Gopnik et al. (2001), Gopnik and Glymour (2002), and Gopnik et al. (2004). An unmodified version of this case will appear again in Chapter 5.

26 Since this is an artifact we are talking about, we might be tempted to say that it is the programmer who is responsible for both. Since this gets us into the tricky field of action explanation, I won’t pursue this objection beyond saying that even if both explanatory stories appeal to a programmer, his reason for wanting the machine to light up etc. will presumably be different than his reason for wanting the lights to flash in this order rather than that.
not: sometimes cubes set off the lights, sometimes they don’t, sometimes pyramids set off the lights, sometimes they don’t, and so on. He then tests for color, and again finds a lack of correlation: sometimes red things set off the machine, sometimes they don’t, and so on for all the other colors. Finally, he tests to see whether the weight of the blocks makes a difference. What he finds is that all of the blocks that weigh over two pounds, regardless of their shape or color, set off the lights, and that none of the blocks under two pounds set off the lights, regardless of their shape or color.

Leopold now has a good deal of evidence at his disposal, but merely amassing this evidence is obviously not the same thing as answering his question. What else is needed? Well, suppose that at some point near the end of his testing it begins to seem to him—alternatively, he seems to see—that the lighting up etc. depends on the weight of the blocks. In other words, it begins to seem to him—alternatively, he seems to see—that the weight of the blocks is the thing that makes a difference to the lighting up etc.

In describing the change in Leopold’s cognitive state we just appealed to two different senses of “seeming.” First, that “it seemed to him that...”; second, that “he seemed to see....” Are these just two different ways of referring to the same change?

Although both expressions have the same ring to them—roughly, the same “phenomenological” ring—in the end they arguably refer to quite different

---

27 Well, for one thing he might think that at this point more testing is in order, depending on how many variables he thinks he has to consider and how confident he is that he has controlled for other factors. But for simplicity we can suppose that to his mind all of the apparently relevant variables have been accounted for.
cognitive states. In the first sense, when we say, of Leopold, that it “seemed to him” that the lighting up etc. depended on the weight of the blocks, what this suggests is that he felt a strong *inclination* to affirm the existence of such a dependency. I will return to this “felt attraction” phenomenology later in Section 4.8, but for our purposes the most important “seeming” to account for is the sense in which he “seems to see” that the one thing depends on the other.

One strong temptation is to suppose that Leopold’s “seeming to see” is essentially a phenomenological episode in the same way that the felt attraction was essentially a phenomenological episode. Specifically, the temptation is to say that in seeming to see that the lighting up etc. depends on the weight of the blocks what Leopold is doing is enjoying a private conscious episode of some kind: a *seeming.*

We have other options, however. For instance, arguably when we say, of Leopold, that he “seems to see” that the one thing depends on the other, what we (or Leopold himself, if he is speaking of himself) are doing is (a) marking the fact that a new capacity or ability of Leopold’s has entered the scene—provisionally, what we can think of as a capacity for “seeing dependencies,” and (b) drawing attention to the fact that this particular exercise of the faculty might be mistaken. Since “seeing” is a success term, it is often safer—at any rate, more prudent—to say of someone that he “seems to see” such a dependency rather than that he simply “sees” it.

---

28 George Bealer (1996, 1999) recommends this interpretation with respect to *a priori* intuitions. According to Bealer, when it seems to Paul as if 2+2=4, for example, we should interpret this to mean that Paul is enjoying “a genuine kind of conscious episode” as-if 2+2=4 (1999, p. 30). For trenchant critiques of this view, see Sosa (1996) and Williamson (2004).
Between these two ways of thinking about Leopold’s seeming to see—on the one hand, of “phenomenalizing” it, and on the other hand, of identifying it as a fallible exercise of his capacity for seeing dependencies—the second seems preferable. Consider an analogy with memory. If Leopold were to say that, “I seem to remember that I left my keys on the kitchen table this morning,” the natural way to interpret his “seeming to remember” is not in terms of a private phenomenal “seeming” episode—an episode that, as several philosophers have pointed out, introspection typically fails to reveal—but rather as an expression of his concern that this exercise of his memory might be mistaken. Even though on the second interpretation of “seeming to remember” the existence on an internal memorial “seeming” episode is not positively ruled out, it nonetheless begins to look like an extra wheel, and one that we can do without.

*Mutatis mutandis*, the same considerations apparently apply to Leopold’s experience of “seeming to see.” What’s more, if we can account for his “seeming to see” in this way—namely, as an indication of concern that a particular exercise of his “seeing” faculty might be mistaken—without appealing to the existence of a suspect private “seeming” episode, then from a theoretical point of view so much the better.

From the point of view of our overall project, moreover, there is a still better reason to favor the second interpretation: namely, because the second interpretation highlights the crucial fact that a new faculty or ability has entered the scene at this point. When Leopold “sees” or “seems to see” that the one thing

\[29\] See, for example, Plantinga (1993, p. 57-64) and Audi (1995).
depends on another, he is no longer just passively taking in the world as it unfolds before him, as the visual connotations of the “seeing” metaphor might suggest. Instead, when he exercises his capacity for seeing dependencies he is in some fundamental sense reaching out towards the dependency in a new and distinctive way—in some sense grasping or comprehending it.

Granted, this is all still metaphor, and in the following section I will try to add more substance to this suggestion.\textsuperscript{30} The important thing to see for now is how naturally and perhaps even inescapably the language of “seeing” in this context seems to require unpacking in terms of this other, more hands-on or manipulative language.

One final point. For convenience in what follows it will help to have a name for this apparent capacity for seeing dependencies. We might call it “the understanding faculty,” but one problem with this name is that it is so broad—as we noted in Chapter 1, we understand concepts, motives, works of art, and on and on. It would be better to have something more focused, something that specifically captures the kind of manipulative ability that seems to be the hallmark of this capacity.

At the risk of being dull, I recommend that we simply call it “the grasping faculty.” This too sounds quite broad, but it is at least broad in a way that focuses on what seems to be distinctive about the ability. It should also go without saying that nothing overly strong should be read into the notion of a “faculty” here; we would do just as well to think of it as a capacity or ability. What the notion of a

\textsuperscript{30} Especially when one notes that the infinitive to comprehend traces back to the Latin comprehendere—to grasp or seize.
faculty captures, however, is the apparent distinctiveness of this capacity or ability, a fact that is worth emphasizing.

4.6 Bernoulli again

So what does all this have to tell us about the kind of epistemic attitude we want to bear towards a correct answer to our why-questions? In light of the previous discussion, here is one promising idea: what is epistemically distinctive about the goal of explanation is that we grasp the correct answer, one consequence of which is that we are able to put the correct answer to use.\(^{31}\)

With respect to our previous question about the plane, we can characterize this way of thinking about the goal as follows:

\[
\text{(6) } [G \text{why can this plane fly?}] = \text{grasping Bernoulli’s principle and applying the principle to the relevant details about the plane.}
\]

What would it take to satisfy (6)? Recall that Bernoulli’s principle states that as the speed of a moving fluid increases (air being just one example of a moving fluid), the pressure within the fluid decreases. Grasping Bernoulli’s principle therefore involves at least the following. First, it involves the ability to anticipate how changes in the value of one of the variables referred to in the principle—the velocity of the fluid, for example—would lead to (ceteris paribus) a change in the

---

\(^{31}\) Henk de Regt (2004) has recently suggested that it is this ability to put a theory to use that constitutes the essence of understanding—and thereby, de Regt would presumably claim, the goal of explanation—and that “subjective feelings” of the sort described by Trout are not even necessary for understanding (p. 104). De Regt writes: “Understanding involves more than mere knowledge of the relevant formula (or in the scientific case: theories, laws, and background conditions), but this ‘more’ should not be sought in any of the subjective experiences that may accompany the appearance of understanding.... Understanding is not only knowing the formula, but in addition being able to use the formula in the case at hand” (p. 101). De Regt’s suggestion is not quite the same as (6), however, because he suggests at the end of the passage that the state he has in mind includes knowing the correct answer. If one is careful about how to spell out knowledge here, I think this is on the right track.
value of another variable—in this case, the pressure of the fluid. Second, it involves the ability to apply the principle to particular situations.

For example, suppose I believe that because the shape of this airplane’s wings is raised along the top and flat along the bottom, the air traveling across the top of the wing will speed up to meet the air traveling along the bottom of the wing. If I have grasped Bernoulli’s principle, and if I then apply the principle to this case, what I will “see” is that, given that the air traveling across the top of the airplane’s wing is moving at a higher velocity, the pressure exerted by the air on the top of the wing will decrease. Moreover, I will “see” that since the pressure exerted by the slower moving air along the bottom of the wing will then be greater than the pressure exerted along the top of the wing, lift—in other words flight—will result.

As this example suggests, although I spoke a moment ago as though grasping a principle involves two distinct abilities—on the one hand, an ability to manipulate the variables expressed by the principle and on the other hand an ability to apply the principle—in a way this is misleading. In reality, it seems to be one and the same ability which allows for both accomplishments. For instance, if I lacked the ability to manipulate the variables expressed by the principle, then it is hard to see how I would be able to apply the principle to particular cases. And if I could not apply the principle to particular cases—if I could not anticipate how inserting particular values into one variable would

32 And, for that matter, a bird’s wings—the same principle applies.
influence the value of another—then it is hard to see how I could genuinely have the ability to manipulate the variables.  

Seeing or grasping how the variables depend on one another in this way therefore requires a distinctive ability or skill. Describing the goal in terms of having an ability to put the correct answer to use in this way thus represents a significant improvement over thinking of the goal along the lines of (5), where the goal was to have something like warm thoughts towards the correct answer. So... is (6) correct?

Although I will soon argue that, understood in the right way, (6) is essentially correct, the notion of grasping still contains an important ambiguity that needs to be clarified. Let us introduce a distinction between conditional and simple grasping. The difference amounts roughly to this: when I conditionally grasp that one thing depends on another, say that A depends on B, although I grasp that A depends on B, I grasp this only in the light of certain assumptions that I either fail to accept or that I take to be importantly incomplete. When I

---

33 Another thing worth pointing out about this ability is something that we have mentioned before at the end of Chapter 2: namely, the general ability seems to carry with it the specific ability to answer “what-if-things-had been-different?” questions. So, for example, if I have grasped Bernoulli’s principle and I am then asked, “Would the plane still have been capable of flight if (ceteris paribus) the top of the wing had been flattened out?” I can and will answer No. What grasping the principle brings with it is the ability to recognize that, if the principle is correct, the lift of the plane depends on the difference in pressure on the wing (specifically, a greater force upwards), and that the difference in pressure depends on the difference in fluid speed traveling over the wing. And in the hypothetical case just presented what I recognize is that if the top of the wing were flattened out there would be no difference in speed, hence no difference in pressure... hence no lift, no flight.

34 Alternatively, I grasp this by imaginatively supposing that the assumptions that I take to be false are in fact true, or by supposing that the assumptions are not, in fact, importantly incomplete. This act of imaginatively taking on certain assumptions and abandoning others in order to see what follows from them should be entirely familiar from our everyday interaction with others, but it should be especially familiar to teachers. When we try to understand the point behind a student’s question, for example, we try to get inside the students head, provisionally taking on his beliefs and abandoning our own in order to see what follows. Notice as well, given
simply grasp that A depends on B, by contrast, none of these qualifications are relevant: in such a case, I grasp that A depends on B in (as it were) a straightforwardly assenting way.

Although simple grasping is perhaps more common, conditional grasping should also be very familiar to us, as the following examples suggest.

First example of conditional grasping. Suppose that Eileen is a contemporary scientist who doesn’t believe a word of phlogiston theory; she does, nevertheless, grasp it very well. Indeed, she might even be the world’s expert! Thus if Eileen is asked, “What is it that accounts for the combustibility of some materials but not others?” she will be able to grasp or “see” what the difference depends on according to phlogiston theory: namely, on the fact that some materials are phlogiston rich and others are not. The qualification, “according to phlogiston theory,” is crucial here, however, because it is not as if she simply grasps or sees that the difference between the combustibility of materials depends on the fact that some materials are phlogiston rich and other are not. Instead, she grasps or sees that the one thing depends on the other, only given certain assumptions—assumptions that, as it happens, she does not accept.

Second example. Imagine that Paul’s question is: “What accounts for the development of species over time?” Or, a little differently: “Why do species develop over time rather than remain static?” Although he grasps Darwin’s answer (that development is driven by natural selection working on random

35 The more natural way of putting things here is that she understands it very well; since my project in this chapter is to try to characterize the goal in independent terms, that notion is not yet available to us.
genetic mutation), he prefers Lamark’s (that development is driven by organisms adapting to their environment and then passing these adaptations on to their descendants). Lamark’s theory just seems more elegant to him, or perhaps more in keeping with his belief that the universe is not such a random, messy place. Nonetheless, he grasps the Darwinian answer to his question perfectly well, and he can competently apply it to particular cases. It is just that Paul fails to accept that species develop in the way that Darwinism suggests, in light of the rest of what he believes.

Unlike in the first example, the grasping in this second example is therefore conditional in a different way; it is conditional not because Paul necessarily thinks the assumptions of the competing explanation are false but rather because he thinks they are incomplete. They fail to include something else Paul believes that, to his mind at least, tells against A’s dependence on B.

As all of this suggests, when we simply grasp that A depends on B, this is because we think that this dependence follows from the rest of what we believe. For the most part (as we will see in greater detail in Chapter 5), this “follows from” inference that we draw is not conscious; it is not as if we typically pause and say, “I grasp that A depends on B, because it follows from these assumptions.” Rather, in light of what we already believe, we normally simply find ourselves grasping (perhaps accompanied by a moment of “Ah, I see...”) that the one thing depends on the other.
4.7 Tidying Up

So how does all this apply to the airplane case? As (6) stands, the basic problem is that if grasping is read here as “conditional grasping,” someone might grasp the correct answer in this way and yet not even believe that the answer he is grasping is correct! Just as one can grasp Darwin’s theory and apply it to particular cases while still thinking that Lamark had the real answer, so too one might both grasp Bernoulli’s principle, as well as apply it to the case at hand, and yet not believe a word of it!

In light of our previous distinction, however, the repair is relatively easy. Thus we might say:

(7) [Gwhy can this plane fly?] = *simply grasping* Bernoulli’s principle and applying the principle to the relevant details about the plane.

Although in the context of the dialectic (7) helps to clarify things, nevertheless for overall reasons of clarity, or perhaps just aesthetics, it is less than ideal.

For one thing, since “simple grasping” is a term of art, it is less likely to remain in our minds than the more straightforward “grasping.” Arguably as well, when we say (without any philosophical build-up) that we grasp that A depends on B, we naturally associate the attitude of grasping with an attitude of assent or affirmation: in addition to “grasping” or “seeing” how the two terms are modally related, we also affirm an apparent fact about the world—namely, that such a dependency exists.

As far as characterizing the goal of our why-question about the plane goes, to my mind it would therefore be best to stick with (6) after all, and when it is

\[ G \text{why can this plane fly?} = \text{grasping} \]

Bernoulli’s principle and applying the principle to the relevant details about the plane.

And to refresh our memory, here it is again: (6) [Gwhy can this plane fly?] = grasping Bernoulli’s principle and applying the principle to the relevant details about the plane.
conditional grasping that we have in mind, in order to avoid confusion we can mark this as a special form of grasping: call it grasping*. In other words, and bearing in mind the previous distinction, we can put things as follows:

simple grasping = grasping

conditional grasping = grasping*

Although grasping* still introduces a term of art into the discussion, since it will not occupy the featured role that grasping will in what follows, it is less likely to clutter things.

If (6)—properly understood—is a good characterization of the goal of our why-question about the plane, at any rate, we can now return now, finally, to our original question about the goal of explanation [G]. Provided that we were right in thinking that our why-questions are inspired in the first place by our desire to figure out why things are one way rather than another, moreover, we can now be more exact both about the kind of why-question that drives explanatory inquiry as well about the kind of epistemic attitude towards a correct answer that we are looking for, as follows:

(8) \([G_{\text{why this rather than that?}}] = \text{to grasp what the difference between these alternatives depends on.}\)

On the assumption that a correct answer to our “Why this rather than that?” question will be one that reveals what the difference between these alternatives depends on, we might also shorten (8) as follows:

(9) \([G_{\text{why this rather than that?}}] = \text{to grasp a correct answer.}\)
Sticking with the more informative (8) for the moment, although it too will eventually stand in need of refinement, most of the essentials now seem to be accounted for. An element of truth, for one thing. For another, an epistemic attitude that combines the attitudes of grasping and affirming the truth.

