Lecture Notes on KNOWLEDGE

Brian Weatherson

2015

Notes to Students

These are the course notes for the first two-thirds of Philosophy 383: Knowledge and Reality, in Winter 2015 at the University of Michigan - Ann Arbor.

They are not meant to replace reading original sources. As the syllabus makes clear, there are other readings that you should be doing most weeks.

You should do even more reading than that on the topics you are writing papers on.

Contents

Last revised: **April 14, 2015**

Chapter 1

Introduction to Knowledge and Reality

In this course we're going to investigate three topics:

1. Arguments for Scepticism.
2. The Analysis of Knowledge.
3. Knowledge and Society

In this chapter, I'll briefly introduce the three topics, and say a little about why we find each of them interesting.

1.1 Arguments for Scepticism

There is an important kind of sceptical argument that traces back, in modern Western philosophy, to Descartes (1641/1996a). (Though we will consider ways in which similar considerations arise in earlier philosophy, as well as in philosophy in other traditions.) Consider some thing that you know, e.g., that you have hands. Now consider some scenario where you are tricked into thinking that is true. To use Descartes's own idea, imagine that an evil demon implants in you the idea that you are sitting in a comfortable chair, in a human sized and shaped body, reading some epistemology notes, etc. Meanwhile, none of this is true; you are an immaterial soul in an immaterial world, and totally without hands.

There is some intuitive support for the idea that we can't rule out such a scenario. And there is also intuitive support for the idea that if we can't rule out such a scenario, then we don't know that we have hands. After all, knowing something suffices for ruling out alternative possibilities. So, we might think, we don't know we have hands.

Now that argument is too quick on many levels. For one thing, I don't find the initial intuition that strong. I can easily rule out the scenario, I think, by simply looking at the world and seeing that it is material. And even if I found the intuition more compelling, I would think it is much less compelling than the negation of the conclusion it is trying to draw. If I had to choose between the intuition that I know a lot about the world, and the intuition that I can't rule out this Cartesian demon, I would certainly ditch the sceptical intuition.

But the sceptical argument isn't merely an appeal to intuition. There are many ways to motivate the sceptical premise, i.e., that we can't rule out the scenario of the Cartesian demon. Looking at those ways is interesting for three reasons. First, we might find scepticism more appealing than we did when it was supported by a raw intuition. Second, we might find some limited forms of scepticism at least somewhat appealing. But finally, we might learn what plausible sounding principles we have to jettison if we are to remain anti-sceptics in good standing.

We will look at five ways of developing the sceptical intuition into a full argument.

One of these ways is **dialectical**. Whatever our reasons are for being anti-sceptical, they don't seem sufficient to convince a sceptic. Indeed, they don't even seem sufficient to move anyone who feels the pull of scepticism, but can't bring herself to accept it. Perhaps there is an argument here, starting with the idea that if we have a good reason to believe something, then we should be able to convince others of it. If we can't be convincing, maybe our reasons aren't that good.

Another way involves **underdetermination**. In some sense, the way things seem to us doesn't determine whether we are in a world like the one we think we are in, or in a demon world. Perhaps when your evidence underdetermines your world, you can't know which world you are in.

Another, related, way is **evidential**. Arguably, we have the same evidence as the victim of a Cartesian demon. And, arguably, people with the same evidence can know the same things. From these premises it follows that we know the same things as the victim of the Cartesian demon. But by hypothesis, he/she/it knows very very little. So we are similarly ignorant.

Another way goes by the idea of **sensitivity**. Robert Nozick (1981) famously argued that to know something, it must be such that if it weren't true, you wouldn't believe it. We'll spend a bit of time on this idea during part one of the course, because Nozick turned this idea into an attempt to analyse knowledge. What matters for our purposes is that our belief that we are not victims of a Cartesian demon is insensitive; even if it were false, we'd still believe it. If insensitivity precludes knowledge, then we don't know that we're not victims. We won't spend much time on this in part 2, since the discussion in part 1 will lead to the conclusion that sensitivity really isn't a constraint on knowledge.

And the way we'll spend most time on is **methodological**. Hume's arguments against the reasonableness of induction (in Hume's somewhat idiosyncratic sense of reasonableness) involved arguing by cases. We can't know by the light of pure reason that induction worked, Hume argued, because sometimes it does not work. And we can't know by experience that induction works, because all our evidence shows is that induction has worked in the past, and concluding that it will keep working would involve an inductive leap. And, prior to knowing that induction works, that would be horribly circular. But all our knowledge comes from pure reason or observation, so there is no way to reasonably believe that induction works.

A similar argument can be given by the sceptic. If you know you're not the victim of a Cartesian demon, then you know this either by reason or observation. But you can't know it by reason, since it is a coherent hypothesis, and reason does not let you rule out coherent hypotheses. And there are arguments that you can't know it by observation either, though the details here will have to wait until we get to this part of the course.

Now I certainly don't want to endorse any one of these arguments. In fact, I think they all fail. But some of them are interesting enough that some people could reasonably think they work. And the rest of us will learn a lot by seeing how they fail.

1.2 Analysis of Knowledge

An analysis of a philosophically interesting concept would satisfy three conditions.

1. It would provide **necessary** conditions for the application of the concept.
2. It would provide **sufficient** conditions for the application of the concept.
3. It would in some way illuminate the concept.

Here is one example of an analysis that does all three of these tasks.

> x is y's **sister** if and only if:
>
> - x is female; and
> - x and y have the same parents.

Both clauses are necessary. If x is not female, then x is not y's sister. And if they don't have the same parents, then x is not y's sister. (I'm assuming here that if they just have one parent in common, then x is at most y's half-sister; if you don't conceptualise siblinghood relations that way, substitute for the second clause the claim that x and y have a parent in common.) But between those two clauses we get sufficient conditions for sisterhood. If you're a female who has the same parents as someone, you are thereby their sister.

This account is potentially illuminating, at least in the sense that you could imagine explaining what a sister is to someone that way. A very young child who has the idea of mommy and daddy, but doesn't have any ideas about siblings (perhaps because they are an oldest or only child) could learn what a sister is this way.

Note that when we say the analysis is illuminating, we don't mean that it provides an operational test for settling all tricky cases of whether someone's y's sister. There are vague cases, and difficult cases, both of being female, and of having the same parents. In those cases, the analysis will say that it is vague whether one person is another's sister.

That's consistent with the analysis being a good one; saying that the analysis must be illuminating is not the same as saying it must clearly settle all hard cases.

I've started with a case where analysis seems to work, but most philosophers are very sceptical that there are particularly many useful analyses of philosophically interesting concepts. The history of philosophy is littered with unsuccessful attempts at analysis. Indeed, a standard plot line in early Platonic dialogues is that one character, often the titular character, will propose an analysis, and Socrates will show that it doesn't work. Against that, there are remarkably few successful analyses of philosophically interesting concepts. Some people insist there are none; this strikes me as too pessimistic about the status of analysis, but the success rate is very low.

In contemporary philosophy, scepticism about analysis was given a huge boost by Wittgenstein (1953). He showed how hard it would be to give a successful analysis, in something like the above sense, of 'game'. And for good measure, he argued that the concept of a game is very important to a lot of philosophical projects, especially projects to do with communication. He was right about both of these points, and we should be hesitant about the prospects for success of other attempts at analysis.

But the fact that analysis has a low probability of success doesn't mean that it isn't worth trying. Even if the project fails, we might be like explorers who find valuable things while not finding what they set out to see. (We don't regard Columbus as a failed navigator because he never found a better trade route to India, for example.) And actually, I think that's exactly what has happened in recent attempts to get an analysis of knowledge. None of the attempts succeed. In my opinion, none of them are particularly close to succeeding. But in the process of not finding a successful analysis, we have learned a lot about knowledge. And imparting those lessons will be a major goal of this part of the course.

We will primarily be looking at analyses which are, vaguely, in the JTB tradition. 'JTB' here stands for **J**ustified **T**rue **B**elief. The JTB analysis says that this is an analysis of knowledge.

> *S* knows that *p* if and only if:
>
> - *S* believes that *p*; and
> - *S*'s belief that *p* is justified; and
> - *p* is true.

This is sometimes referred to as the 'traditional' account of knowledge, though to be honest I have never seen much historical evidence to warrant that designation. It was defended as an analysis in A. J. Ayer (1956), and then refuted by Edmund Gettier (1963). We'll start our exploration with Gettier's paper, and then look through the voluminous literature it led to.

Gettier's examples have what Linda Zagzebski called a 'double luck' structure. The subjects are unlucky in a certain way, then end up with a true belief because of a second piece of luck. Much of the literature on the analysis of knowledge since Gettier has concentrated on these cases, and whether something like the JTB analysis can be amended to deal with them. There are two other important strands in the literature relevant to whether knowledge can be analysed. We'll only have time to look at one of them, unfortunately.

One strand concerns whether there are counterexamples to the JTB account that don't share this 'double luck' structure. We will some time on these examples, in part because they are in some ways more relevant to everyday life and scientific practice than the double luck cases.

Another strand concerns whether some analysis that is nothing like the JTB analysis can work. For instance, many philosophers in recent years have been attracted to theses like the following.

> *S* can properly use the fact that *p* as a reason in her practical or theoretical deliberations if and only if *S* knows that *p*.

For example, imagine that I am trying to decide where a group of us should go for dinner, and I am thinking about whether a new nearby restaurant would be suitable. The thought is that if I'm to use the fact that the restaurant has a good vegetarian selection as a reason to go there, I have to know it has a good vegetarian selection. If I don't know that, I'm not properly reasoning, rather I'm just guessing about where we should go.

Now that inset claim doesn't look much like an analysis of knowledge. But note that 'if and only if' is symmetric, so if that thesis is true, so is its converse.

> *S* knows that *p* if and only if *S* can properly use the fact that *p* as a reason in her practical or theoretical deliberation.

To be sure, most philosophers who endorse the truth of this thesis would say it could not be an analysis because it isn't particularly illuminating. They say that it is the fact that *S* knows that *p* that explains why *S* can use *p* as a reason, and not the other way around. But there is also a large thread in recent ethics, deriving in part from work by Derek Parfit (1984), on the idea that reasons are a relatively primitive explanatory notion, and that it is useful to explain other philosophical concepts in terms of reasons.

Now whether or not these biconditionals are true, and whether or not they are illuminating analyses in the relevant senes, are both huge questions. I'm raising them here largely to set them aside. The reason for raising them is to try and prevent the following inference: No analysis of knowledge that looks like the JTB analysis is working, so there is no analysis of knowledge. Perhaps there is a very different analysis which works, though it is well beyond the scope of this course to say whether it is.