### 4.8 Phenomenology

I mentioned at the beginning of the chapter that the *phenomenological* aspect of the goal of explanation has been largely either dismissed as irrelevant or neglected altogether. The reason for this, I think, has been the concern to emphasize that the goal of explanation is fundamentally an *epistemic* goal, one that (in some way) involves being appropriately connected to the truth. So far in this chapter I have likewise focused on trying to give the epistemic dimension of the goal its due. In this section I will now consider how, if at all, phenomenological considerations enter into the picture.

Although trying to characterize the phenomenology of any mental state is something of a fool’s game, two elements stand out as specially connected to the goal of explanation. The first element is the “felt attraction” phenomenology that is familiar to us from Section 4.5, and the second is the celebrated *Aha!* experience.

In the normal course of things the felt attraction phenomenology is so familiar that it generally escapes our attention; what we are focused on is the claim we believe, not the felt attractiveness of the claim. Taking a step back, however, what we see is that when we are inclined towards a certain answer to

---

37 Specifically, we will see in Chapter 6 that one can grasp a correct answer just by luck; arguably, however, we want more than that: something along the lines of *safe* or *secure* grasping.
our question, something about this answer will attract us or pull us in a way that another answer will fail to attract or pull us. Moreover, there is something that it is distinctively like to be attracted or pulled towards a claim in this way. As Plantinga (1993) notes:

Indeed, how could it [i.e., assenting with such a feeling] be otherwise? Could it really be that you should believe a proposition, even though it had none of this phenomenal attractiveness, this seeming-to-be-true—even if, that is, there was no felt or feelable inclination to believe it on your part? (p. 192)

Even though experiencing the felt attraction of a certain answer may not strictly speaking count as a necessary condition for achieving the goal, as Plantinga suggests, for beings like us it is almost impossible to imagine what it would be like to affirm an answer without it.

In contrast to the felt attraction phenomenology, the Aha! experience as it is usually characterized is much more dramatic. The physicist Steven Weinberg describes the Aha! experience, and defends its centrality to the process of explanation, as well as anyone:

It is not the job of philosophers or anyone else to dictate meanings of words [such as “explanation”] different from the meanings in general use. Rather than argue that scientists are incorrect when they say, as they commonly do, that they are explaining things when they do their work, philosophers who care about the meaning of explanation in science should try to understand what it is that scientists are doing when they are explaining something. If I had to give an a priori definition of explanation in physics I would say, ‘Explanation in physics is what physicists have done when they say Aha!’ (2004, p. 24-25; underline my emphasis)

In his earlier 1994 book, Dreams of a Final Theory, Weinberg describes the experience in even more explicitly phenomenological terms:

38 Perhaps there are Oliver-Sacks-style cases in which someone lacks this phenomenology altogether; there seems no reason to think they would not be interested in asking and answering explanation-seeking why questions.
It is a tricky business to say exactly what one is doing when one answers such a [why-] question. Fortunately it is not really necessary. Scientific explanation is a mode of behavior that gives us pleasure, like love or art. The best way to understand the nature of scientific explanation is to experience the particular zing that you get when someone (preferably yourself) has succeeded in actually explaining something. (p. 24; my emphasis)

When Weinberg writes in the first of the two passages that the connection between the goal of explanation and the Aha! experience can be known a priori, this suggests that he thinks the Aha! experience is a necessary condition on meeting the goal, or at any rate that for beings like us it is in some way essentially tied up with the goal. Is there anything that can be said for this?

Given the dramatic nature of the Aha! experience, the reasonable thing to say is that Weinberg overstates the case—certainly relative to everyday instances of inquiry and arguably even with respect to the “scientific” inquiries that are his explicit focus. Consider, for example, any one of the countless small episodes of inquiry that we engage in throughout the course of a day. Why didn’t my alarm go off this morning? After looking around for a moment I arrive at an answer: Ah, I see, because I forgot to set it. In other words, what I grasp (or seem to grasp) is that my failure to set the alarm is the thing that made the difference between the alarm’s going off and its remaining silent. But no dramatic Aha! moment was experienced. No special Zing! At least, if there were experiences of

39 According to theorists such as Herbert Feigl (1949), for example, the sign that one has reached the goal of explanation is a certain feeling of intellectual satisfaction or relief: a feeling of “Aha!” Indeed, according to Feigl, the deeper the explanation, the stronger the “Aha!” one experiences. Peter Godfrey-Smith (2003) likewise suggests that the Aha! at least characteristically accompanies the goal: “Indeed, it seems clear to me that what produces an ‘Aha!’ reaction is often the realization that some odd-looking phenomenon is really a case of something more general” (p. 196). See also Turney (2004).
this kind going on and I didn’t notice them, then the experiences are not all that they are cracked up to be.

Yet as the alarm clock example suggests, there does seem to be some distinctive phenomenology that accompanies the little graspings or seeings that we enjoy throughout the day. As a tribute to the “Ah, I see” moment from the alarm clock case, we might think of this distinctive experience as an “Ah experience”—the less showy cousin of the Aha! experience. One suggestion worth considering, then, is that even when an episode as dramatic as the Aha! does not accompany the goal of explanation, some such moment of illumination, if only the Ah, always does—if not always verbalized (that would be quite rare), then at least always felt.

Since first-person introspective research of this sort is bound to be less than conclusive, and since reasonable people might therefore disagree about the pervasiveness of the Aha!/Ah experience, this suggestion would be on much stronger footing if there were independent reason to think that it would in some sense be fitting for the moment of grasping or seeing to be accompanied by a distinctive phenomenology of its own. Alison Gopnik (1998) has recently suggested just such a reason.

As Gopnik points out, once one appreciates the potential power that comes from grasping how things depend on one another in the way specified above, it is not particularly surprising both (a) that grasping these dependencies should be uniquely desirable as an epistemic goal, and (b) that, over the course of time, a special kind of phenomenology should have come to be associated with the exercise of our grasping faculty. Just as, Gopnik suggests, in order to encourage
reproduction it “made sense” (as it were) for Mother Nature to associate sexual orgasm with its own pleasurable phenomenology, so too it would seem to make just as much sense for graspings or seeings to acquire their own distinctive, and distinctively desirable, phenomenology. Since the ability to grasp dependencies in this way (and thus to manipulate one’s environment) is obviously a fitness-enhancing characteristic, it is quite natural to suppose that, over time, evolution would have found a way to impress the goodness of such graspings upon us.

Suppose at any rate that we are persuaded either by our own introspective research or by the plausibility of Gopnik’s just-so story (or both) that something along the lines of the Aha!/Ah experience characteristically accompanies our graspings of the correct answer. Whether we should say that, strictly speaking, these phenomenological elements are necessary for the achievement of the goal or whether instead we want to say only that there is a more contingent connection—perhaps that normally we experience these ingredients in conjunction with the goal—is a question I will not try to settle here. For our purposes it is enough to appreciate the pervasive role that these phenomenological elements play in our acquisition of the goal.

### 4.9 How Does it Match?

Given that we now have quite a fairly complete picture of the goal of explanation, we can finally ask: How well does understanding resemble the picture?

In two central respects at least the resemblance looks remarkably good. First, there is general agreement that, like the goal of explanation, understanding
is a state that essentially involves an element of grasping or seeing. As virtually every theorist who has written on the topic has noted, if nothing else understanding involves the ability to bring together various bits of information and to “see” how they are connected. Moreover, understanding is often thought to involve a distinctive phenomenological element, one that has typically been characterized in terms of the by-now-familiar Aha! or Eureka! moment. So in this respect too understanding is a good match for the goal.

As we noted at the outset of the chapter, however, the basic reason to think that understanding is not the goal of explanation—in other words, that it fails to match the profile—is that whereas understanding is possible in the absence of truth, truth is essential to the goal of explanation. To put the point a different way: what we are aiming for in our explanatory inquiries is the truth. Since understanding can be found apart from the truth, however, then we are not aiming for understanding.

Consider the CIA conspiracy theorist again (call him Andy), and suppose that he wakes up one morning and finds that the tapwater tastes funny, leading him to wonder why it tastes funny. Given that Andy believes that the CIA is behind everything suspicious in the world, he will more or less automatically “see” (or seem to see) that the tapwater tastes funny because of the involvement of the CIA. In other words, given his background beliefs, he will “see” (or seem to see)

---

40 In our discussion of the goal of explanation, moreover, we have (I hope) made some progress on what this seeing amounts to: namely, the ability to identify dependencies and to manipulate the elements that constitute the dependencies—specifically, to anticipate how a change in one of the elements will result (ceteris paribus) in a change in the other. “Seeing,” in this sense, is thus at least partly demystified; it calls on the same kind of manipulative abilities that we are familiar with from our hands-on interaction with the world, though now raised to the level of mental manipulation.
see) that it was the involvement of the CIA that made the difference between the tapwater’s normal taste and its current taste.

According to the objection considered at the beginning of the chapter, the problem now is that even though Andy apparently enjoys some form of understanding with respect to why the tapwater tastes funny, with respect to the goal of explanation he clearly falls short. The goal of his inquiry was to find the truth and he missed the truth, instead ending up with something significantly less valuable: a misbegotten (albeit still genuine) case of understanding.

4.10 Understanding is Factive

To see where the objection goes wrong consider how a parallel argument might go with respect to memory. Imagine that one of our goals is to acquire accurate information about the past. If so (the parallel argument continues), it would be a mistake to think that remembering things about the past could give us what we want. After all, people have remembered all kinds of false things—even crazy things! For example, UFO enthusiasts remember being abducted by aliens! Since our goal is to acquire accurate information about the past, remembering things about the past is therefore not what we are after.

The problem with this objection, naturally, is that since remembering is a factive mental state, the UFO enthusiasts do not actually remember being abducted. Instead, they only seem to remember being abducted.41 We should therefore classify cases of this sort not as:

---

41 Notice: this will be a still different sense of “seeming” than the various senses of "seeming" we distinguished with respect to Leopold’s “seeming to see” experience. Sometimes we use “seeming” language to remain neutral between success and failure: or, better, to express a
(a) a case of remembering that sadly missed the truth;

but rather as:

(b) an exercise of memory (of the memory faculty, if you will) that sadly missed the truth.\footnote{If you are inclined to think this is more an exercise of the imagination than of memory, then substitute the following example: I cannot remember that my childhood phone number was 635-9688, because it wasn’t (it was 635-9689). If I did seem to remember that my phone number was 635-9688, however, then presumably this would be an exercise of my memory that sadly missed the truth.}

In the same way, I suggest, what we should say of Andy’s story about the CIA and the tapwater is not that:

(c) this is a legitimate case of understanding that sadly missed the truth;

but rather that

(d) this is an exercise of his grasping faculty that sadly missed the truth.

The reason why the two are apt to be confused, moreover, is that (d), even though correct, is not commonly distinguished from (c). Most people, presumably, fail to give the existence of a "grasping faculty" a second thought. Since they sense that there is something true in the neighborhood, however, they are drawn almost by default to (c). But (c) is just as incoherent as (a)! In both cases the right answer lies with the second alternative: (b) and (d).

What we should say of Andy, therefore, is not that he understands why the tapwater tastes funny but that he \textit{seems to understand} why the tapwater tastes funny. At the same time, we can acknowledge that despite the fact that Andy falls short of understanding, his experience is—from an internal perspective—
essentially identical to the experience of someone who genuinely understands. What’s more, we can recognize Andy’s state of seeming to understand as a kind of genuine accomplishment. Andy does exercise his grasping faculty with respect to a possible dependency, after all; thus he can anticipate, for example, how a change in the value of one of the elements in the dependency will result in a change in the value of another element of the dependency. It is just not a dependency that actually exists.

4.11 Summing Up

A gradual case has been made that the goal of explanation—the thing we want when we pose our why questions—is to grasp the truth about why things are one way rather than another. For creatures like us, moreover, the act of grasping is accompanied by a distinctive phenomenology, which I have described both in terms of a “feels right” phenomenology and in terms of the Ah or Aha! phenomenology. In all of these respects, however, the resemblance between the goal of explanation and understanding is remarkably strong.

43 A final comparison with memory is again helpful. When I misremember my childhood phone number I am nonetheless exercising the same ability or capacity—namely, memory—that I exercise when I genuinely remember my childhood phone number. It is just that in the former case, my ability has temporarily failed me. From the inside, however, this seeming to remember will often “feel” exactly the same as actually remembering.
CHAPTER 5
RELIABILITY AND UNDERSTANDING

How reliable is this grasping faculty? Looking at the historical record, you might think that—even if we agree that understanding is factive—whatever graspings come from this faculty won’t have much if any positive epistemic status because the faculty itself is so poorly connected to the truth.

J. D. Trout (2002) cites the following historical examples as evidence against the reliability of the faculty.

Pierce identifies the distinctive cognitive experience of explanatory understanding by isolating the final moment of acceptance; the good explanation “is turned back and forth like a key in a lock.” This description alone should supply little solace to those holding that good explanations are epistemically reliable. After all, alchemists surely felt the key turn, but once inside we find only false descriptions of causal mechanisms. And when Galen arrived at a diagnosis of melancholy due to black bile, his sense of understanding was so gratifying it must have balanced his humors. (p. 213)

Later he is even more emphatic:

The fact is, our history is littered with inaccurate explanation we confidently thought were obviously true: the explanation for mental illness in terms of demonic possession, the humoral theory of illness, and so on. The sense of understanding would be epistemically idle phenomenology were it not so poisonous a combination of seduction and unreliability. It actually does harm, sometimes making us squeamish about accepting true claims that we don’t personally understand, and more often operating in the opposite direction, causing us to overconfidently accept false claims because they have a kind of anecdotal or theoretical charm. (pp. 229-30)

By Trout’s lights, apparently, the unreliability of our “sense of understanding” is just obvious; at any rate, even a casual look at the history of science confirms it.
Ptolemy, Galen, the alchemists, the astrologers... the list of people who confidently identified dependencies that either didn’t exist or that were entirely unrelated to the thing they wanted to explain goes on and on.

To bring Trout’s objection into line with our discussion so far, a bit of translation is needed. When Trout claims that “the sense of understanding” is unreliable, what he seems to mean by this is that the *phenomenological response* we feel with respect to some explanatory stories but not others is poorly correlated with the truth of those stories. In other words, that more often than not (or at least, not often enough for some other reliability threshold Trout has in mind) when a certain explanatory story “feels right” to us, it turns out to be wrong. In light of our earlier discussion, we can reframe the objection as follows: according to Trout, *our grasping faculty* as a whole is poorly connected with the truth. In other words, the dependencies that it “sees” (grasps, identifies) turn out to be false more often than not.¹

If the exercise of our grasping faculty is as unreliable as Trout suggests, at any rate, then we would have a serious problem: on the uncommon occasion when our grasping faculty identifies a *genuine* dependency, the identification will be too accidental to have much if any positive epistemic status. For comparison, suppose you claim to have the ability to determine whether someone is honest just by looking at them. If it turns out that most of the people that you classify as

---

¹ Of course, as we saw in Chapter 4, for beings like us the very act of grasping or seeing will be accompanied by a distinctive phenomenology—among other things, it will “feel right” to us that there is such a dependency—so Trout is right that these phenomenological elements will be in the picture. But on our picture the source of the unreliability (if there is unreliability to be found) will not be the phenomenology per se but rather the exercise of the grasping faculty that the phenomenology accompanies.
honest are deceitful crooks, then on the few occasions when you identify a genuinely honest person we would not say that your identification had much going for it in the way of positive epistemic status. To earn *that* a more secure link to the truth is required.

In this chapter I will argue, first, that a bit of armchair empirical research suggests that Trout dramatically exaggerates the unreliability of our grasping faculty. Second, I will point out that current clinical research reinforces what our armchair empirical research suggests: namely, that as a simple matter of record we are quite good at identifying dependencies. One tempting conclusion to draw from this research is that, *contra* Trout, our grasping faculty simply *is* reliable. In Sections 5.4 and 5.5 I will suggest that this conclusion too is mistaken. What we need is a compromise position, one that recognizes the presumed positive epistemic status of most of the exercises of our grasping faculty while also making sense of the fact that astrologers, conspiracy theorists, and the like can go so systematically wrong.

Since the purpose of this chapter is to evaluate whether and in what sense our grasping faculty is reliable, it is important to bear in mind how we normally assess reliability. If we are evaluating whether a certain thermometer is reliable, for instance, it is no strike against the reliability of the thermometer to point out that it would fail to give an accurate reading on the sun.² Or again, if we are evaluating whether Ichiro Suzuki is a reliable hitter of baseballs, then it is no strike against Ichiro’s reliability to point out that he would fail to make contact

with the ball while in the dark or while blindfolded.\(^3\) Judgments about reliability, in other words, are always made relative to a certain environment. I will simply assume throughout that the environment that is relevant to the assessment of our grasping faculty is the environment in the “normal” world—which is to say, at least, not a Cartesian demon world, or a world where all the walls have been replaced with funhouse mirrors, and so on.\(^4\)

### 5.1 Armchair Evaluation

What should we make of Trout’s objection? One immediate worry is that he fails to specify a reference class for his unreliability claim. In pointing out that astrologers, alchemists, and the like have “seen” or “grasped” dependencies that aren’t there, he then immediately attributes the unreliability of these apparent seeings to the unreliability of “the” sense of understanding, as if our grasping faculty were in some way unreliable at the root. But that doesn’t follow. Just as from the fact that the visual system of cataract sufferers is unreliable nothing follows about the reliability of “the” visual system, so too from the fact that the grasping faculty of a certain sub-group of people is unreliable nothing follows about the reliability of “the” or “our” grasping faculty.\(^5\)

---

\(^3\) Cf. Greco (2000, pp. 211-27).