1.3 Epistemology and the Social

At least since Descartes, the focus on epistemology in the Western European tradition has very much been on the individual. For Descartes, the ideal subject of epistemology is someone locked away in a quiet room, thinking hard about the world and their relation to it. There has been a substantial rebellion in the Western tradition against this individualistic focus in the last few decades, with a stronger focus by contrast on social epistemology.

We're going to look at two topics in particular. The first is the nature of **testimony**, and the second is the relationship between epistemology and questions about **justice**.

There are two closely related questions about testimony that will guide our discussion.

The first is whether testimony is in any way *special*. When we think about the ways in which people learn from other people, do we just have to take general principles about learning and apply them to the case of testimony, or do we have to theorise about it in a distinctive manner. The individualistic focus of much epistemology pushes towards a negative answer; other people are sources of information just like other things that measure the world. But a positive answer would perhaps undermine that focus.

The second is whether testimony is *basic*. Do we need to have reason to trust, or rely, on others in order to gain knowledge from their testimony? Neither answer to this question seems particularly easy to defend. On the one hand, saying that we can trust people for no reason whatsoever seems an invitation to gullibilism. On the other hand, saying that we cannot makes it somewhat mysterious how we could ever learn as much about the world as we actually do.

Finally, we'll look at some questions about epistemic **justice**, focussing on an important recent book by Miranda Fricker. In particular, we'll look at the intersection of two issues that she raises. One concerns *stereotypes*. On the one hand, there is something that strikes us as deeply wrong about reasoning like *He's from Liverpool, so he's probably a criminal*. On the other hand, reasoning from the existence of one characteristic to another that is statistically correlated with it is the paradigm of good empirical reasoning. And at least some stereotypes, not all but definitely some, are grounded in correlations that really obtain. So there is a delicate balance to be drawn here; what's the difference between bad, even pernicious, stereotyping, and good statistical reasoning?

Another question concerns the harms that arise from various kinds of epistemic prejudice. Consider a simple case. A wants to learn about the French Revolution, and sees that there is a course being taught on the French Revolution in the history department. So he enrolls for the course. But when he gets to the first lecture, he finds that the professor is a black woman. A is a racist and a sexist, so he doesn't think he could learn anything from a black woman, so he drops the course and enrolls in something less interesting. A's racism and sexism has obviously harmed A himself; he hasn't had a chance to learn something that was genuinely interesting. But has it also harmed the professor, and if so how? Once we see the idea that people are harmed by not being taken seriously as a source of information, what other work can it do in explaining features of the world, or perhaps revealing priorities for changing the world?

Our topics here reflect an interesting shift in recent philosophy. Traditionally, epistemology was thought to be the neighbouring discipline of metaphysics. (The very title of this course somewhat reflects that thought.) In contemporary philosophy, there are much tighter and deeper connections between epistemology and ethics than between epistemology and metaphysics. This shouldn't be surprising; questions about how one ought to think (epistemology) and how one ought to act (ethics) look like they should be related. And we'll treat them, at least for the final part of the course, as being very closely relatd.

Part I

Scepticism

Chapter 2

Pyrrhonian Scepticism

In this part, we'll be looking at arguments for scepticism. The primary kind of argument we'll be looking at is what is usually known as **academic** scepticism. But we'll start with two chapters on the other classical form of scepticism, often known as **Pyrrhonian** scepticism. This view traces back to Pyhhro, who lived from roughly 365–375 BCE, but the most important ancient discussion of it is due to Sextus Empiricus, who lived from roughly 160–210 AD.

2.1 The Pyrrhonian Sceptical Challenge

Pyrrhonian scepticism starts from the idea that everything we know, we come to know by some means or other. Every belief comes about through a belief-forming method. In some cases, the method might be something fairly unsophisticated. It might be that the method is the one of taking appearances at face value, or the method of believing what you would like to think is true. In some cases it might be indeterminate which of many methods you are using. (Am I using the method of trusting my appearances, or of trusting my eyes, or of trusting my eyes in good light, etc?) But this won't matter for the Pyrrhonian argument.

There is something odd about using a method that we do not have a reason to believe is reliable. Consider this exchange.

> A: It will snow tomorrow.
> B: Why do you think that?
> A: This machine said it will snow tomorrow.
> B: Why do you believe what the machine says?
> A: Oh, no reason. I just decided to use it.

This seems like it is bad. A can't really know that it will snow tomorrow because some random machine, which he doesn't have any evidence about, says it will snow.

It's worth distinguishing A's situation from more realistic situations, because we can easily understate the force of the Pyrrhonian reasoning if we aren't careful. If the machine A bought was labelled as a weather forecasting machine, and it was available for sale, then A has some reason to trust it. After all, there are laws against false advertising, so if the machine was completely fraudulent, there is some chance that it would have been barred from sale. Or if A had used the machine a few times in the past, and it had worked, then A would have some reason to believe it is reliable. Or if machines like this are in common use, and are often reliable, A would have a reason to trust it. (Compare the reason you have to trust the gas gauge in a car you just rented.) I don't want to insist at this stage that this would have been good reasoning on A's part. Indeed, the Pyrrhonian will deny that either is good reasoning. What I want to stress is that there is a difference between these cases and the case where A has no evidence whatsoever, and no reason whatsoever to believe the machine. In that case it does seem odd to think that A can know that it will snow simply by consulting the machine. Indeed, something stronger seems plausible. A doesn't have any reason to believe that it will snow tomorrow because some arbitrary machine says so. After all, it is trivially easy to produce another machine that says it won't snow. (Consider the beginning computer science exercise of producing a program that says "Hello world"; just change it to say "It won't snow tomorrow".)

Reflection on cases like this might suggest the following principle as plausible.

- A belief-forming method can only give a person a reason to believe something if the person has a reason to believe that the method is reliable.

But now we are on our way to a nasty looking sceptical argument. Assume that S believes that *p* using method M1. We just said that if this is to be at all reasonable, then S must have some reason to believe that M1 is reliable. Call that proposition *p1*. S must believe that, and the belief must be based on some method. Call that method M2. If M2 is M1, then it looks like we have something problematically circular. S believes something using a method, and believes that method because the method itself says it is plausible. Imagine if A backed up his belief by saying "Well, the machine says it is reliable". That doesn't seem any good. So M2 must be distinct. But if M2 is to produce reasonable belief, S must believe it is reliable. And that belief must be produced by some method M3. If M3 is M1 or M2 it seems we have the same kind of worry about circularity that we just saw. So M3 must be distinct from M1 and M2. And that means there is one more belief S must have, namely that M3 is reliable. And that must have been produced by some method M4. And for similar reasons M4 must be distinct from M1, M2 and M3, and S must believe it is reliable, by some new method M5, and so on.

The conclusion seems to be that S can't ever get going. So none of S's beliefs are at all reasonable! It's worth noting how radical a conclusion this is. When we get to the Academic sceptics, we'll worry about things like the possibility you are currently dreaming. Maybe you can't know that you're awake right now. Let's grant that, just for the sake of argument. It's still consistent with that that you know something, namely how things seem to be, either in a dream or in reality. And it's consistent with that that you can reasonably believe you're awake, even if this reasonable belief does not amount to knowledge. But the Pyrrhonian sceptic suggests that you can't have even a reasonable belief that you're awake, and in fact you can't even have a reasonable belief that you're having a certain appearance right now. That's a radical kind of scepticism!

2.2 Responses to the Pyrrhonian Challenge

There are, as Sextus noted, three prominent ways of getting out of the sceptical argument here. Let's label them first, then say what they are in more detail.

1. Infinitism
2. Coherentism
3. Foundationalism

The infinitist response is the easiest to state. It says that there simply isn't a problem here. Yes, we must have an infinity of beliefs; that M1 is reliable, that M2 is reliable, and so on. And each of them must be grounded in a way that is dtsinct from the others. But there is nothing wrong with this. Most people's reaction to this view is that it is obviously crazy; if that is what we have to say the Pyrrhonian has already won. This is too quick. Peter Klein has shown how to make infinitism plausible. (See the last section of (Klein, 2013) for more references.) But we will set this option aside.

The coherentist says that after a while, the argument that the methods must be distinct gets too weak to do the work the Pyrrhonian needs. It's true, they'll concede, that M2 should be different to M1. It doesn't help to have one method that endorses itself. But this doesn't mean that the same argument works all the way up the chain.

This kind of thought can be supported by reflection on our senses. It would seem objectionably circular to test our eyes by using our eyes. (Though we'll come back to that point in a bit.) But imagine the following way of checking that your eyesight is working. You see what looks like a wall. You knock on the wall, and feel the wall against your hand just when you see your hand touch the wall. Simultaneously you hear a sound of a hand hitting a wall. This shows that, in a small respect, your sight, touch and hearing are cohering with one another. And this coherence can make you more confident that each is working.

Or take a more general case. We usually form beliefs by a combination of direct observation, projection of past regularities, inferring from some data to its best explanation, testimony from friends, and expert opinion. In practice, these methods usually point in the same direction. Or, more carefully, to the extent that they point at all, they point in the same direction. And we have a history of them all pointing in the same direction. On any occasion, if one of them appeared completely different to the other four, we would discount it. (This is even true for visual appearance. If the other four sources tell me something is a visual illusion, I will believe that current appearances are not accuracte.) We don't have to answer the question of which of these methods is prior to the others, because we're constantly in the process of checking each of them against the other.

But the problem is that coherence is too easy. Again, take an everyday example, the kind of theories that are usually derided as 'conspiracy theories'. If you want a particular example to concentrate on, think of the theory that the moon landings were faked.[1] The thing about these theories is that they tend to cohere extremely well. If you push a proponent of them on one part, they'll tell you something else that makes perfect sense within the theory. The 'something else' might strike you as completely crazy, but it's hard to deny that the story hangs together.

And that's the general problem. A collection of crazy theories that hang together well is still crazy. So it seems that mere coherence cannot be a reason to believe a theory. So even if the theories that, say, M1 through M10 are reliable are coherent, each is well supported by the others, that fact alone doesn't seem to make it reasonable to believe them. They might collectively form a crazy conspiracy theory.

2.3 Foundationalism

So the most popular response to the Pyrrhonian is a form of **foundationalism**. The foundationalist thinks that there are some methods that can be relied upon without having prior justification for relying on them. These methods are the foundations of all of our other beliefs.

Different foundationalists offer different explanations of what the foundations are,. Here are some candidate methods that have been proposed as foundational.

- Trusting your senses as a guide to the outside world, e.g., believing that there is a table in front on you in a case where there actually is a table in front of you, and you see it.
- Trusting your introspection, e.g., believing that you are in pain when you can actually feel the pain.
- Trusting testimony, e.g., believing what someone says in the absence of reasons to the contrary.
- Trusting basic logic and arithmetic, e.g., believing that if there is one thing here, and one thing there, then there are at least two things.
- Trusting your memory, e.g., believing that you were at the football last weekend on the basis of a memory of being there.