\(^4\) As we’ll see, however, the reliability of someone’s grasping faculty will depend heavily on the accuracy of the person’s background beliefs, and I will make no attempt to specify what counts as normal with respect to background beliefs.

\(^5\) Of course, for this last analogy to hold there would have to be something distinctive about the grasping faculties of this sub-group of people, some reason why their faculties identified the wrong dependencies. Starting in Section 5.4 I will suggest just such a reason, but for now it is enough to point out that the transition from the one group to the other is not automatic.
A further problem is that even at the individual level the reliability of someone’s grasping faculty is apt to vary. Consider Ptolemy again. Trout, needless to say, is right to point out that Ptolemy’s grasping faculty led him astray on occasion, sometimes spectacularly. One thing that focusing on the spectacular failures overlooks, however, is that there were presumably also countless other occasions throughout Ptolemy’s life when his grasping faculty served him well. We can only speculate about what exactly these occasions might have been, but we can imagine that there were many: that the fire was failing to catch because the wood was wet, for example, that the horse was limping because it just suffered a bad fall, that Ptolemy Jr. was slurring his words because he just drank a bowl of wine, and so on. Assuming this is right, then the correct thing to say about Ptolemy’s grasping faculty is not that it was unreliable simpliciter but rather that, at most, it was unreliable with respect to some domains and reliable with respect to others.

Applying this point more generally, what a casual glance from the armchair suggests is that most people—perhaps overall and at least with respect to specific domains—are quite reliable at identifying genuine dependencies. Does any non-armchair, clinical research support this suggestion?

Evidently, Yes. In the following section I will consider the results of this research in some detail, both for the light that it sheds on our reliability question but just as importantly because it will help to put into sharper focus many of the quasi-technical expressions (for example, the notion of dependence, the notion of a “grasping faculty,” and so on) that to this point we have used in only a rough and informal way.
### 5.2 Clinical Evaluation

In a series of publications, the cognitive psychologist Alison Gopnik (along with a number of research partners\(^6\)) has argued that human beings are built with a distinct cognitive system for identifying causal dependencies in the world. According to Gopnik, the system basically functions as an input-output device: it takes as its input patterns of observed contingency and correlation among phenomena, and then mines these observations to identify apparent dependencies among the phenomena. Alternatively, it mines them in order to construct what Gopnik calls “causal maps” of the world.

Here is one example of such a causal map, which I will first describe in general terms and then fill in with a particular example.

![Causal Map](image)

**Figure 1. Causal Map**

The nodes of the graph, designated by capital letters, represent variables whose values are properties of the system to which the net applies.\(^7\) A particular capital letter might designate *color*, for example, and it might be a variable with several

---


\(^7\) Both the figure and the following description of the figure are taken (with some modifications) from Gopnik and Glymour (2002, pp. 126-27).
values or with only two (red and green, say). Alternatively, a capital letter might
designate weight, and this might be a variable with two values, heavy and light,
or it might be a variable with a continuum of values.\(^8\)

With respect to the arrows, finally, an arrow connecting one node or
variable to another—Z to A, for example—signifies that an intervention that varies
the value of Z, but does not otherwise alter the relationships among the variables,
will change the value of A. More briefly, that changing the value of Z will (ceteris
paribus) lead to a change in A.

Applying the map to a particular example should help to clarify things.
Suppose that what we are trying to construct is a map that captures the
dependencies among the various elements relevant to local storm activity, and
that the previous map is in fact the result of such a process. Specifically, suppose
that we take A to be a variable designating atmospheric pressure, and that we
assign it two possible values: rising or falling. Similarly, suppose we take S to be
a variable designating storm, with the two possible values present or absent.
Finally, suppose that B represents the barometer reading, and that it likewise
has two values: rising or falling.

What the previous map tells us is that, given a change in the value of A,
even though a corresponding change in the value of B will be correlated with a
corresponding change in the value of S, the value of S does not actually depend on
the value of B. More exactly, what the map tells us is that if we were to intervene

\(^8\) The fact that there is some openness in how we might specify the range of possible
values reflects the openness we are familiar with when we pose our explanation-seeking why
questions. For example, I might ask why a certain box is heavy rather than light, or I might ask
why it weighs 10 pounds as opposed to some other number.
to change the barometer reading—if we were to make the setting fall with our own hands, for example—this change would not lead to a change in the value of A or S.

The general idea is that in our interaction with the world one of our cognitive system (or systems) has the task of constantly constructing and reconstructing maps of this kind as it tries to identify genuine dependencies among things in the world. Put a little differently, through testing and observation the system is constantly trying to discriminate correlations that reveal actual dependencies from correlations that fail to reveal actual dependencies—either because the latter correlations are sheerly accidental, or because they are due to the influence of a common cause, or what have you. Of course, to speak of “the system” doing all this is a bit artificial; it is really we who are doing all this, by means of the system. But one benefit of speaking in this artificial way is that it underscores how below-the-radar much of this discrimination and evaluation is.

So how reliable is this system—are we—in terms of constructing accurate causal maps? What clinical studies on young children suggest is that the system is, in fact, quite reliable. As described by Gopnik and Glymour (2002):

In several studies, we have begun to show that children as young as two years old, in fact, do swiftly and accurately learn new causal relations—they create new causal maps. They do so even when they have not themselves intervened to bring about an effect, and when they could not have known about the relation through an evolutionarily determined module or through prior knowledge. We present children with a machine, the ‘blicket detector,’ that lights up and plays music when some objects but not others are placed in it. Children observe the contingencies between the objects and the effects and have to infer which objects are ‘blickets.’ That is, they have to discover which objects have this new causal power. Children as young as two years old swiftly and accurately make these
inferences. They identify which objects are blickets and understand their causal powers.... [E]ven very young children use a form of ‘screening off’ reasoning to solve these problems. (p. 131)\(^9\)

In other words, what studies suggest is that not only do we have a special cognitive system dedicated to constructing causal maps of the world around us that is apparently hard-wired (given our ability to construct these maps from a very young age),\(^10\) but that we also have a natural aptitude for the task. Constructing accurate causal maps is the kind of thing, apparently, that we are naturally good at doing.

Alongside Gopnik’s research with children, Mark Steyvers and his colleagues (2003) have recently reproduced similar findings with adults. In order to weed out potential differences in background information (a point that will reappear as significant later), Steyvers designed an experiment in which subjects were asked to identify causal relationships among three hypothetical “alien mind-readers.” At the beginning of the experiment subjects were shown a picture displaying the thoughts of three aliens (of the Hollywood variety), and they were asked to identify which alien was reading the minds of the others.\(^11\) In one of the experiments, subjects simply observed how changes in the thought bubble of one alien either did or did not change along with the thought bubbles of the other aliens. In another experiment, subjects were allowed to intervene in


\(^10\) Or, more broadly, to identify dependencies in general (including, for example, mereological dependencies).

\(^11\) The alien “thoughts” were displayed in thought bubbles, and were quite basic. Thus in one experiment the first alien was thinking “TUS,” the second “POR,” and the third “POR.”
order to change, via manipulation, the alien thoughts themselves, in order to see how a change in one would affect the others.

What Steyvers found was not only that subjects were quite good at identifying dependencies based on passive observation, but that when they were allowed to manipulate the interactions their reliability increased dramatically. As he summarizes:

Faced with the challenge of inferring the structure of a complex causal network, and given no prior expectations of what causes what, people bring to bear inferential techniques not so different from those common in scientific practice. Given only passive observational data, people attempt to infer a system’s underlying structure by comparing what data they see to what they would expect to see most typically under alternative causal hypotheses. Given the opportunity to learn actively from observing the effects of their own interventions on the system, they make more accurate inferences. (p. 486)

Combined with Gopnik’s work, what both studies therefore suggest is that the very methodology that a mature science uses to identify dependencies in the world—of searching for correlations in the data, of using control groups, and so on—is in some sense the built-in method that our grasping faculty uses, that we use, to identify dependencies. Given that we think the scientific method is a reliable way to identify dependencies in the world, it should therefore come as no surprise that our grasping faculty turns out to be as reliable as it is.

5.3 Corruptions

One reasonable conclusion to draw from these studies is therefore that, contra Trout, our grasping faculty simply is reliable at identifying dependencies—at least characteristically and under normal conditions. We might think of this as the “optimistic conclusion.”
In the remainder of this chapter I will argue that the optimistic conclusion is misleading; if it can be sustained at all, it needs to be qualified significantly. What the optimistic conclusion suggests is that even though our grasping faculty is susceptible to unreliability (the astrologers and the alchemists have to be accounted for, after all), the unreliability is the result of some sort of external corruption rather than the result of a problem inherent in the normal operation of the system—just as, we might think, although our visual system is susceptible to unreliability via external corruption (cataracts, mind-bending drugs, etc.), in the absence of such external corruption it is basically reliable. To put the point another way, what the optimistic conclusion suggests is that when our grasping faculty is functioning properly in its normal environment, it will be largely reliable, just as when our visual system is functioning properly in its normal environment, it too will be largely reliable.

What I want to propose instead is that even a properly functioning grasping faculty in its normal environment might be thoroughly unreliable at identifying dependencies. In addition to its intrinsic interest, one further reason why, if true, this would be an interesting result is that it would help to explain why eminent thinkers such as Ptolemy and Galen did identify and affirm extremely strange dependencies, in a way that would not have to attribute some sort of malfunction in their grasping faculty to them.

In the remainder of this section I will consider two reasons—or, as I will put it below, “types” of reasons—why even a properly functioning grasping faculty might systematically produce bad results.
5.4 Type I: Intellectual Vices

Drawing on recent work by Christopher Hookway (2001, 2003), the basic idea behind Type I sources of unreliability is that even in normal environments, and even when the information we have at our disposal is quite good, the influence of so-called “intellectual vices” can significantly diminish the reliability of our grasping faculty.\(^\text{12}\)

By way of illustration, suppose that Lisa is tracking the progress of 100 cold-sufferers who are trying a new product, Cold-Away. As it happens, all 100 of the sufferers recover from their cold within a week.\(^\text{13}\) Based on the strong correlation between the taking of the Cold-Away and the timely relief from their colds, Lisa concludes that it was the Cold Away that made the difference to the group’s improvement.

This will be a bad inference, naturally, and the reason is the lack of anything resembling a control condition. At a minimum, in order to determine whether it was the Cold-Away that made a difference to their recovery, Lisa should have checked to see how cold-sufferers who did not take Cold-Away fared. In this case, what she would have found is that even those who failed to take Cold-Away—in other words, even those who took nothing—would have likewise

\(^{12}\) Several epistemologists, especially those working in the virtue epistemology tradition, have stressed the connection between the intellectual vices and unreliability. See, e.g., Zagzebski (1996), Riggs (forthcoming), and Baehr (forthcoming).

\(^{13}\) The basic example is from Salmon (1970, sec. 8); see also Bishop and Trout (2005, pp. 147-48).
recovered within a week. Colds are simply the kinds of things that, given a little
time, run their course. But then, presumably, it was not the Cold-Away that
made the difference after all.

Crucially, however, the source of the failure in this case does not seem to
be Lisa’s grasping faculty—it is not as if, for instance, her grasping faculty was
misled by the equivalent of cataracts or occlusions—but rather Lisa’s handling of
her grasping faculty. To draw again on an analogy with the visual system, the
more time you give a properly functioning visual system to take in the world, the
more likely it will be to represent the world accurately. By contrast, the more you
rely on quick glimpses, etc., the less likely it will be to represent the world
accurately—hence the less likely that beliefs formed on the basis of such
information will be accurate. In the same way, however, the more information
Lisa feeds into her grasping faculty—in the form of wide and varied samples, in
the form of control groups, etc.—the more likely it will be that she identifies
genuine dependencies. By contrast, the more she fails to provide her grasping
faculty with this information—the more she jumps to conclusions based on a
small amount of data, for example—the less likely it will be to identify genuine
dependencies.

The way I have characterized the source of Lisa’s unreliability is in terms
of a kind of intellectual carelessness, perhaps even a kind of intellectual laziness:
she jumped to conclusions without bothering to acquire all of the relevant data.
Needless to say, however, there are dozens of other intellectual faults that are

---

14 In reality, of course, it would be very hard to pin the blame on the character flaw rather
than the faculty; for all we know she might have some cataract-equivalent problem with her
faculty. In this case, though, we can simply stipulate that she doesn’t.
likewise capable of contributing to the unreliability of someone’s grasping faculty. For example, George might be someone with a wide and varied set of information at his disposal, but he might nevertheless be guided by his prudential interests (his interest in “seeing” certain patterns, for whatever reason) to emphasize certain parts of the data and to de-emphasize other parts. Or again, from intellectual arrogance Ralph might fail to consider alternative hypotheses, or might reject as faulty data that fails to support his own hypothesis. In all of these cases, however, the problem will lie not with the faculty itself but rather with someone’s careless (or perhaps simply incompetent) handling of the faculty.

### 5.5 Type II: Bad Background Beliefs

Another reason—perhaps the more likely reason—why even those with properly functioning grasping faculties might systematically go wrong is because of bad (i.e., false) background beliefs. Even when everything else is ideal—even when, for example, one has the benefit of a rich and varied amount of data, even when one pours through the data in the most careful and attentive way—one’s background beliefs will still have an enormous influence on the sorts of possibilities that one takes seriously in the first place.\(^{15}\)

The following case helps bring this out. Suppose that Albert is an atheist and Paul a theist, and that both are wondering why a certain group of patients, Group A, recovered from their illnesses while another group, Group B, failed to

\(^{15}\) A point that is, in another form, familiar to us from our discussion of proto-understanding in Chapter 3.
recover. Albert and Paul subsequently learn that all of the patients in Group A were being prayed for by a group of pious nuns several thousand miles away and that none of the patients in Group B were being prayed for by the nuns. Moreover, they learn that the nuns’ prayers for the patients in Group A were in fact part of rigorous, double-blind experiment. None of the patients in either group were aware that they were being prayed for, and other possible countervailing influences were carefully screened out.

As an atheist who believes that petitionary prayers are useless, learning about the results of the study leads to no graspings or seeings for Albert—no sense that the prayer-healing connection “feels right.” By his lights, the thing that is identified as a difference maker here—petitionary prayer—simply isn’t qualified to be a difference maker. He therefore comes to think not that the recovery of the patients in Group A depended on the nuns’ petitionary prayers, but rather than there was something wrong with the study.

Paul, by contrast, not only believes in God but likewise believes in the power of petitionary prayer. While he may not immediately accept that it was the nuns’ prayers that made the difference between Group A and Group B, he will

---

16 This case is loosely based on real life. In 1999, the American Medical Association’s Archives of Internal Medicine published a better-designed study of nearly a thousand consecutive patients who were newly admitted to the coronary care unit of a hospital in Kansas City. The researchers created a 35-item score sheet that was used to measure what happened to the patients during a 28-day period in which 15 groups of 5 persons (“intercessors”) prayed individually for about half the patients. The intercessors were given the patients' first names and were asked to pray daily for "a speedy recovery with no complications." The prayed-for group had a 10-11% reduction in total scores even though their average length of hospital stay was similar to that of the "usual-care" group. For further the details see Harris et al. (1999).

17 At least, not from a distance; the fact that patients who pray for their own health tend to improve as a result has been widely documented, and is now generally accepted. Whether the improvement is due to the patient’s sense of optimism, or whether instead to divine intervention, is an open question.
nonetheless find the results worth taking seriously. If he learns of further studies that also indicate the existence of such a correlation, moreover, he will then likely conclude that it was, in fact, the prayers that were making the difference on these occasions.