[1] I mean to focus on extreme theories, but it can be tricky to say in general what makes a theory extreme. The theory that the CIA and Mafia conspired to invade Cuba has most of the characteristics of an extreme conspiracy theory, and is basically true.

Some foundationalists say that the foundations are **indefeasible**. That is, they say that if M is a foundational source, and someone believes p on the basis of M, then there is no other evidence that can make this belief unreasonable. Other foundationalists say that foundations are merely **defeasible** grounds for belief. That is, they say that even the output of foundational methods can be overturned. What makes the methods foundational, though, is that they can be trusted in the absence of independent reason for doubting them. And, typically, these foundationalists will say that such reasons for doubt do not come easily. The more defeasible one makes the foundations, the more plausible it is that there is a very wide variety of foundational methods. It is crazy to think that anyone telling us that *p* provides a knock-down reason to believe that *p* is true. It isn't crazy to think that testimony that *p*, in the absence of reason to believe either that *p* is false or that the person telling us this is untrustworthy, provides a sufficient basis for belief that *p*.

Of course, the foundationalist needs a response to the Pyrrhonian sceptical argument. The best response, I think, is to say that it equivocates in a certain way. The first of the following principles is true, the second of them is false.

- For S to get a reasonable belief that *p* by using method M, there must be some reason that M is reliable.
- For S to get a reasonable belief that *p* by using method M, S must have a reason to think that M is reliable.

The first of these could easily be true, while the second is false. Consider the beliefs that an infant gets by visual perception. There is a good reason this is reliable. The infant is the product of evolution, and if humans didn't have accurate visual perception, they would have died out. Now the infant can't possibly know this, or even understand the nature of evolution. But no matter. What's important is that there is a good reason, available to the theorist, that the infant's beliefs are reliable. It doesn't matter whether the infant herself is aware of this.

Chapter 3

Easy Knowledge

3.1 Cohen's Challenge

In recent years, this kind of foundationalist response to Pyrrhonian scepticism has come under sustained fire from Stewart Cohen (2002, 2005). Cohen's position is motivated by stories like these two.

> Suppose my son wants to buy a red table for his room. We go in the store and I say, "That table is red. I'll buy it for you." Having inherited his father's obsessive personality, he worries, "Daddy, what if it's white with red lights shining on it?" I reply, "Don't worry-you see, it looks red, so it is red, so it's not white but illuminated by red lights." Surely he should not be satisfied with this response. Moreover I don't think it would help to add, "Now I'm not claiming that there are no red lights shining on the table, all I'm claiming is that the table is not white with red lights shining on it". But if evidentialist foundationalism is correct, there is no basis for criticizing the reasoning. (Cohen, 2002, 314)

> Imagine my 7 year old son asking me if my color-vision is reliable. I say, "Let's check it out." I set up a slide show in which the screen will change colors every few seconds. I observe, "That screen is red and I believe that it is red. Got it right that time. Now it's blue and, look at that, I believe its blue. Two for two…" I trust that no one thinks that whereas I previously did not have any evidence for the reliability of my color vision, I am now actually acquiring evidence for the reliability of my color vison. But if Reliabilism were true, that's exactly what my situation would be. (Cohen, 2005, 426)

The theories that Cohen mentions at the end of each anecdote, evidentialist foundationalism and Reliabilism, are forms of foundationalism in the sense we've been using the term here. They are both theories that say that we can stop the Pyrhhonian regress by using a method without justifying it. (The theories are very different, and we'll come back to the differences below.)

3.2 The Easy Knowledge Argument

Cohen's stories are meant to induce a kind of incredulity, and I think they're quite effective at doing that. But it isn't hard to extract arguments from them. Here's one way of doing this. Assume that M is a method with the following three characteristics.

- M is foundational in the sense that S can use it to produce reasonable beliefs without having a reasonable belief that M is reliable.
- S does not actually have reason to believe M is reliable.
- S can tell, at least in a wide range of cases, that M is the source of her beliefs that are produced by M, and more generally is reasonably reliable about what M is saying at any given time.

The last condition is meant to distinguish between people like us, who know when we get a belief via, say, vision, and babies, who may not always know how they are getting information. (Of course, even adults quickly forget the source of their information, but we can often tell what the source is while thinking about it.)

Now imagine that all this is true, and S uses method M to get information *p1* and *p2*, and knows that she's doing all this. Then she can reason as follows.

1. *p1* (Via method M)
2. M says that *p1*. (By the assumption that S can reliably tell what M says.)
3. M is correct about whether *p1* is true. (from 1, 2)
4. *p2* (Via method M)
5. M says that *p2*. (By the assumption that S can reliably tell what M says.)
6. M is correct about whether *p2* is true. (from 4, 5)
7. So we have some evidence that M is generally reliable. (from 3, 6)

I've only listed two cases where M worked, but obviously that problem could be finessed by using M over and over again. So that shouldn't be the problem with step 7. But surely step 7 is absurd. This is 'easy' knowledge that M is reliable. But telling that a method is reliable can't be this easy. So something in the foundationalist picture must have gone wrong.

3.3 Responses to Cohen's Argument

I'll go over five possible responses on behalf of the foundationalist.[1] None of these is clearly correct, but I think between them we can see some ways out for the foundationalist. Here are the four options.