Both Albert and Paul are therefore interestingly like Ptolemy in one respect at least. Presumably Ptolemy’s belief that it was the earth that stood still and that it was the sun that moved was so well entrenched (after all, just take a look! which one is doing the moving?) that this belief automatically ruled out any explanatory story that portrayed the earth as a mover. What’s more, one of the two—either Albert or Paul—is also like Ptolemy in the following sense: not only is one’s belief about God false (as Ptolemy’s geocentric belief was false), but it will be false in a way that negatively effects many of the other conclusions he draws. If God doesn’t exist, then Paul will be identifying countless dependencies that simply aren’t there (petitionary prayer being only one of them), not to mention (perhaps) overlooking the dependencies that are there. If God does exist, however, then it will be Albert who will be missing out on countless dependencies: he will be failing to grasp what actually makes the difference in all of these cases.¹⁸

¹⁸ Naturally, the unreliability of someone’s grasping faculty can be attributable to more than one source. Thus someone might be unreliable at identifying dependencies both because he is intellectually careless and because he has bad background beliefs. If Andy has been brainwashed into believing that the CIA is behind all the suspicious activity in the world, this will become part of his background beliefs and, as per Type IV cases, it will influence the kinds of explanatory stories that he finds worth considering. If he has good intellectual habits, however, in virtue of which he took all the data seriously, considered alternative hypotheses, etc, then eventually his belief about the CIA should begin to lose credibility as he comes to observe that countless suspicious (or apparently suspicious) things happen for reasons other than the intervention of the CIA. If he has bad intellectual habits, however, he will focus on the data that
5.6 Conditional Reliability

Bearing all this in mind, to the question, “Is our grasping faculty reliable?,” our answer should therefore be: It depends. More precisely, as I will now suggest, our answer should be: The faculty is conditionally reliable, and this in several respects.¹⁹

One respect in which the faculty is conditionally reliable is clearcut: if you get bad data in you will almost inevitably get a bad grasping out. Specifically, if the information that you are mining is full of inaccuracies then the dependencies you identify on the basis of the information will likewise be inaccurate. As we saw with respect to Type II cases, this is also true in a more subtle sense: if you have bad background beliefs then scores of accurate explanatory stories will not even receive consideration from you. In both of these cases, however, it would be a mistake to say that the problem lies with the faculty per se. Instead, the problem has something to do with the quality of the information that is being fed into the faculty.

With this in mind, moreover, when we look back to the Gopnik-Steyvers experiments from Section 5.2 we can now appreciate one of the reasons why the subjects in the experiment, first the children reasoning about blickets and then

doesn’t rule out of the intervention of the CIA and thus convince himself that—in this case too—the CIA is responsible.

¹⁹ In Alvin’s Goldman ([1976] 2000) words, “A process is conditionally reliable when a sufficient proportion of its output beliefs are true given that its input beliefs are true” (p. 347). Deductive reasoning therefore counts as an example of a conditionally reliable process, for example, in the sense that if you start with garbage (falsehoods) you will end up with garbage (more falsehoods), whereas if you start with truth you will end up with truth.
the adults reasoning about alien-mind reading, performed so uniformly well. For notice: *in both experiments the problem of false background beliefs was essentially factored out*. For the children, the background beliefs were doubly factored out: first, because children lack a large stock of background beliefs to begin with, and second because the novelty of the blickets experiment required them to approach the situation without any preconceptions. With the adults, the peculiar alien-mind reading experiment was likewise chosen precisely because of its novelty. Lacking any preconceived notions of alien mind-reading skills, the adults were able to evaluate the relationships among the variables with few if any of the preconceptions that usually shape our view of the world.

We can also now appreciate how our grasping faculty is conditionally reliable in a still further respect: namely, the reliability of our grasping faculty depends not just on the quality of our information but also on how the information is *used*. Intellectual character is relevant to reliability because, as Type I cases show, even if one’s background beliefs are true and the information one has about a particular situation is large and varied, there is always a temptation to train one’s focus on some parts of the data rather than others, either out of a vain desire to protect one’s own views or out of a prudential desire to find certain patterns rather than others.\(^{20}\)

\(^{20}\) The opposite sort of person—namely, someone with a *good*, and therefore reliable, cognitive character—will therefore be someone who takes all of the data seriously, does not restrict their attention to a limited range of information that might favor a certain pattern, does not rush to conclusions from small sample sizes, and so on. In short, someone with a good cognitive character will be someone who has essentially internalized the scientific method! What’s more, as Alison Gopnik (1998) notes in the following passage, even if someone has not internalized the method in this way, one might still use the method as a kind of external corrective to compensate for the unreliable tendencies in one’s intellectual character, much as one corrects the results of an unreliable visual system by using glasses. Gopnik writes: “It appears that one of
5.7 Coherence Renewed

Our grasping faculty will therefore be reliable to the extent that our background beliefs are good, to the extent that we have good, varied data, and to the extent that we have a good cognitive character that uses this data appropriately.

Bearing in mind how all of this impacts on the reliability of this faculty, we can now make sense of one final point about the nature of our graspings or seeings—and thus about the nature of understanding—that several commentators, including Jonathan Kvanvig (2003), Wayne Riggs (2003), and Catherine Elgin (1996, 2004) have stressed: namely, that coherence considerations have a significant role to play in the exercise of our grasping faculty.21

the differences, perhaps the most important cognitive difference, between organized science and spontaneous theory formation is precisely that science contains additional normative devices that are designed to supplement the basic cognitive devices of the theory formation system, and to protect them from error. We might think of science as a kind of cognitive optometry, a system that takes the devices we usually use to obtain a veridical picture of the world and corrects the flaws and distortions of those devices. The fact that most people over forty wear glasses is not, however, usually taken as an indictment of the visual system. In fact, the analogy might be taken even further, perhaps science compensates for our deteriorating adult theory formation abilities the way optometry compensates for our deteriorating adult vision” (p. 113; my emphasis).

21 So too, to a lesser extent, with Ernest Sosa ([1997] 2000): “[B]eyond ‘animal knowledge’ there is a better knowledge. This reflective knowledge does require broad coherence, including one’s ability to place one’s first-level knowledge in epistemic perspective. But why aspire to any such thing? What is so desirable, epistemically, about broad coherence? Broad coherence is desirable because it yields integrated understanding, and also because it is truth-conducive, even if in a demon world broad coherence fails this test and is not truth conducive” (p. 280; my emphasis). If I understand Sosa correctly, the suggestion is therefore that understanding (or at least “integrated understanding,” if that is a distinct type of understanding) comes from seeing, in some sense, how one’s beliefs cohere with one another. It also bears pointing out that Kvanvig et al. might be reluctant to speak of a grasping “faculty,” as we have done here; but I think the basic point remains the same.
According to Riggs, for example, understanding a particular bit of information is essentially, perhaps even exclusively, a matter of seeing how it “fits together” or “coheres with” the rest of one’s beliefs. Riggs writes,

An important difference between merely believing a bunch of true statements within subject matter M, and having understanding of M (or some part of M), is that one somehow sees the way things ‘fit together.’ There is a pattern discerned within all the individual bits of information or knowledge.... The epistemological notion of ‘coherence’ and the idea of ‘explanatory coherence’ in particular seems to be getting very close to something characteristic of understanding. (p. 218)

According to Kvanvig too:

The central feature of understanding, it seems to me, is in the neighborhood of what internalist coherence theories say about justification. Understanding requires the grasping of explanatory and other coherence-making relationships in a large and comprehensive body of information. One can know many unrelated pieces of information, but understanding is achieved only when informational items are pieced together by the subject in question. (p. 192)

Later Kvanvig expands on this claim:

I want to focus on this crucial difference between knowledge and understanding: that understanding requires, and knowledge does not, an internal grasping or appreciation 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. (pp. 192-193)

I do not want to endorse—in fact, in the following chapter I actively dispute—the suggestion here that understanding is fundamentally an internal matter, a matter of grasping how one’s beliefs cohere with one another. That said, we can now see the sense in which coherence considerations do have a crucial role to play in the exercise of our grasping faculty. For one thing, as we mine our information about the world in a search for dependencies, our mining is necessarily influenced by our background beliefs. As we have seen, if you are
convinced that it is the sun that moves and the earth that is still, then when you are trying to explain (say) stellar parallax, you will focus on explanatory stories that don’t portray the earth as a mover and rule out the rest. Or again, if you have been brainwashed into believing that the CIA is responsible for everything suspicious in the world, then when you wake up one morning to find your tapwater tasting funny, you will take seriously explanatory stories that involve the CIA and disregard the rest. Our background beliefs therefore constitute a kind of first filter (cf. Lipton 2004, ch. 3), and possible dependencies do not even get a hearing at this stage if they fail to pass the test.

After the first (possible/impossible) filter, moreover, we evaluate possible dependencies in terms of how well they fit with the correlations that are already familiar to us, either via testimony or via direct observation. To appeal to a simple example, if someone suggests that your lawn is browning-out because of the absence of rain, but you believe it rained last night, you will fail to accept the alleged dependency. Likewise, if you think the lawn often browns out this time of year because of grubs, then even if you agree that it hasn’t rained recently you will likewise fail to accept the alleged dependency: instead, you will think the matter deserves more looking into.

Whether or not we accept that a certain dependency exists will therefore be a matter of how well it makes sense of the rest of what we believe. If our background beliefs are accurate, however, if we have a rich and varied amount of information to sift through (whether consciously or not), and if we have good intellectual habits, then, finally, we have reason to be optimistic: our chances of identifying genuine dependencies (with any luck) will be very good.
CHAPTER 6

IS UNDERSTANDING A SPECIES OF KNOWLEDGE?

Among philosophers of science there is a general agreement that understanding represents a *species of knowledge*. Thus, for instance, we find Peter Lipton (2004a) arguing that a good theory of understanding should characterize the topic in a way which is “unmysterious and objective,” and that the first step towards reaching this goal is to see that, “Understanding is not some sort of super-knowledge, but simply more knowledge: knowledge of causes” (p. 30; cf. Lipton 2004b). Peter Achinstein (1983) likewise writes that, “Explaining q has been defined as uttering something with the intention of rendering q understandable.... Such understanding I take to be a form of knowledge” (p. 23).\(^1\) Of course, as these quotes effectively remind us, philosophers of science have hardly agreed on just which sort of knowledge it is that characterizes understanding: whether understanding why P comes from knowing the cause of P, or from knowing that P can be subsumed under laws, or what have you, is still

\(^1\) Other philosophers of science who seem to assume, essentially without argument, that understanding is a species of knowledge include Wesley Salmon (1989, pp. 134-35), James Woodward (2003, p. 179), and Philip Kitcher (2002, *passim*).
widely debated. But that understanding is a kind of knowledge is something that—so far at least—has rarely been doubted by those working in the field.2

That makes it all the more striking that virtually every major epistemologist who has thought seriously about the nature of understanding—including Catherine Elgin (1996, 2004), Linda Zagzebski (2001), and Jonathan Kvanvig (2003)—has come to the conclusion that understanding is not a species of knowledge.3 They variously claim, for example, that understanding is immune to Gettier problems whereas knowledge is not, that understanding is transparent whereas knowledge is not, and that understanding is possible even in the absence of truth, whereas this is an impossibility for knowledge. All in all, they argue, understanding and knowledge are simply pulling in too many different directions for the former to be thought of as a species of the latter.

In this chapter I will evaluate the case against taking understanding to be a species of knowledge by focusing specifically on the accounts offered by Zagzebski and Kvanvig, and I will argue against this growing consensus in epistemology that—just like knowledge—understanding requires truth, is not transparent, and can be Gettiered. In short, that understanding certainly seems to behave a lot like a species of knowledge. To show that the current arguments in favor of the claim are wanting, however, is not the same as to show that the

---

2 This is to set to one side issues raised by those who, like J. D. Trout (2002), claim that understanding is not really an epistemic good at all but rather merely a “subjective feeling.” I will not be dealing with that sort of understanding skepticism in this chapter.

3 And, for that matter, not just epistemologists. According to John Haugeland (1998): “Understanding is always “of” something—objects, in a broad sense—but this of-ness is not the same as that of beliefs or desires. Thus, understanding is not the same as knowledge, a special kind of knowledge, or even a complex structure or totality of knowledge” (p. 1).
claim is false, and later in the chapter I will take up the question of what, on
balance, we should say about this subject.

6.1 Zagzebski

According to Zagzebski (2001), one of the most striking differences
between knowledge and understanding is that while understanding is
*transparent*, knowledge is not. On her view, understanding is fundamentally a
matter of grasping how various pieces of information relate to one another; it is a
matter of making connections among them, of seeing how they hang together.
But if that’s the case, she argues, then it is natural to think that the *object* of
understanding—apparently, the connection that we “see”—must be internal.
After all, how can one “grasp” or “see” something that isn’t right there, open to
our mental view?

In this as well as other respects, however, Zagzebski claims that
understanding differs from knowledge. As she writes:

> Understanding has internalist conditions for success, whereas knowledge
does not. Even when knowledge is defined as justified true belief and
justification is construed internalistically, the truth condition for
knowledge makes it fundamentally a concept whose application cannot be
demonstrated from the inside. Understanding, in contrast, not only has
internally accessible criteria, but it is a state that is constituted by a type of
conscious transparency. It may be possible to know without knowing that
one knows, but it is impossible to understand without understanding that
one understands.... [U]nderstanding is a state in which I am directly
aware of the object of my understanding, and conscious transparency is a
criterion for understanding. Those beleaguered by skeptical doubts
therefore can be more confident of the trustworthiness of putative
understanding states than virtually any other state. (pp. 246-47)

It is not hard to see what Zagzebski is driving at. For one thing, we can grant that
there is some sort of connection that is immediately open to view when we take
ourselves to understand; it seems to be a hallmark of having understanding that we can typically articulate (or explain) what it is in virtue of which we take ourselves to understand, for instance, and that kind of articulacy might be thought to require a direct apprehension of the thing understood. For another, it is plausible that there are some types of understanding that do allow for the kind of transparency Zagzebski has in mind. Our understanding of concepts might fall into this category.

But to grant that some components of understanding are transparent in this way, or even that some types of understanding might allow for this kind of transparency, is a long way from granting that understanding in general allows for the kind of transparency described by Zagzebski. In particular, it is a long way from granting that our understanding of natural phenomena—arguably the paradigm case of understanding—allows for this kind of complete transparency.

A basic example helps to illustrate the point. Suppose that you open the refrigerator and notice that the light has gone off, so you put your hand inside and feel around, finding it warm to the touch. While looking around for an explanation eventually you notice something unusual: the cord has been unplugged. You now take yourself to understand why the refrigerator stopped working, but it should be obvious that you might be mistaken. It might have broken down as the result of a short circuit, and the cord might have been unplugged for fear of fire. In that case, despite your sense that you understand why the refrigerator stopped working, in fact you don’t understand. You’ve mischaracterized how things stand in the world.
In keeping with the lesson from the end of Chapter 4, the basic problem with Zagzebski’s account is therefore that our understanding of natural phenomena is surely factive—what we are trying to grasp is how things actually stand in the world, and there is no reason to think that how things stand in the world is consciously transparent to us. To this it might objected that even when we miss out on understanding in this way due to some failure of fit with the world, there is nonetheless a desirable form of understanding that we can nonetheless lay claim to. In the case of the refrigerator, for instance, even though you fail to understand how things stand with respect to this particular situation, you still seem to understand something else of value: roughly, that refrigerators depend on electricity, and in the absence of a source for electricity, they stop working. But generalizations of this sort are themselves substantive claims about how things stand in the world; they claim to describe real physical dependencies. And, again, there is no reason to think that we have any kind of transparent access to whether or not such dependencies actually obtain.  

6.2 Kvanvig

Kvanvig’s (2003) account of understanding is therefore an improvement on Zagzebski’s, in this respect at least, because he insists on the factive character of understanding. For Kvanvig, what this means is that in order for a mental state to count as a state of understanding, the beliefs that constitute the state

---

4 Zagzebski has more to say here, about how understanding is internal to (what she calls) practices, and about how practices allow us to identify and correct mistakes along these lines. I will not address this further claim here, beyond pointing out that there are some practices—astrology, voodoo, etc.—which would confirm very strange dependencies indeed, and that from the inside these might well have all of the same conscious marks as more conventional practices (such as, to cut things thickly, contemporary science).
must be true. As he sees things, however, what is important and distinctive about understanding is entirely a matter of what one makes of one’s true beliefs, once one has them. Specifically, Kvanvig insists along with Zagzebski that understanding is fundamentally a matter of grasping or seeing connections among one’s beliefs. But on his view what this entails is that, so long as the beliefs happen to be true, facts about how one’s beliefs were acquired—in Kvanvig’s terms, external facts about the etiology of one’s beliefs—are irrelevant when it comes to assessing whether or not one understands.

Kvanvig argues for this claim by means of the following example (pp. 197-98). Suppose you pick up a textbook on Native American History and read through a chapter documenting the Comanche dominance of the southern plains, until eventually you seem genuinely to understand why the Comanches dominated the southern plains. But suppose as well that while the book you happen to pick up is accurate, most other books on this topic are full of errors. If you had picked up one of these other books instead (and we can imagine that they are all within easy reach!), your beliefs about the Comanches would have been almost entirely false.