[1] These don't exhaust the responses. For more responses to Cohen and to related arguments, see Weisberg (2012) and Pryor (2013).

1. Say that access to M undermines the efficacy of M.
2. Deny that this is a good use of induction.
3. Deny that there is a problem here.
4. Say that Cohen's arguments don't generalise to all kinds of foundationalism.
5. Say that the defeasibility of M solves the problem.

Let's take these in turn.

3.3.1 Access and Undermining

There is something odd about the general structure of the puzzle here. It's meant to be a puzzle for people who are using a foundational method, and know which method they are using, but have no evidence about the reliability of that method. That is, to a first approximation, never the situations humans find themselves in. For most foundational methods, we start using them as infants. At that time we may not even have the concept of a method, and certainly aren't in a position to actively think about the reliability of our methods. By the time we have the capacity to think those thoughts, we have lots of evidence that the method works.

So the following position is both coherent, and consistent with humans as we find them having a lot of justified beliefs. It's fine to use a method without knowing it is reliable, as long as you don't know that's the method you're using. In that situation, you won't be able to make any problematic circular inferences, as in the easy knowledge argument. If you later use the collected evidence you got in infancy to tell that M is reliable, that won't be problematically circular.

This response does get the data right, but there's something unsatisfying about it. Why should it be that knowing you are using M undermines the force of M? Perhaps this question has an answer, but without it, the response seems insufficient.

3.3.2 Induction and Evidence Collection

There is something fishy about step 7 in the reasoning, and not just because of the small sample size. In general, projection from a sample requires that the sample be representative. So going from the premise that M has worked in all these cases, to the conclusion that M generally works, requires that 'these cases' be a representative sample. And that isn't obviously the case.

Let's think about an example where this requirement is not met. Imagine that the Olympics are on, and I know that the University of Michigan newspaper is focussing on what competitors from the local area do. Their rule for publication, I know, is that they will print the result of any event where a current Michigan student competes, or an event where a Michigan alumnus wins a medal. I read the Michigan newspaper, and no other news about the Olympics. And while I see lots of current students competing without winning a medal, every alumnus I see reported on wins a medal. It would be crazy to infer from that that every Michigan alumnus who competed at the Olympics won a medal, even if I have a very large sample. It would be crazy because the sample reported in the University of Michigan newspaper is obviously not representative. Only the ones who won medals were ever going to be reported.

The same thing is happening in the easy knowledge argument. If M makes a mistake, the method I'm using (which only relies on M and my ability to detect what M outputs) will never detect it. And I shouldn't make inductive generalisations from a one-sided sample like that.

I think this is a perfectly good objection to the last line of the reasoning, and one that doesn't rely on foundationalism being false. But I also don't think that it really solves the problem. After all, line 3 is pretty strange too. It's odd to think that you could tell M is working on even one occasion by using M. At least, it seems odd at first. Maybe that should be questioned though.

3.3.3 Biting the Bullet

The phrase 'biting the bullet' is usually used, in philosophy and elsewhere, for the action of simply accepting what strikes most people as an absurd consequence of one's view. And we should think about how plausible the bullet-biting strategy is in this case. Here's one way to make it more palatable.

What's the best way to tell that something works? Run it a few times, and see it gets the right results. That's what one does in the easy knowledge case. And the method works. So we infer that it's reliable. What's wrong with that?!

My view is that at this stage we need to distinguish between different kinds of foundationalism. In particular, we need to distinguish between **internalist** and **externalist** varieties.

The internalist foundationalist says that we can tell 'internally' whether the method we are using is one of the foundational ones. For these methods it is usually hard to impose by fiat any kind of success constraint on them. (Hard, but not impossible; see the next subsection.) A victim of an evil demon could really be using the good methods, although they are not at all reliable or even successful. And if this is how we understand methods, then the bullet biting strategy does seem implausible. (What Cohen calls the 'evidentialist foundationalist' is a kind of internalist in this sense.)

The externalist foundationalist says that whether a method M is foundational depends on facts external to the subject. So it might be part of our philosophical theory that a method is foundational only if, as a matter of fact, it is reliable. S can't tell from the inside whether her visual perception, say, is reliable. But if it is actually reliable, she can reasonably use it - though she might not be able to tell in advance whether she is being reasonable in using it. (What Cohen calls reliabilism is externalist in this sense.)

If we put this kind of reliability condition on M, the bullet biting strategy gets a little more plausible. If an agent really is using a reliable strategy, and comes to have a reasonable belief that it is reliable in part because it is actually reliable, then it isn't clear what's gone wrong.

Still, there is a whiff of circularity here. The next two responses can be thought of as ways to reduce this 'whiff'.

3.3.4 Kinds of Foundations

Cohen's stories both involve the assumption that visual perception is one of the foundational methods. But it is clear from his discussion that he does not mean to assume that the foundationalist thinks this. He wants to argue against all forms of foundationalism. But not all foundationalists think that visual perception is foundational. And this might affect the argument.

Some foundationalists think that appearances, not perceptions, are foundational. So if I look at a table, the foundational method is 1, not 2.

1. From how things appear to me, come to believe that it looks like there's a table in front of me.
2. From how I see things, come to believe that there is a table in front of me.

We can run an 'easy knowledge' argument against the first type of foundationalist. Oh look, it looks like it looks like there's a table in front of me, and it does look like there's a table in front of me. I'm good at telling how things look to me. Oh look, I feel like I feel like I'm in pain, and I do feel like I'm in pain. I'm good at telling how things feel to me. Does this seem objectionably circular? If not, you might think that Cohen doesn't have an argument against foundationalism here, as much as an argument against a certain kind of foundationalism. That is, what Cohen really gives us is an argument against taking perceptions of the *external* world to be foundational. He doesn't have an argument against taking perceptions of the *internal* world to be foundational. But the general anti-foundationalist conclusion he draws, and that the Pyhhronian needs, requires both.

3.3.5 Defeasibility

There's another aspect of the foundationalist picture that might be relevant to responding to Cohen's challenge. Remember that many foundationalists say that foundational methods are **defeasible**. If you have a positive reason to think that M is unreliale, then you can't get reasonable beliefs from using M.

So that suggests that there's a two part response to the 7 year old. If there is a positive reason to think that the lighting might be deceptive, or in general that M is unreliable, then even the foundationalist should simply concede that we need independent grounds for checking M. That is, the imagined response to the 7 year old that Cohen gives is inappropriate, even by foundationalist lights. On the other hand, if there is no reason to think the lighting might be deceptive, that's what one should say to the 7 year old. It is silly to use M to confirm M. What you should say is that there's no reason to worry about trick lighting. And you should say that because, by assumption, there is no reason to worry about trick lighting.

Part of what makes Cohen's argument so hard to respond to, part of what makes it such a good argument, is that things like trick lighting are right on the border between things we do have good reason to worry about, and things we don't. Trick lighting happens; it's not like evil demons controlling your entire life. But we usually get by safely assuming that the lighting is normal enough. Once we decide which side of the line this possibility falls on - realistic enough that we need a reason to respond to it, or unrealistic enough that we don't - the puzzle can be solved.

3.4 For Next Time

We'll move onto a discussion of academic scepticism. But we won't leave the issue of easy knowledge behind entirely. In particular, we'll come back to it at two points. It will come up when we consider the 'dialectical' argument for scepticism. And it will come up when we think about how we can know we're not in a sceptical scenario.

Chapter 4

Demons and Disasters

4.1 Academic Scepticism

Academic scepticism takes off from fear of certain scenarios, what we'll call **epistemic disaster scenarios**. For Descartes (1641/1996b), the disaster scenario was an evil demon who deceived him into thinking that there is an external world.[1] In Descartes's scenario, we are all immaterial souls, who are being given signals as of a non-existant world. In the 20th century, lots of philosophers and science-fiction writers worried about the brain-in-a-vat scenario. In this scenario, you are a brain in a vat of nourishing liquids, given just the electrical stimulation you need to feel as if you are seeing, touching, etc an external world. And while there is a world out there, it is not the one you think it is; you are a disconnected brain sitting (in a vat) on a lab bench. There is a version of this scenario in the movie *The Matrix*.

Bertrand Russell (1921/2008) suggested a very different disaster scenario.

> There is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it then was, with a population that "remembered" a wholly unreal past. There is no logically necessary connection between events at different times; therefore nothing that is happening now or will happen in the future can disprove the hypothesis that the world began five minutes ago. (Russell, 1921/2008, Lecture IX)

This scenario doesn't threaten our knowledge that there is an external world, but it does threaten our knowledge that we woke up and showered this morning. Along similar lines, David Hume (1739/1978) worried about scenarios where the future behaves radically different to the way the past has behaved. Again, there is no threat of the external world failing to exist, but it does suggest a scepticism about what we can know about the future.

[1] Descartes's importance to the history of academic scepticism is so great that this kind of scepticism is often called Cartesian scepticism in his honor. But I'll stick with the older name.

It's worth reflecting a bit on what makes these disaster scenarios so useful for the sceptic. Why is the evil demon, or the brain-in-a-vat so helpful to the sceptic's cause? Or, put another way, if you wanted to come up with a new disaster scenario, in what ways would it have to resemble the 'classic' disaster scenarios? It seems to me that part of what is so special about disaster scenarios is that they have so many features that are useful to the sceptic. Here are a few things we might say about a disaster scenario, noting that several of these claims are theoretical claims that we might want to retract on closer theorising.

- An agent in the real world would have exactly the same beliefs were she in the disaster scenario, but in the disaster scenario those beliefs would be false.
- An agent in the real world and her counterpart in the disaster scenario have the same evidence, but the agent in the disaster scenario does not know some things we take to be central to our knowledge of the world.
- If an agent, even an agent in the real world, were confronted with someone who thought that they might be in the disaster scenario, there is no non- circular argument the agent could give to convince her friend that they were in the real world.
- None of the methods by which we ordinarily seem to get knowledge seem like methods that an agent could use to acquire knowledge that she's not in the disaster scenario. Deductive methods seem too weak to prove a contingent claim, such as the claim that the agent is not in the disaster scenario. And non-deductive methods seem to presuppose that the agent is not in the disaster scenario, so cannot be used to show that she is not in it.

4.2 Varieties of Sceptical Argument

Each of these points can be used to generate an argument for scepticism from the disaster scenario. In what follows, we'll use *S* for the name of the agent in question, *SH* for the sceptical hypothesis, i.e., the hypothesis that *S* is in the disaster scenario, *DS* for the disaster scenario, and *O* for an ordinary proposition that we ordinarily take S to know, e.g., that she has hands, but which is false in the disaster scenario. Then the following five sceptical arguments all have their appeal.