What then should we say about the status of your Comanche beliefs? Do they amount to knowledge? Many would say that they fail to amount to knowledge because the fact that your beliefs about the Comanches hit the truth seems to be a mere accident. Unfortunately for you, you have stumbled into a Gettier environment, where a bit of good epistemic luck (happening by chance upon accurate information) has canceled out a bit of bad epistemic luck (being in an environment full of misinformation).
But now, as Kvanvig notes, we still have a further crucial question to consider: Once we see how easily your beliefs might have been false, should we take the same negative view towards your (alleged) understanding? Should we say that it fails to amount to a case of genuine understanding? According to Kvanvig, the answer is No. The specific story about how one’s beliefs came to be true—whether as a result of accident or luck or some other means—has no impact on their ability to contribute to understanding. Once the truth condition has been met, understanding is a purely internal matter. He sums up these claims as follows:

Understanding does not advert to the etiological aspects which can be crucial for knowledge. What is distinctive about understanding, once we have satisfied the truth requirement, is internal to cognition. It is the internal seeing or appreciating of explanatory relationships in a body of information which is crucial to understanding. When we think about knowledge, however, our focus turns elsewhere immediately, if we have learned our lessons from the Gettier literature: we think about the possibility of fortuitousness, of accidentality, of being right only by chance. We focus, that is, on what kind of further external connections there are between mind and world, beyond the fit required for the belief to be true.

The basic idea here is that, though knowledge is incompatible with a certain kind of epistemic luck, understanding is not. Upon learning of the disturbed etiology of beliefs about the Comanches, as in the case imagined here, we might say that the person has true beliefs or even true justified beliefs, but no knowledge, if we have heeded our lessons from Gettier. We would not, at least we should not, say that because of these factors, she is lucky to have the knowledge she has, for knowledge rules out this kind of luck. But we needn’t say the same thing about the claim of understanding. If the etiology were as imagined, one would be lucky to have any understanding at all of the Comanche dominance of the southern plains. So such understanding would count as understanding not undermined by the kind of luck in question. (pp. 198-99)

Understanding is therefore a paradigm instance of a luck-proof cognitive state.

And since knowledge is essentially vulnerable to luck, we can conclude that understanding is not a species of knowledge.
In the remainder of this chapter I will try to show that this general way of thinking about understanding—what we might call “understanding internalism”—is misguided. Not only is understanding susceptible to luck, I will argue, but etiology clearly matters when it comes to understanding. This will then leave us with the following question: What should we make of Kvanvig’s Comanche case? Specifically, has he really shown that there can be understanding in the absence of knowledge? I will argue that the Comanche case shows none of these things. As we’ll see, it matters how we unpack the details of this kind of case. But on any way of filling in the details knowledge and understanding seem to sway together.

Finally, I will ask whether we can conclude from this that understanding is, after all, a species of knowledge. At this point the chapter takes a new turn: There appears to be independent reason, apart from the considerations Kvanvig offers, to think that understanding in fact is not a species of knowledge. In Sections 6.11 and 6.12 I explore this suggestion at greater length.

Beginning with the claim about luck, there are at least three different genres of counterexamples to understanding internalism, all of which help to show how etiology matters to understanding. We can call the first the **bad inductive base problem**, the second the **veridical hallucination problem**, and the third the **bad environment problem**.

### 6.3 Genre I: The Bad Inductive Base

Suppose you take yourself to understand why Sam suddenly died: because he ingested XX. And suppose that you take yourself to understand this based on
your belief that XX is lethal, a belief grounded in turn on your first-hand experience of several people ingesting XX and dying straight away. As it happens, however, suppose that (a) XX truly is a lethal substance, but (b) all the prior deaths you observed were improbably caused instead by something else simultaneous with the victim's ingestion of XX: a heart attack in one case, an aneurysm in another, etc. Finally, add (c): that Sam's death—the death you take yourself to understand—was indeed caused by his ingestion of XX.

Given this fleshed-out description of the case, should we say that you really understand why Sam died? I take it that the answer is No. Your beliefs about the cause of Sam's death are accurate, sure enough, and hence they satisfy Kvanvig's truth requirement: XX really is a lethal substance, and it really was the ingestion of XX that did him in. But the fact that your belief that XX is a lethal substance is true is more or less a matter of luck; contrary to appearances, your inductive evidence does not in fact support your belief that XX is a lethal substance.

One formula for counterexamples to Kvanvig’s account is therefore straightforward: find a subject whose inductive base seems to support the true generalization that P, though only due to luck or coincidence, and then allow this generalization to play a central role in the subject’s explanatory account.5

5 One might object that a problem with using this formula is that it illicitly smuggles in false beliefs, and that this goes against Kvanvig’s truth condition: for example, it smuggles in the false belief that the first victim died because he ingested XX, that the second victim died because he ingested XX, and so on. But we can simply stipulate that you did not have any beliefs to this effect, never mind false ones: in other words, we can stipulate that you (truly) believed that the first victim ingested XX and died straight away, and likewise for the second, etc., and then formed the (true) belief that XX is lethal on the basis of this evidence.
6.4 Genre II: The Veridical Hallucination

Suppose that the CIA slips a hallucinogen into Albert’s coffee, and that as a result he “sees” his dog bump into the table, causing a vase to crash to the floor. Putting things together, he then takes himself to understand why the vase fell from the table and crashed to the floor: because the table was bumped by the dog. As it happens, his hallucination exactly matches the events that are actually unfolding in front of him.

When we add the pieces up, we again seem to have all the elements for understanding that Kvanvig requires. There is the true belief that the vase fell from the table, as well as the true belief that the stand was bumped by the dog. There is, in addition, the genuine dependence that holds between the two: that the vase fell because the table was bumped by the dog. But still, we’re tempted to say, Albert doesn’t really understand why the vase fell—specifically, he doesn’t understand that the vase fell because the table was bumped by the dog. And the reason is that he might too easily have misidentified the cause of the fall: if the drug had caused him to hallucinate something else, had it portrayed his cat bumping into the table instead of his dog, then he would have believed that the vase fell because the table was struck by the cat.

6.5 Genre III: The Bad Environment

The final genre of counterexamples is modeled on one of Mackie’s (1980) well-known examples. Suppose that while wandering through a blacksmith’s shop Becky notices a chestnut lying on top of an anvil, and she pauses to watch as

---

6 Thanks to John Turri for the basic example.
the blacksmith moves to strike the chestnut with his hammer. At the very instance the hammer touches the chestnut, the chestnut explodes into fragments.

Again, the details matter. Suppose first, then, that as a rule the blacksmith heats the anvil to extremely high temperatures, so that after a certain amount of time the chestnuts placed on the anvil eventually explode from heat stress. Moreover, imagine that the blacksmith enjoys testing his timing so that his hammer grazes the top of the chestnut at the precise moment that the chestnut is due to explode.

In this case, however, as Becky is walking by, things don’t unfold normally. Either because he was in the mood for a change or simply because he forgot to heat up the anvil, as the hammer makes contact with the chestnut it is not on the verge of exploding from heat stress; so it is, in fact, the force of the hammer blow that shatters the chestnut.

Watching all this, but unfamiliar with the blacksmith’s usual routine, Becky concludes that the chestnut shattered because of the blow from the blacksmith’s hammer. Does she genuinely understand this? Once again, the answer seems to be No. To clarify why, suppose that Becky lingers for a while and watches as the blacksmith performs his chestnut trick in the usual way. In each case, the blacksmith times his swing perfectly so that the hammer impacts the chestnut just as the nut is about to explode. The first time it happens, we can imagine Becky thinking to herself: “That chestnut just shattered because of the impact of the hammer.” And so on for the second time, and the third time, etc.

Indeed, we can imagine that all the blacksmiths in the area perform the same
trick, so that time after time the chestnut explodes precisely upon impact by the hammer.

If we are reluctant to ascribe understanding to Becky on the lone occasion she gets it right, moreover, I suggest that this is for precisely the same reason that we are reluctant to ascribe knowledge to someone driving through fake barn country on the lone occasion when they happen to spot a real barn. For example, if while driving through fake barn country Becky pauses repeatedly to say “That’s a mighty fine barn!,” then when she comes across the lone real barn in the area, we are reluctant to say that she know this is a barn. Knowledge, as Nozick observed (1981, p. 213), should be made of sterner stuff than that—not so easily should the knower be capable of mistake. But now, in light of the previous example, the same seems to hold true of understanding. If Becky would have failed to identify the genuine cause of the shattering 99 out of a 100 times in the blacksmith shop, the natural temptation is to say that, on the lone occasion when she gets things right, she fails to understand. Like knowledge, understanding too should be made of sterner stuff.

6.6 Comanche Cases

I think these examples show that etiology matters to understanding, and thus that there is good reason to think that it is not a purely internal affair. It is important to see, however, that Kvanvig can accept all this and still have his fundamental claim go through: namely, that understanding is not a species of

---

7 Where by “fake barn country” I mean the kind of place famously described by Alvin Goldman in his (1976): namely, a place where the majority of barn-like structures in the area are in fact barn facades, set up to make travelers think that the area is more prosperous than it actually is.
knowledge. All he needs to show in order to establish *that*, after all, is that there are cases where understanding and knowledge can come apart, specifically, that we can have cases of understanding that are not cases of knowledge. And Kvanvig claims to have produced a case of this sort: namely, where someone understands all about the Comanches even though, because of the chancy way they came to believe the relevant information, none of the relevant beliefs amount to knowledge.

Nothing special hangs on the fact that the example concerns the Comanches *per se*, of course, but for convenience we can hereafter refer to the broader type to which this case belongs as a “Comanche-style case” and characterize it as follows. A Comanche-style case is one in which we form true beliefs on the basis of trusting some source, and either (a) the source is unreliable, or (b) the source is reliable, but in the current environment one might easily have chosen an unreliable source. Cases along the lines of (a) are easy enough to imagine: they are cases in which we trust a bad source (a crystal ball, a pathological liar, etc.) that just happens to get it right. Cases along the lines of (b) are also fairly straightforward: Suppose by luck you happen across the only honest man in a room of pathological liars, and the man explains Einstein’s theory of general relativity to you. If all goes well you will then understand the theory, but you will (apparently) fail to know it: too easily might you have trusted some other source of information.

Since Kvanvig is not entirely clear about whether his Comanche example is supposed to be an instance of (a) or (b), to see how knowledge and understanding
sway together throughout Comanche-style cases we will need to lay out the various possibilities a bit more carefully and consider them in turn.

The following chart is one way to make sense of the landscape. I will say more about what I mean by sources of information and information environments in a moment.

**TABLE 1**

KVANVIG’S CHART

<table>
<thead>
<tr>
<th>Information Environment</th>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>knowledge, understanding</td>
<td>no knowledge, understanding</td>
</tr>
<tr>
<td>Bad</td>
<td>no knowledge, understanding</td>
<td>no knowledge understanding</td>
</tr>
</tbody>
</table>

On this way of looking at things, a *source of information* is the primary notion, and we can distinguish between a good source and a bad source as follows. A good source of information with respect to a certain domain is a source which is more likely than not to provide accurate information about that domain.\(^8\) A bad source is one that is more unlikely than not to provide accurate

\(^8\) Or, at any rate, that places above whatever the reliability threshold happens to be.
information. For example, a good source of information about the past will be someone with a reliable memory, and a bad source of information about the past will be someone who speculates about the past by using a Ouija board. A good source of information about the (well-lit, normal) immediate environment will be someone with normal perceptual abilities; a bad source will be someone who has just been taking hallucinogenic drugs.

The notion of an information environment, as I am understanding it, builds on the notion of a source of information. A good environment, information-wise, is an environment where one’s sources of information—the sources of information within easy epistemic reach, as it were—are by and large good (i.e., reliable). A bad environment, information-wise, is an environment where the sources of information within one’s easy epistemic reach are not by and large good. Suppose you are at an intersection with a dozen people, and that you need to ask for directions. A bad information environment will be one in which only one person is reliable about the directions in question, while the other eleven will send you off on a wild goose chase. A good information environment would be one in which virtually all of the people at the intersection are good sources of information, in the sense that virtually all will tend to get things right with respect to the domain in question.

---

9 Or that places below whatever the reliability threshold happens to be.

10 I am cutting things quite thickly here: someone could have reliable beliefs about their college years, for example, and unreliable beliefs about their childhood.

11 On this view, things like books will count as derivatively reliable.

12 There are complications here about people’s willingness to share their accurate information. If this is your worry, imagine that you are choosing among a pile of twelve maps instead, only one of which contains accurate information.
With this in mind, Kvanvig’s claim is that knowledge and understanding come apart in three of the four quadrants. On his view, whenever you have an element of luck involved, either from a bad information source or from a bad information environment, although knowing on the basis of this information is ruled out, understanding is still entirely possible.

As this last comment suggests, when we say that for Kvanvig knowledge and understanding come apart in these areas, we need to be careful about interpreting this in the right way. What he means, presumably, is something along the following lines: that the potential for knowledge or that the possibility for knowledge can come apart from the potential or possibility for understanding. For example, just because one bases one’s (true) belief on a good information source in a good information environment, it does not follow—on anyone’s view—that the belief amounts to knowledge. You might have defeaters for the belief that you are overlooking, for instance, or (a misleading) reason to think that the source, in fact, is not reliable. We should therefore read Kvanvig as claiming that believing on the basis of an unreliable source, or on the basis of a reliable source but in a bad information environment, permits or allows for understanding, but does not permit or allow for knowledge, in three of the four quadrants described above.

6.7 Bottom Half

In this section I will argue that Kvanvig is mistaken about the bottom half of the chart: just as basing one’s beliefs on a bad source does not allow for knowledge, so too it does not allow for understanding. In other words, the two
states in fact do not come apart. Although I will try to show that this holds for any Comanche-style case, for concreteness it will help to work with a specific example, and Kvanvig’s own (Comanche!) case will do as well as any.

The first thing to bear in mind when evaluating these possibilities, especially when understanding is ultimately at issue, is that there are two different ways to think about the nature of the information—in this case, the information contained in the Comanche book—we are considering. On the one hand, we can imagine that the book consists, at least in part, of explanations—or better (in order to emphasize their potential fallibility) of explanatory stories. On the other hand, we can imagine that the book does not contain explanatory stories at all. Instead it is the kind of book that just offers descriptive information and is silent about how this information should be interpreted or about how one bit of information illuminates the rest.

6.8 Case I: Explanatory Story from Bad Source

Let’s take the first way first, supposing that Kvanvig’s Comanche book offers explicit explanatory stories. Perhaps, for example, the book claims (and here I am frankly just making things up) that the Comanches dominated the southern plains of North America because their economy was based on a barter system, and that economies based on barter systems were more likely to produce members who were competitive and cunning, traits which in turn were enormously conducive to success in battle.

Suppose that all this is true (i.e., that the alleged dependencies are genuine), that I accept this explanatory story, and that I take myself to
understand why the Comanches dominated the southern plains on the basis of “seeing” or “grasping” how the various elements described in the story depend on one another. In keeping with the bottom half of the chart, what we now need to add is that the author of the textbook came to accept this explanatory story on the basis of extremely shoddy research. For instance, we can imagine that the author concluded that economies based on barter systems were more likely to be competitive and cunning on the basis of only one sample, and without controlling for the influence of other factors. In these circumstances, then even if it turns out that the story is entirely correct, and that there is a genuine dependence between barter systems and competitiveness on the one hand and between competitiveness and success in battle on the other, then presumably the author of the theory doesn’t genuinely understand why the Comanches dominated the southern plains. And if the author doesn’t understand this, it is very hard to see how the reader of the textbook could understand in turn.\(^{13}\)

### 6.9 Case II: Descriptive Information from Bad Source

For completeness we also need to consider a Comanche textbook that doesn’t traffic in explanations. In other words, to consider a book that is entirely descriptive, one that lays out the facts, and leaves the insightful reader free to construct explanations on his own. Since the relevant accomplishment—the piecing together, as it were—would then appear to be entirely internal, perhaps

\(^{13}\) Indeed, if understanding could be acquired in such a slipshod way (by consulting only one sample, without controlling for the influence of other factors, etc.), then it is hard to see why we would have the need for the sociologists, economists, and historians—with their established methodologies and established canons for constructing and evaluating explanations—that we evidently have.
the facts about how the descriptive information was acquired would then become irrelevant to the possibility of understanding.

But that does not follow. Suppose that the entire textbook is the product of baseless reporting, written by a 19th century version of the New York Times’s Jayson Blair—a negligent writer who could not be bothered to do the research and simply invented his claims about the Comanches out of the blue. Even if all the stars line up in his favor and his claims turn out to be entirely correct, it is hardly the case that the textbook reader who pieces things together and develops an account of Comanche dominance would really come to understand why the Comanches dominated the Southern Plains.

Of course, in saying this we do not have to deny that there is a kind of internal accomplishment involved in piecing together the facts and constructing an explanatory story in this way. All we have to deny is that the accomplishment is really one of understanding why the Comanches dominated the Southern Plains. For that accomplishment, apparently, a firmer (less accidental) connection to the Comanches is required.¹⁴

¹⁴ There is perhaps an even more straightforward way to illustrate the problem with thinking that understanding can flow from poor sources of this kind, one that can be made to fit either Case I or Case II examples. Thus consider a swampbook version of Kvanvig’s Comanche book, that is, an exact replica of a normal book brought into being by a stray lightning bolt and a patch of swampy gunk. We can imagine either that the swampbook contains explicit explanatory stories as in Case I, or we can imagine a swampbook that just presented the facts, as in Case II (all entirely accurate, of course). On either way of thinking about the swampbook, however, we would be hard pressed to say that the readers of swampbook understood anything about the actual Comanches. It is much more plausible to say instead that what they understood was something about the logic of the story, or perhaps about the patterns of dependency that held among some possible tribe. To understand the actual Comanches, again, we seem to need a firmer connection to the Comanches themselves. (If one thinks the failure of understanding here depends on failure of reference, then it is easy enough to imagine that one has originally learned about the Comanches in a more orthodox way.)
6.10 The Upper Right

Assuming that it is likewise impossible to acquire knowledge from such poor sources, so far we have the following results. First, that the possibility of knowledge and the possibility of understanding sway together on the upper left quadrant, where a good source meets a good environment. Second, where we have a bad—i.e., unreliable—source (a terrible methodology, a Jayson Blair, a swampbook, etc.), just as knowledge on the basis of such information seems ruled out, so too does understanding. So on all of the ways of filling in backstory that we have considered so far, Kvanvig’s point fails to hold: knowledge and understanding do not come apart.

Arguably, however, the best case for Kvanvig remains to be considered: namely, the upper right quadrant, or the case in which our beliefs are based on accurate information from a good source, but in the midst of a bad information environment.

Here again, we can fill in the details of this quadrant in one of two ways; we can either suppose that the information from this good source traffics in explicit explanatory stories, or we can suppose that it only provides descriptive information. It is simpler to evaluate Kvanvig’s claim, however, if we suppose that the information is purely descriptive.\textsuperscript{15}

Suppose then that the accurate information contained in the textbook comes from a reliable source: someone who is not only a Comanche expert, but whose methods are impeccable, etc. Moreover, all of the other information

\textsuperscript{15} Moreover, although it is hard to say exactly, this seems to be the model he has in mind.
sources about the Comanches within easy reach are unreliable: if they hit upon the truth, it is only by accident. For the sake of argument, suppose we agree that coming to understand on the basis of such information (i.e., the information from the good source) is possible. What I want to argue now is that, pace Kvanvig, there is a perfectly legitimate sense in which such a person can come to know on the basis of this information as well. And if that can be established (and a few more details sorted out), then Kvanvig will not have shown that understanding and knowledge can come apart after all.

The key point to appreciate here is that Kvanvig overstates his case in claiming, without qualification, that knowledge is incompatible with luck. In reality, knowledge is compatible with some forms of luck and incompatible with others. For example, although believing on the basis of information from a bad source never seems to allow for knowledge, even when by luck the bad source happens to get things right, believing on the basis of a good source in a bad information environment does seem to allow for knowledge. In suggesting in a blanket way that knowledge is incompatible with luck Kvanvig therefore paves over significant differences.

John Hawthorne (2003) has recently proposed a case that illustrates the compatibility nicely.

I give six children six books and ask them each to pick one of the books at random. All but one contains misinformation about the capital of Austria. I ask the children to look up what the capital of Austria is and commit the answer to memory. One child learns ‘Belgrade,’ another ‘Lisbon,’ another ‘Vienna,’ and so on. I ask an onlooker who has witnessed the whole sequence of events (or someone to whom the sequence of events has been described) ‘Which one of the schoolchildren knows what the capital of Austria is?’ or ‘How many of the children know what the capital of Austria is?’ It is my experience that those presented with this kind of case will
answer, not by saying ‘None of them,’ but by selecting the child whose book read ‘Vienna’—even though that child was only given the correct answer by luck. (Note in this connection that if I make a five dollar bet on a certain child knowing that Vienna is the capital of Austria, you will pay up as soon as you are convinced that the child believes the capital to be Vienna. You will not inquire further about how the child came by that information—whether by dumb luck or from an informant that normally lies—even if you have reason to suspect such an unreliable source.) (pp. 68-69)

Hawthorne’s judgment about the case seems very plausible; it seems right to credit the child with knowledge. Hawthorne does not go on to discuss why this seems right, but a few things can be said to try to clarify the case.

First, when (as third party evaluators) we have reason to believe that the source of the information is good—here, that the textbook came from a reliable author, etc.—we tend to focus on the sense in which the belief is not lucky: it was no matter of luck, we think, that the textbook author identified Vienna as the capital of Austria, even if it was a matter of luck that this particular textbook ended up in the student’s hands.

But now suppose we learn that the author’s identification of Vienna was itself the product of luck. Perhaps, for example, the information in all of the textbooks, including the one the lucky boy received, was produced entirely with the help of a Ouija board. As a result, one book contained the information that the capital of Austria was “Oog,” another “Tyuy,” and another “Vienna” (or we could substitute the names of real capitals if you like; the point remains the same). If we fill in the details of Hawthorne’s case in this way, our inclination to judge that the boy knows the capital of Austria drops considerably. He has a true belief about the capital of Austria, naturally, but we draw the line: you simply
can’t gain knowledge from a Ouija board, regardless of whether the board happens to get things right.

My intention in raising Hawthorne’s case is not, needless to say, to launch into an extended account of the various ways in which knowledge is and is not compatible with luck. The intention instead is simply to point out that, if we are tempted to think that someone can come to understand various things about the Comanches on the basis of good information source in a bad information environment, then in cases of this sort there is also a strong tendency to say that the person knows these things about the Comanches as well.

Naturally, there are still other ways to imagine the case—cases where we make the luckiness of the belief seem more salient—in which we would be inclined to deny that basing one’s belief on a good source might allow for knowledge. If we learn that before picking up the good-source Comanche book you picked up 20 other books on different subjects (dinosaurs, space exploration, etc.), all of which were tissues of lies, and that you accepted everything these books had to say, then when you come across the good-source Comanche book, our inclination to say that the beliefs you form about the Comanches amount to knowledge diminishes significantly.¹⁶

That point is no solace to Kvanvig, however, just so long as our judgments about the possibility of understanding sway along with our judgments about the possibility of knowledge. And they certainly seem to. If I pick up the same 20

¹⁶ And the same thing for Hawthorne’s capitals case. If we feed the child 20 books filled with false information about world capitals, and he blithely accepts all of this, then there is a strong tendency to think that when he comes across the one book that says “Vienna is the capital of Austria” he fails to know it.
books and I take myself to understand various things about these subjects on the basis of this information (in the first book about dinosaurs, and in the second book about space exploration, etc.), then when I come across the lone accurate book about the Comanches, the temptation to say that I genuinely understand these things about the Comanches drops significantly again, even though all of the connections that I draw might be perfectly correct.

We can sum up these results with the following revised chart.

**TABLE 2**

**REVISED CHART**

<table>
<thead>
<tr>
<th>Information Environment</th>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information Source</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>knowledge understanding</td>
<td>knowledge* understanding*</td>
</tr>
<tr>
<td>Bad</td>
<td>no knowledge no understanding</td>
<td>no knowledge no understanding</td>
</tr>
</tbody>
</table>

I have placed a star ("*") next to the contents of the upper right quadrant because these circumstances have clearly emerged as a special case. According to the ground rules I laid out in Section 6.6—where a quadrant counted as knowledge-friendly or understanding-friendly just in case it was possible for knowledge or understanding to emerge under those conditions—in one sense we should simply remove the stars: if we accept the Hawthorne-style cases, then it is possible for knowledge as well as understanding to emerge under those conditions.
As we just saw, however, cases involving a good information source in a bad information environment quickly become complicated, and it would be misleading to suggest that (defeaters aside) knowledge and understanding are the norm here. Although this starred status makes the chart less theoretically elegant, in terms of the larger project it is in a way beside the point: for us, all that matters is that when knowledge and understanding do shift with particular circumstances, they shift together. And this last fact has been repeatedly established: where the connection with the truth has been strong enough to classify a mental state as a state of understanding, the beliefs that constitute the state have also been strong enough to count as knowledge. When, on the other hand, the connection has been too weak for understanding, it has likewise been too weak for knowledge.

6.11 So is Understanding a Species of Knowledge?

So is understanding a species of knowledge, then? From what we’ve seen so far there seems to be a truth component to understanding; if you identify the cause of the refrigerators shutting down as the pulled plug when in fact it was a short-circuit, then no matter how vivid or transparent the pulled-plug story might seem to you, you will fail to understand why the refrigerator shut down. There also seems to be an anti-luck component: hitting upon a true explanatory story by accident—as through a Jayson-Blair-style source, or as in a blacksmith’s shop in the midst of eccentric nut-smashing activity—rules out understanding. In all of these cases, a stronger connection to the truth is needed.
Although the parallels between knowledge and understanding therefore run quite deep, to this point we have failed to consider another component of knowledge, one that threatens to upset the comparison: namely, the psychological component of knowledge.

The psychological component of knowledge is generally taken to be belief. When we believe something we take it to be the case, we hold it to be true.\footnote{Cf. Schwitzgebel (forthcoming): belief is “the attitude we have, roughly, whenever we take something to be the case or regard it as true” (p. 1).} A slightly more illuminating way to characterize belief is in terms of the act of assent—to believe that something is so is to assent to the claim that things are so. On this view, just as denying some claim is like saying No to it, so too believing some claim is like saying Yes to it.\footnote{Not that these attitudes are exhaustive—there are still withholdings to consider.}

But now notice: If we think that knowledge is a species of belief, and that understanding is a species of knowledge, then understanding too would turn out to be a species of belief. And given what we’ve just claimed about belief, it would then follow that the psychological component of understanding would likewise be thought of along the lines of an act of assent. It is at this point, however, that the parallels between knowledge and understanding start to look questionable.

To see the problem, suppose for a moment that the psychological component of understanding could be characterized as an act of assent. What would be the thing to which we were assenting? The most plausible answer here is a claim of the form $A$ because of $B$. For example, if you are curious about why the tides exist, then (assuming the appropriate epistemic background is in place)
your curiosity will be satisfied when you learn that the tides exist because of the moon’s gravitational pull. From a psychological point of view, the thing you will be assenting to—saying Yes to—in this case will therefore be a claim to the effect that the tides exist because of the moon’s gravitational pull.

The difficulty is that an act of assent of this sort (or even, arguably, several acts of assent of this sort) seems too thin to capture what is going on when we take ourselves to understand. Suppose that one day I get into my 1991 Volkswagen and none of the gauges—the speedometer, tachometer, etc.—come to life; they’re all dead. I drive the car to the garage and my mechanic tells me that the reason why the gauges are dead is because I have a bad ignition switch. I then seem to have excellent reason to assent to the claim that my gauges are dead because of a bad ignition switch, even though I fail to grasp how a bad ignition switch might lead to this result. In other words, even though I currently lack the epistemic wherewithall to grasp how A depends on B, I nonetheless now have an excellent basis for assenting to the fact that A depends on B.19

Examples from mathematics also bear this out. If you are my teacher and you tell me that a certain proof explains a theorem, then as a result of your testimony I will assent to the fact that the theorem can be explained because of, or in virtue of, the things you cite in your proof. But it seems obvious that I may still fail to understand how the one explains the other. Grasping the way in which

19 Imagine that I see another switch lying around. Will this one be an improvement? I don’t know. I don’t know what it is about the one that’s tied to the other. Or again, suppose I am helping you with a chemistry experiment and you tell me that the reaction we were looking for failed to occur because of an excess of oxygen in the area around the experiment. As a naïf, on the basis of your testimony I will accept this claim as true—more exactly, I will assent to the claim that the reaction failed to occur because of the excess of oxygen in the area around the experiment—but I will nonetheless fail to grasp, fail to understand, how the one depends on the other.
the theorem depends on the elements you cite in your proof is different from assenting to the claim that the dependency holds.

As this discussion suggests, and in keeping with what we have argued in previous chapters, when trying to offer an account of understanding the notion of grasping arises almost irresistibly. Moreover, when we grasp some claim we are apparently doing something significantly different from merely saying Yes to it.$^{20}$

It might be said that this attitude of grasping is nothing significantly different from an act of assent: it just involves more assent. But it seems clear that one can pile up assents as high as you like without getting a grasping. In the mathematics case, for example, I can be told that the key to the proof lies in the fact that the first element is divisible by the second, and I can assent to this (again, based on the testimony of someone I trust). But I still might not understand the proof; I might fail to see or grasp how the truth of the theorem depends on the fact that the first element is divisible by the second. And it seems like this game could go on indefinitely. You can continue to spell out to me how the various dependencies are supposed to work. Moreover, based on your testimony I can assent to these claims at every step of the way. But none of these assents, in and of themselves, add up to a genuine grasping on my part. A new kind of cognitive achievement is needed.

---

$^{20}$ Common usage also points to this connection between understanding and a distinct mental act along the lines of grasping. According to the *Oxford English Dictionary*, for instance, “to understand” is “to grasp the idea of or to apprehend clearly the character or nature of,” with “to apprehend” meaning “to lay hold of, to seize, to take possession of.” *Webster’s* likewise defines “to understand” as “to achieve a mental grasp of the nature, significance, or causal explanation of something.”
Following James Woodward (2003), one way to spell out the requisite notion of *grasping* is in terms of having an ability to answer “what-if-things-had-been-different?” questions. To have an ability to answer questions of this sort, Woodward argues, is to be able to anticipate the sort of change that would result in the thing we want to explain (the explanandum) if the factors cited as explanatory (the explanans) were different in various ways. Brian Skyrms (1980) expresses this same point in a similar way. According to Skyrms, grasping involves having an ability to anticipate that “wiggling” one variable will characteristically lead to a wiggling of another variable (p. 11). If you fail to see how things are connected in this way, however—specifically, if you fail to see how a change in the value of one variable will lead to a change in the value of another variable—then although you might be able to assent to the fact (based on testimony, etc.) that the two variables depend on one another, but you will not have *grasped how* they depend on one another.21

21 Grasping therefore seems to involve a different kind of ability than assent, a point which is especially clear when we think of what is involved in grasping, as opposed to merely assenting to, a *rule*. As Colin McGinn (1984), commenting on Wittgenstein, nicely puts the point: “Grasping a rule essentially involves knowing how to follow it; so there is a direct conceptual connection between the concept of understanding and concepts of action” (p. 122). Indeed, one helpful way to think about what is involved in grasping a claim along the lines of *X because of Y* is that this involves grasping the way in which *X* depends on *Y*—and presumably the nature of this dependence can be expressed in rule-like form: changing the value of the variable *Y* leads in some regular (i.e., rule-like way) way to a change in the value of the variable *X*. Here is a simple illustration of this point. Suppose you are teaching me a new board game, “Cops and Robbers,” and you tell me that if one of the robbers lands on a black square he will be thrown into jail. I can then assent to this fact even if I am not particularly good (perhaps even particularly bad) at applying it. I might consistently fail to put my robber piece in jail when he lands on the black square, for example, or I might mistakenly put the piece in jail when he lands on a blue square. All the while, however, I might genuinely be said to know the rule (if you asked me to cite the black square rule I will report it accurately 100 out of 100 times, let’s say) even though because of my practical incompetence in applying it we would be very reluctant to say I understand the rule; if I grasp it at all, my grasp seems obviously inadequate or deficient. Being able to apply or manipulate the rule therefore seems to involve a new and distinct kind of cognitive achievement, and one that cannot be characterized as a clustered sum of assents.
There is one final reason to suppose that thinking of the psychological component of understanding in terms of grasping (rather than merely assenting) is correct: namely, it would help to shed light on why the epistemic gain we experience when we understand is so universally valued. For notice: if understanding involves the ability to answer “what-if-things-had-been-different?” questions, then this ability brings with it the promise of tremendous control. If there are various ways in which, from your point of view, the world might be, and if you are able to identify the features which determine (in some sense) the various possible states of the thing you want to explain, then it seems like you’re in a very good position. Specifically, if you know how to wiggle variable A to produce a variation in variable B, and if the state of variable B is important to you, then you will be in an excellent position to produce the desired state and avoid the undesirable state.

6.12 A False Choice

What I want to suggest in closing is that the choice between thinking of understanding on the model of belief/assent and thinking of it on the model of grasping is not a forced one. More specifically, I want to suggest that, while belief can be quite a thin psychological state, some forms of belief are thicker than others. On this way of looking at things, belief emerges as a kind of umbrella category for “ways of saying Yes” to some claim, and within belief there are simple acts of assent and then there are also acts of assent that, as it were, combine an element of grasping (or, if you prefer, acts of grasping that combine an element of assent).
In Chapter 4 we already provisionally adopted a model of this kind when we stressed that grasping (or simple grasping) differed from grasping* (or conditional grasping) because the former brought with it an element of assent or affirmation. After all, grasping* a dependency without assent seemed compatible with even denying that the dependency actually holds, as when someone grasps how Darwin’s theory accounts for the development of species but prefers Lamark’s because it makes better sense of the rest of what he believes.

Perhaps the most powerful reason, however, to accept that some beliefs are thicker than others—and in particular that it makes sense to think of acts of graspings that combine an element of assent—is that we already have a nice model of how this works: namely, a priori knowledge. A priori knowledge is not distinguished merely by its content but also—perhaps primarily—by one’s attitude towards this content. My knowledge that 2+2=4, for example, does not qualify as an instance of a priori knowledge simply in virtue of the fact that 2+2=4 is a necessary truth. If I were to assent to this claim simply on the basis of accepting the testimony of a reliable authority, for instance, then I would thereby know it, but I would not know it a priori. What makes it an instance of a priori knowledge for me (assuming it is) instead depends on the presence of some other psychological attitude that I bear towards the content in addition to (or in some way mingled with) the attitude of assent. What exactly this other attitude amounts to is a topic I will not explore here (thankfully!), but in light of what we have said so far it is impossible not to mention that in this respect too talk of
“grasping” the truth or “seeing” that things could not have been otherwise emerges almost unavoidably.22

If we are willing to count a priori knowledge as a species of knowledge, at any rate, as we manifestly are, then the reservations we raised earlier about the inherent “thinness” of the psychological state of belief seem misguided. Although belief can be quite thin—can simply amount to a simple saying Yes to a claim—it can also have additional layers of richness. Moreover, so long as we are willing to accept that one of these layers contains an element of grasping, then all told there seems to be no good reason deny that understanding is a species of knowledge and, in light of our findings about how etiology matters to understanding, plenty of good reason to accept that it is. More directly, we have every reason to think that understanding is a species of knowledge.

22 Roderick Chisholm (1989), for example, begins one of his discussions of the a priori with the following passage: “There are propositions that are necessarily true and such that, once one understands them, one sees that they are true. Such propositions have traditionally been called a priori” (p. 26).
CHAPTER 7

EPISTEMIC GOALS AND EPISTEMIC VALUES

Reading the sports page over breakfast a few months ago, I came across the following quote from Ricky Williams, former running back for the Miami Dolphins who walked away from his million-dollar salary to pursue a career in holistic medicine. “I’m going to search for the truth,” Williams said. “Everything I’m doing in my life is about finding the truth.” Naturally, Williams is only one of the latest in a long line of thinkers who have been passionate about pursuing the truth. Socrates, for instance, could think of no greater compliment to pay his dialogue partners than that they entered the discussion only for the sake of truth. And when Socrates lost his patience with someone, it was almost always because they placed more importance in protecting their reputation or impressing the crowd than in finding the truth for its own sake.

There are a variety of conclusions that epistemologists are tempted to draw from such examples. That truth is the goal of inquiry, for example. Or that truth has an intrinsic or standing value for us, insofar as we are cognitive agents. Or even that truth is the only thing, from an epistemic point of view, that has an intrinsic or standing value.

These apparently straightforward conclusions lead to deep worries, however. If we say that truth is the only intrinsically valuable thing from an epistemic point of view, for example, then why do we think that knowledge is
better than mere true belief, or that a justified true belief is better or more valuable than an unjustified true belief? A justified true belief that it is raining is not more true than an unjustified true belief (a lucky guess, for example) that it is raining, so why do we take the former to be more valuable than the latter? There are other problems as well: most prominently, if we think that the truth has a standing value for us, then why are we unapologetically indifferent to so many truths? If you propose an evening memorizing the phone book for Topeka, Kansas, and I decline, have I really missed an opportunity to enrich myself, from an epistemic point of view? If truth has a standing value for us, it seems that I have. And yet that conclusion seems ridiculous.

In this chapter I will suggest a new way of thinking about the role of truth in our epistemic lives that promises to clarify the nature of our epistemic goal and to suggest a way to avoid some of the problems just mentioned.

7.1 Truth for Its Own Sake

Our interest in finding the truth is often motivated by our practical goals. If one of my practical goals is to board my flight to South Bend, for example, I will be interested in finding out when my flight is leaving, where the plane is docked, and so on. Practical goals aside, however, we also seem to have a purely epistemic interest in finding the truth. As Alvin Goldman (1986) notes,

Truth acquisition is often desired and enjoyed for its own sake, not for ulterior ends. It would hardly be surprising, then, that intellectual norms should incorporate true belief as an autonomous value, quite apart from its contribution to biological or practical ends. (p. 98)

---

1 For more on the so-called value problem in epistemology see, for example, DePaul (1993, 2001), Zagzebski (2003), Sosa (2003), and Greco (2003).
In later work Goldman (1999, 2001) goes further: not only does truth acquisition have an intrinsic value, in the sense that it is often desired for its own sake, on his view truth is indeed the only intrinsically valuable thing from an epistemic point of view. Thus he argues that although there are a variety of traits we value from an epistemic point of view—having beliefs that derive from reliable sources, for example—all of these other traits are valuable only insofar as they tend to lead us to the truth.

But let’s take a step back for a moment: Why think that we have this interest in the truth for its own sake in the first place, or that truth has a more than merely instrumental value for us? The following passages from Richard Foley (1987) and Marian David (2005) help to clarify things. According to Foley, Might not S lack the goal of now believing truths and not believing falsehoods? Perhaps not. Perhaps this is impossible, at least if the notion of a goal is understood in a suitably broad sense. After all, the vast majority of us are intrinsically curious about the world; we intrinsically want to have true beliefs. (p. 11)

And according to David,

We want to possess truths in part for their own sake (we tend to be simply curious about things) and in part, or maybe mostly, because we think that having truths will increase our chances for securing other, non-epistemic, goals (in particular, we want truths about the most effective means for securing our other goals). (p. 302)

For our purposes, the particularly significant part of both these claims is the appeal to curiosity. According to Foley and David, the reason why we desire truth for its own sake, quite apart from our practical goals, is because we are naturally curious about things. Even when nothing of practical importance seems to ride on finding out how things stand with respect to a certain subject, we seem to have
a natural—or, as it were, purely epistemic—interest in finding out how things stand.

I will return to the role of curiosity in a moment, but first I want to point out how seductive a certain way of thinking about the value of truth begins to seem once these innocent-looking premises are in place. Notice that we began this section with the claim that we often desire truths for their own sake. Even when nothing of practical importance seems to hinge on these truths, they still seem worth pursuing. Moreover, the explanation seemed straightforward: the truth as such interests us because we are the kind of beings who are naturally curious.

But now notice how tempting it seems to conclude that any truth is worthy of our interest, or has a standing value for us, from an epistemic point view. For, after all, what does it mean to be curious? Fundamentally, it seems, two things: first, that we desire the truth for its own sake, and second, that this desire is open-ended or unrestricted. From the point of view of our natural curiosity, any truth would seem worthy of our interest. Thus, for instance, John Stuart Mill writes in *Utilitarianism* that,

A cultivated mind—I do not mean that of a philosopher, but any mind to which the fountains of knowledge have been opened, and which has been taught, in any tolerable degree, to exercise its faculties—finds sources of inexhaustible interest in all that surrounds it: in the objects of nature, the achievements of art, the imaginations of poetry, the incidents of history, the ways of mankind, past and present, and their prospects for the future. ([1861] 2001, p. 14)

Since Mill does not place a limit on his “all that surrounds it” claim, the implication is that his list of topics is representative rather than exhaustive. A
curious mind, a cultivated mind, finds an interest in how things stand, without exception, or in everything that is the case, without limit.

It is because this conclusion seems so tempting, I suggest, that the following deeply important question seems so stupid: namely, what is it about a given truth—or better, about a given subject—in virtue of which it sparks our curiosity? The question seems stupid because it looks like at this point we already have an answer: A given subject sparks our curiosity because it holds the promise of truth, and as curious beings we want the truth. There is not something in addition to the truth that we want, or that explains why we want certain truths. After all, we've already said that insofar as we are curious we want the truth as such, the truth for its own sake.

In the following sections I will try to spell out why this isn’t a stupid question after all, but perhaps the best way to illustrate the need for asking the question is first to consider an example of a theorist who seems to embrace the seductive conclusion just mentioned. Thus according to Jonathan Kvanvig (2003):

[W]e do have an interest in the truth, both pragmatic and purely intellectual. It is the nature of interests to lack specificity: We do not have an individuated interest in the truth of the claim that our mothers love us, that the president is not a crook, that Wyoming is north of Mexico, and so on. What we have is a general interest in the truth, and that interest attaches to particular truths in the manner of instantiation in predicate logic. The default position for any truth is that our general interest in the truth applies to it, though, of course, there can be special circumstances involved so that the general interest in the truth is overridden by other factors.... We are finite beings, with limited time and resources for enhancing our well-being; and without some special situation in which counting the grains [of sand] brings pleasure to a person, perhaps only by passing the time in a way not completely boring, our general interest in enhancing our well-being comes into conflict with our general interest in the truth. (p. 41)
Although Kvanvig does not explicitly use the word, along with Foley and David he would presumably attribute the general value we place in finding the truth to something along the lines of our natural curiosity. It seems to be a hallmark of our curiosity, again, that we are not interested in the truth for practical purposes but rather for its own sake. Moreover, it seems to be a hallmark of our curiosity that we are interested in the truth in an open-ended way; a truly curious person seems to have a desire to learn about *anything*.

Of course, Kvanvig is realistic. He acknowledges that the standing value we place in learning the truth can be overridden; the vast majority of the time, it will be more important to pursue truths that promise a practical payoff of some kind. But on his view this does nothing to undercut the standing value that learning the truth—indeed, any truth—holds for us. Were it not for the fact that we are “finite beings”—had we world enough and time, as it were—we would apparently be able to exercise our curiosity in its purest form and pursue the truth in any entirely unrestricted way.

### 7.2 Curiosity

But why think that? Specifically, why think that, in virtue of our curiosity, it is the truth *per se*, or finding out how things stand with respect to *any* subject, that has a standing value for us? Cases along the following lines might be thought to lend support to this idea. Suppose as you sit down at a desk in the library you notice a piece of paper turned face down in the upper right corner. Naturally, you turn the piece of paper over and take a look at it. Why? The answer seems obvious: you’re just curious! Or again, suppose you’re wandering
just below the crest of an unfamiliar hill. Many of us, perhaps not so many as would turn over the paper, but many nonetheless, will naturally be interested in what lies beyond the crest of the hill. Why? Again, because we’re just curious! Even when there is no obvious practical benefit that attaches to these actions, we seem to be driven by a desire to find out how things stand.

This all seems to move too fast, however. When we turn over the piece of paper, for instance, what is it, really, that drives us? Why does this particular subject spark our curiosity? On the view just considered, this subject sparks our curiosity because finding out how things stand in general sparks our curiosity. Or, as Kvanvig might put it, this is interesting because the truth in general is interesting, and this is an instance of the truth. But this explanation matches the facts quite poorly. If it were truth simpliciter that sparked our curiosity, then why does our interest focus on the piece of paper rather than on the countless other truths within easy reach? Suppose there are presently 53 motes of dust on the desk. Why not spend the next few minutes counting them? For that matter, why even focus on one’s immediate surroundings at all? Assuming you have a healthy stock of beliefs, why not just start conjoining them with one another? Or disjoining individual beliefs with any random proposition? The number of potential truths within easy reach quickly begins to look stunningly large.

But if it not my general interest in the truth that explains my interest in the paper, then what is it? Why is learning the truth about this subject worth pursuing? A natural answer seems to be that there is something else, apart from

\[\text{As Sosa asks in his (2003).}\]
truth, that explains the worthwhileness of this subject. Perhaps, for example, my interest in finding out how things stand with respect to the paper can be explained by the general value I place in the well-being of others: someone may have left this behind, I think, and if so perhaps I can help them recover it. More likely, perhaps, it can be explained by the standing value I place in finding out other people’s business, combined with my sense that this paper holds the promise of such information. All of this is to emphasize, however, that acquiring these beliefs is not valuable just insofar as they are true. Instead, they are valuable because finding out how things stand with respect to the paper promises to lead to other things we value: for example, the value we place in altruistic acts or voyeuristic discoveries.

To this a theorist like Kvanvig might protest that while these additional values might explain the particular salience of finding out how things stand with respect to the piece of paper, that does not in itself tell against the value that the subject had all by itself, simply in virtue of its promise of truth. Adding extra value to a subject will help it to stand out from the crowd of subjects with this promise, but that is not to say that from an epistemic point of view it was not the truth per se that we found worth pursuing. What’s more, it is in this sense that counting the motes of dust on the desk can claim the same standing value, from an epistemic point of view. Again, the idea would be that had we world enough and time we would realize that these truths too were valuable and worth learning, but as things stand our other interests tend to crowd out our purely epistemic interests.
But that just seems hard to accept. Suppose we take away my finitude, at least in the sense of making me immortal. If at some point counting the motes of dust on my desk seemed worth doing from a purely intellectual point of view, then I can only conclude with Bernard Williams\textsuperscript{3} that immortality would be a tedious and dreary prospect indeed, and itself not worth having. When the only data we have to go by tells us that there is nothing worthwhile at all about counting motes of dust or memorizing phone books, then we should take these data at face value and look for a better way to explain our interest in the truth.

### 7.3 A New Start

Clearly, we have hit a glitch somewhere. We began with Goldman’s observation that from a purely intellectual point of view we are often interested in learning the truth as-such, or for its own sake. This then spun off into various claims about value. Most importantly, that the reason why we desire truth for its own sake is because the truth \textit{per se} has a standing value for us, in virtue of our natural curiosity. But it doesn’t take much argument to show that finding out how things stand with respect to countless subjects seems to hold no standing epistemic value for us at all.

We need to find a new approach, then, one that makes sense of both these results. Again, on the one hand that as intellectual beings we often desire the truth for its own sake. And on the other that even those of us with a passably cultivated mind seem indifferent to countless truths.

\textsuperscript{3} Williams (1976).
To get a sense of where things went off track it will help to consider two analogies. In both cases, the tactic will be to start with ends that we seem to desire for their own sake, and then to ask why the ends seem to have this status.

So, first, suppose that,

[1] Reading *Don Quixote* is something I desire for its own sake.

Why think [1]? Well, among other things, the goal of reading *Don Quixote* seems to be something for the sake of which I do many other things. For example, it is something for the sake of which I will go to the bookstore and buy it, clear away my evenings so that I might have free time, and so on. Moreover, it is certainly conceivable that in reading it I might have no other practical goal in mind: no literature course to prepare for, no friends I hope to impress at cocktail parties, and so on. It seems to be the kind of end that I desire as such, or for its own sake.

All that said, however, a number of questions still seem perfectly sensible. For instance, Why is reading *Don Quixote* something that I desire for its own sake? Why is it something I find worth doing? On the Kvanvig model considered in the previous section, the answer would seem to be because:

[2] Reading [anything] is something that has a standing value for me.

But, patently, [2] is false: reading anything does *not* have a standing value for me. Reading Paris Hilton’s *Confessions of an Heiress*, for instance, is not something that I find worth doing, nor is reading the fine print of the latest Sears catalogue. And by this I mean that reading the fine print in the latest Sears catalogue does not even have *pro tanto* value for me; it is not the kind of thing
which has some weight, and which happens to be outweighed by other things I value.\(^4\)

Then what is it about the reading of *Don Quixote* that makes it something I desire for its own sake? I’m sure there are many sophisticated ways of answering this question, but a commonplace answer will do: so let us simply say, crudely, that reading *Don Quixote* has this standing value for me because it is very funny and poignant. The reading of *Don Quixote* is valuable, therefore, not because it is an instance of reading but rather *because of* these other features. Provisionally, we can therefore say that the proper explanation for [1] is something along the lines of

\[3\] Reading [very funny, poignant things] is something that has a standing value for me.

So we might say: some reading is worth pursuing for its own sake, but the reason why it is worth pursuing for its own sake is not primarily because it is an instance of reading but rather *in virtue of* its other features.\(^5\)

Another analogy should help to illustrate this point. So suppose that,

\[4\] Kissing my wife is something I desire for its own sake.

Again, we can clearly ask: Why is this so? What explains why it is desirable for its own sake? Naturally, the mistake would be to try to explain [4] by appealing to

\[^4\] Embarrassingly, that’s probably not quite right about the Paris Hilton memior: acquiring gossipy information about celebrities does seem to have at least *some* standing value for me. So let’s stick with the Sears catalogue.

\[^5\] By the same token, it should be pointed out, the reading of the fine print in the Sears catalogue, although it has no standing value for me, can certainly *become* valuable (worth doing) in light of the other goals and desires I might acquire. For example, if I have ordered some tools from Sears and I want to return them, the fine print is where I want to go if I want to find out how things stand with respect the Sears return policy. The reading is then valuable—or, more exactly, the acquiring of this information—is then valuable in light of my other goals and desires.
[5] Kissing [anyone, or kissing simpliciter] is something that has a standing value for me.

Because, patently, [5] is false. I have no desire to kiss Henry Kissinger, for instance, or to kiss a brick wall. And by this I mean that kissing Henry Kissinger or kissing a brick wall does not even have pro tanto value for me; it is not the kind of thing which has some weight, and which, as it happens, is outweighed by other things I value.

What is it then that makes kissing my wife something I desire for its own sake? Here again, a crude answer should be good enough: So let's simply say that kissing my wife is valuable because I love her. In other words, that [4] is true because of [6]:

[6] Kissing [someone I love] is something that has standing value for me.

Naturally, it is in the nature of a standing value that it can be overridden by other things we value. So it is not as if accepting [6] commits us to the ridiculous idea that it is always all-things-considered preferable to kiss those we love. What it does help to explain is what it is that we are responding to when we say that kissing someone we love is desirable in itself or for its own sake.

If these analogies are apt, at any rate, they suggest a way through the thicket we identified earlier in this section. Specifically, they show how it might be possible to reconcile the fact that (a) from an epistemic point of view, we often desire the truth for its own sake, with the fact that (b) countless truths seem flatly indifferent to us, considered from an epistemic point of view. For if what we've said here is correct, even if we acknowledge truth as a goal worth pursuing for its own sake, there is still a significant explanatory question to be asked: Namely,
what is it, in virtue of which, learning certain truths is desirable for its own sake?
Or better, why is finding out how things stand with respect to certain subjects
worth pursuing while others are not?

7.4 Questions

Let us suppose provisionally, then, that the value of truth can be accounted
for in a similar way. In other words, let us suppose that there is some way to fill
in the brackets along the lines of [3] and [6], so we can acknowledge that it is
often the truth as such that we want, rather than the truth for some particular
purpose, while also acknowledging that these truths might have their value at
least in part in virtue of possessing other features.

The outstanding question then is this: What are these additional features,
in virtue of which certain truths are interesting? I think this question can be
answered most fruitfully if we polish or tweak it a little. Rather than asking
“What is it that makes certain truths interesting?,” we would be better off asking
“What is it that makes certain subjects or topics interesting?” Not only does the
former question sound poorly formed, it seems to get things back to front. What
interests us is certain subjects or topics, and what we want to find out is how
things stand with respect to these subjects or topics; more briefly, we want to find
out the truth with respect to these subjects or topics. But it is the subject or
topics which interests us initially, and the truth about the subject or topic which
promises to satisfy this interest.
So with respect to those subjects or topics which offer the promise of truth, why are some worth pursuing and others not? Philip Kitcher (2004) has recently addressed a question along these lines. As Kitcher points out,

The resolute efforts to ban talk of values from the philosophy of science have obscured the fact that certain types of questions arise for us, and we seek explanations that answer these types of questions. (p. 216; emphasis in original)

He elaborates this thought as follows,

Thoughtful and perceptive people throughout history have sometimes entertained a question not because the answer would enable them to do something practical, something they couldn’t have managed without it, but simply because the question itself fascinated them. When we view a completely pragmatic account of the sciences as inadequate, I think we’re responding to this (almost?) universal human sense of curiosity. Our aim... [is] simply to answer the questions that matter to us. (p. 216)

Two suggestions from these passages seem particularly important, one of which we have discussed already: namely, that certain questions interest us simply in virtue of our curiosity. The other suggestion remains to be explored: namely, that the questions or topics that interest us in this way tend to share certain common features: that they belong to a certain type. Unfortunately, Kitcher’s account of the kinds of questions or topics that interest us is rather vague. The questions that pique our curiosity are questions that matter to us, he tells us, and they

---

Christopher Hookway (2001) likewise adopts a similar question-first approach. On his view, our primary epistemic business essentially consists in inquiry, and (in his words) “We have identified inquiry as an attempt to answer a question” (p. 199). Ernest Sosa (2003) is skeptical that anything general can be said about the kinds of questions that intrinsically interest us: “We shall be interested in a huge variety of questions, as family members, as citizens, and just as rational, naturally curious beings. Is there anything general that can be said here? Can some general desire for the truth be recommended? It is hard to see what it could be” (p. 158). Perhaps what Sosa is resisting here is the “recommendation” aspect of the claim in the penultimate sentence. There is no unique form of epistemic flourishing, Sosa later suggests (p. 159), so it would be a mistake to suppose that only one kind of question is worth pursuing. I mean my claim here to be innocent of the kind of recommendations Sosa mentions; more exactly, I take the claim I offer here to be descriptive, not prescriptive.
matter to us in a non-pragmatic way. This doesn’t seem to tell us much that we don’t already know. Can we do better?

Here is one promising proposal with respect to our why-questions at least: a situation elicits our curiosity and therefore prompts our why questions in virtue of our sense that the subject of the situation might have been otherwise, given the kind of thing the subject is. This suddenly brings some new philosophical machinery into the picture, machinery that I have tried to explain in earlier chapters, especially Chapter 3. The basic idea behind the proposal, however, is fairly simple. For example, when the moon goes through different phases, one day appearing as a sliver and shortly thereafter appearing full, we naturally have an interest in finding out—we are naturally curious about—what it is that accounts for this difference. Again, when the tides lapping up on the shore are sometimes high, sometimes low, we want to know what it is that accounts for the difference between the high and the low tides. Or if some metal bars rust and others do not, we want to know what it is that accounts for the rusting in some rather than others.

The most venerable version of this type of question concerns the nature of the universe itself. As Derek Parfit (1998) points out, even on the supposition that the universe has always existed, its mere existence still seems to call for explanation:

---

7 I leave open here whether there might be other types of questions, other than why-questions, that naturally pique our curiosity in the way Kitcher suggests. For example, it might be that we place a standing value in answering a “what-question” along the lines of “What’s this stuff made of?” In other words, with respect to any substance you might think of, we have a natural interest in pulling it apart in order to find out what it is composed of.
Suppose first that the universe has always existed. Some believe that, if all events were caused by earlier events, everything would be explained. That, however, is not so. Even an infinite series of events cannot explain itself. We could ask why this series occurred, rather than some other series, or no series.

He continues,

Of the supporters of the Steady State Theory, some welcomed what they took to be the theory’s atheistic implications. They assumed that, if the Universe had no beginning, there would be nothing for a Creator to explain. But there would still be an eternal Universe to explain. (p. 24)

Why, given the Steady State theory, is there still an eternal universe to explain? Apparently, because we can still ask a sensible question along the lines of: Why this (eternal universe) rather than nothing at all? It seems possible that there might have been nothing, after all. So why something? Granted that the sense of possibility that drives this question is quite different than the sense of possibility that drives us to ask (say) why are the tides high rather than low—nonetheless, in both cases the question arises, and the desire for an explanation is stimulated, because of our sense that things might have been otherwise, and we are naturally interested in finding out why one alternative obtained rather than another.

Of course, as a matter of record not everyone is interested in, or curious about, such things. People might spend their whole lives observing the tides and noticing how they differ from day to day, for instance, and not be interested in what it is that accounts for the difference. Their lives might simply be too consumed with work or sickness or depression to bother. But this is consistent with the claim that these questions have a standing or pro tanto value for us, because by their very nature pro tanto values are the sorts of things that can be

8 For that matter, why this eternal universe rather than a non-eternal one?
weighed against one another, and thus are capable of being outweighed. It also bears pointing out that the kind of pure intellectual interest or curiosity we are thinking about here is something of a luxury item. In the face of a grinding work schedule, it may simply not be possible to pause and ask why things are one way rather than another. But this again hardly tells against the pro tanto value that such questions possess.

More importantly, once we understand why certain things stand in need of explanation for us, the number of these things begins to balloon very quickly. To return again to a passage cited earlier in Chapter 3, as William Alston (2005) points out,

> Whenever any claim is made about something other than an explanation, one can seek an explanation of its being so rather than otherwise. We can ask, “Why does this object have this property [rather than some other]?,” “Why did this happen where and when it did [rather than elsewhere and at another time]?,” “Why does this proposition entail that proposition [rather than some other]?,” “Why does this substance weigh more than that one?,” and so on. So explanation is one sort of thing we can seek with respect to any subject matter whatever. (p. 165)\(^9\)

Indeed, appreciating Alston’s point allows us finally to make sense of something that had been motivating alternative accounts of the value of truth from the beginning: namely, that our sense of curiosity, of wanting to learn about the world, seems essentially unbounded. We are now in a position to recognize the sense in which this is true. For, as Alston points out, in principle there appears to be no limit to the kind of “why this rather than that?” questions that can fascinate us, and apparently fascinate us for their own sake. Alston’s insight also helps to explain why someone might be tempted to think that we have a (standing)

---

\(^9\) In adding the brackets, I have tried to draw out the contrastive (“why this rather than that?”) structure to these questions that Alston calls attention to in the first sentence.
interest in finding out how things stand with respect to *any* subject or question. Since the number of “why this rather than that?” questions that can fascinate us seems unlimited, there is a strong tendency to think that any fences one might attempt to draw between types of questions will be artificial. Unlimited curiosity, it is tempting to think, is precisely the kind of thing that can’t be fenced in.

This thought rests on a mistake, however, for as we know from mathematics, there can be *orders* of infinite things. Thus just as there are more natural numbers than there are real numbers, even though there is an infinity of natural numbers, in the same way, I want to suggest, there are more questions out there that hold the promise of answers than there are questions that interest us, even though the number of questions that interest us seems essentially unlimited.

All that said, it will help in closing to see how this plays out with respect to a particular example. So consider the following claim, which at least was true before I came to learn about the relationship between the moon and the tides:

[7] Finding out why the tides are high rather than low on a given day is something I desire for its own sake.

But what explains why [7] was true—in other words, why I had that desire? The unsatisfying answer, again, is to say that [7] was true because the truth *per se* had a standing value for me, and this inquiry promised to lead me to the truth. A significantly better answer would instead have been something along the lines of:

[8] Finding out [what accounts for the difference between the various ways in which, in virtue of my beliefs about the kind of thing it is, the subject of the situation might have been] has a standing value for me.
[8] certainly seems true, especially considering the fact that standing or *pro tanto* values are capable of being overridden.

### 7.5 Summing Up

One conclusion to be drawn from this is that the truth goal is itself a vessel heavily laden with value. Less metaphorically, one conclusion to be drawn is that when we desire the truth for its own sake, we often desire the truth *because of* these other things we value, even if not straightforwardly *for the sake of* these other things. In the remaining chapter I will briefly consider how the results of this discussion might bear on broader issues of epistemic appraisal.
CHAPTER 8
CONCLUSION

When Aristotle famously begins his *Metaphysics* with the claim, “All men by nature desire understanding (episteme),” presumably part of what he is trying to capture is the idea that all human beings have a natural desire not simply to know *that* such-and-such is the case but to figure out *why* it is the case. As Jonathan Lear (1988), commenting on Aristotle, writes:

> From earliest childhood humans display an innate curiosity. Indeed the British psychoanalyst Melanie Klein once called this childhood curiosity *epistemophilia*—love of episteme. But curiosity is not, I believe, the best way to conceptualize what drives men on. Perhaps it is better to think of man’s natural capacity to be puzzled. We tend to take this capacity for granted. Yet it is a remarkable fact about us that we cannot simply observe phenomena: we want to know *why* they occur. We can imagine beings who simply watched the sun set and the moon rise in the heavens: they might come to expect regular transitions, but they would lack curiosity as to why the changes occur. We are not like that. The heavenly motions cry out (*to us*) for explanation. (3)

Although perhaps Lear is right to suggest that mere curiosity—say, curiosity about who won the Mets game last night—is not what drives us at a fundamental level, curiosity about why things are one way rather than another does seem to drive us. It is in this sense that the heavenly motions, no less than the squeaking wheel on my bicycle, cry out (*to us!*) for explanation.

Given that we are driven by a desire to find out why things are one way rather than another, in this dissertation I have argued that the satisfaction of this
desire will involve grasping what it is that the difference between (or among) these alternatives depends on. If this grasping is tightly enough connected to the truth, moreover, then that great good—understanding—will result.

In closing I will suggest that this way of thinking about understanding leaves us with two important questions about the nature of epistemic value and epistemic appraisal.

First, if what we have said so far is correct, then someone (especially someone with a naturalistic frame of mind) might be tempted to conclude that the very notion of *intrinsic epistemic value*—of truths that, as Goldman claims, we desire strictly for their own sake and not for ulterior prudential or biological ends—needs to be rethought. If, as I suggested in Chapter 7, understanding is one of the few, and perhaps the only, epistemic goals that we seem to desire for its own sake or just insofar as we are intellectually curious, then in light of the prospect of tremendous environmental control that understanding offers there is reason to wonder how pure the desire for understanding really is. In other words, perhaps understanding too is not desired for its own sake but rather—at the most fundamental level—for the sake of ulterior prudential or biological ends. If that’s true of understanding, however, then the venerable idea that as epistemic agents what we value above all is “truth for its own sake” begins to seem suspect.

Second, the results of the last chapter likewise suggest that a prominent way to think about the nature of epistemic appraisal—call it “the teleological approach” to epistemic appraisal—is misguided.¹ On this approach, truth is taken

---

to be the thing with intrinsic epistemic value and other forms of epistemic status—for example, having a belief that is rational, or justified, or reflectively defensible—are regarded as derivatively valuable, or valuable just to the extent that beliefs with this status are oriented to the intrinsically valuable thing.

If the results of Chapter 7 are correct, however, then at best only a certain subset of truths (or, better, subjects) will have this kind of intrinsic epistemic value. On the teleological approach to epistemic evaluation, however, this result leaves us at a loss for how to appraise beliefs that are directed towards truths or subjects that lack intrinsic epistemic value. Is my belief that there are 53 motes of dust on my desk unjustified, for example, or irrational, just because believing such a thing is so apparently devoid of epistemic value? That hardly seems right, but if so then whether or not we appraise someone’s belief positively or negatively cannot be a function of how well they do with respect to a goal that has intrinsic epistemic value.²

In light of the problems with the teleological approach to epistemic appraisal, what I suggest instead is that in appraising beliefs what we are really doing is evaluating whether the source of the belief is a good (i.e., reliable) one.³ In other words, we are asking ourselves how good/reliable/trustworthy the source of the belief is with respect to this information. We are interested in the reliability of sources of information, moreover, not because the truths that

² For further criticism of the teleological approach to epistemic evaluation, see Sosa (2003) and Kelly (2003).

³ This is perhaps just another way to make a point that John Greco makes in his (2003) and in his (forthcoming): epistemic evaluations are first and foremost agent evaluations. What we are evaluating is how trustworthy people are as sources of information, not how well they are doing with respect to a goal which may or may not have intrinsic epistemic value for them. An important forerunner to this view can likewise be found in Craig (1990).
reliable sources tend to generate have an intrinsic epistemic value—as it were, a kind of magical aura. Rather, we are interested in the reliability of sources—or, perhaps better, of agents—because as information-sharing and information-dependent creatures we rely on others to provide us with accurate information so that we might effectively carry out our other projects and concerns.

This way of looking at epistemic appraisal, finally, brings with it the following advantage: namely, it helps us to see the sense in which epistemic appraisals are not just evaluative but normative. In other words, it helps us to make sense of the fact that when we appraise someone’s beliefs we often blame them, find them irresponsible, suggest that relative to their evidence they ought to have believed a certain way. What all of this suggests, however, is that epistemic appraisal is likewise more than just a matter of evaluating performances—performances that, as it were, we may or may not have a stake in, as when we praise the accurate serve of a tennis champion but do not blame (instead, merely acknowledge the inferiority of) the unreliable performances of the novice. Given that, as information-sharing creatures, we essentially depend on others for good information, it makes sense that our terms of epistemic appraisal should have developed a sharper, more deontological edge.

Perhaps, then, Chisholm was right after all in suggesting that as “intellectual beings” we have an obligation to the truth. If this is so, however, it is because we first have, as potential sources of information, an obligation to each other.


Baehr, Jason. Forthcoming. “Character, Reliability, and Virtue Epistemology.”


Harris, W. S. *et al*. 1999. “A Randomized, Controlled Trial of the Effects of Remote, Intercessory Prayer on Outcomes in Patients Admitted to the Coronary Care Unit.” *Archives of Internal Medicine* 159.