Sensitivity Argument

1. Were *DS* actual, then *S* would still believe $\neg SH$, but that belief would be false.
2. If her belief that $\neg SH$ is to constitute knowledge, then in must be *sensitive*. That is, it must be such that it wouldn't be held if it weren't true.

C. *S* does not know that $\neg SH$.

Circularity Argument

1. If *S* were challenged to provide a reasonable, non-circular argument that $\neg SH$, she could not provide one.
2. Anything that an agent knows, she can offer a reasonable, non-circular argument for if challenged.

C. *S* does not know that $\neg SH$.

Indiscriminability Argument

1. For *S* to know that she is not in *DS*, she must be able to discriminate the disaster situation from the actual scenario.
2. *S* cannot discriminate her actual scenario from the disaster scenario.
3. Indiscriminability is symmetric, i.e., for any two situations o_1 and o_2, if *S* cannot discriminate o_1 from o_2, she cannot discriminate o_2 from o_1

C. *S* does not know that she is not in the disaster scenario, i.e., she does not know $\neg SH$.

Methods Argument (Quantified)

1. There is no means by which *S* could know $\neg SH$.

C. *S* does not know $\neg SH$.

Underdetermination Argument

1. *S* has the same evidence in the disaster scenario as in the real world.
2. Knowledge supervenes on evidence. That is, two agents with the same evidence, even in different worlds, have the same knowledge. So if *S* has the same evience in the disaster scenario as in the real world, she only knows things in the real world that she knows in the disaster scenario.

C. *S* does not know that *O*.

Each of these five arguments, especially the fourth, could do with some tightening up of the premises. Among other things, I've left tacit the premise that knowledge is factive, and that *O* is false in the disaster scenario. But I think they're useful enough to start with.

Note that the arguments do not all have the same conclusion. The first four conclude that *S* does not know that $\neg SH$. But one of them, the underdetermination argument, concludes that *S* does not know that *O*. We can convert the other four arguments into arguments that *S* does not know that *O* in two ways. First, we could try to defend the following premise.

- If *S* knows *O*, then *S* knows that $\neg SH$.

But that premise is actually fairly implausible. If *S* is a young child, she might well have never considered *SH*, so it is not at all obvious that she knows $\neg SH$. So this way of modifying the arguments seems to introduce a false premise. A better alternative is to add the following premise.

- If *S* knows *O*, then *S* could know that $\neg SH$.

The thought is that although *S* might not have considered *SH*, if she knows *O*, she is at least in a position to infer $\neg SH$, by simple deduction from *O*. And that very plausible claim is equivalent to this one, which looks like it will be helpful.

- If *S* could not know $\neg SH$, then she does not know *O*.

But now we face a problem. What the earlier arguments (at least the first four) had concluded was that *S* **did** not know $\neg SH$. We have here a conditional that gets triggered if *S* **could** not know $\neg SH$. Can we overcome this gap?

In practice, this looks like a fairly easy hurdle to get around. The premises of the first, third and fourth arguments don't just show that, as a matter of contingent fact, *S* is ignorant of whether she's in *SH*. They seem to posit deep philosophical difficulties that get in the way of ever coming to know whether she's in *SH*. So if we accepted those arguments, we should probably find the stronger argument that *S* could not know whether she's in *SH* plausible too.

The arguments above are fairly generic. They can be, and have been, used to argue for scepticism about the external world. But they also can be, and in many cases have been, used to argue for narrower sceptical hypotheses. For instance, leting the disaster scenario be the Russell-world, where everything just popped into existence a few minutes ago, we could imagine using the underdetermination argument to show that we don't know that the past exists.

More restricted forms of scepticism seem more plausible in general. Indeed, several restricted forms are even held by a number of philosophers. (I'm sympathetic to a reasonably strong sceptical view about consciousness in non-human animals.) So you might think that these arguments will be more helpful to a 'limited' sceptic, who simply wants to argue for scepticism about other minds, or about unobserved scientific entities, rather than scepticism in general.

The problem with this approach is that the second premise in each of the first three arguments states a fairly general philosophical claim. And that premise is no more plausible when it is restricted in some way than it is in general. If you deny in general that knowledge supervenes on evidence, you're not likely to be more sympathetic to the view that inductive knowledge supervenes on inductive evidence. It's logically possible that the restricted supervenience thesis holds, but really the only reason we would have to believe that the restricted supevenience thesis holds is that the stronger thesis holds. So most of these arguments look like they'll either provide an argument for a sweeping version of scepticism, or they won't help even the restricted sceptic. There are exceptions to these generalisation, exceptions that we'll come back to, but I think these will genuinely be exceptions; the rule is that the arguments work for a very strong sceptical conclusion, or they don't.

4.3 Why Care about Scepticism?

One common response to any detailed work on scepticism is that it's all a waste of time, since it is completely obvious that scepticism is false. This kind of position is sometimes associated with Wittgenstein, or at least with modern Wittgensteinians. I think there's a grain of truth in this position, but that the conclusion is radically mistaken.

It's true that scepticism is obviously false. Or, at least, it is true that the most general forms of scepticism are false, and clearly so. But this doesn't mean that we have nothing to learn from thinking about scepticism. Indeed, there are four things that we might be interested to discover from thinking about sceptical arguments.

One is whether scepticism is true anywhere, and if so where it is true. Now *some* kind of scepticism is sure to be true; there is lots we simply cannot know. But there are other kinds of scepticism that are more plausible. Perhaps we should endorse scepticism about the future, or about other minds, or about non-human minds, or about ethics, or about scientific unobservables, or about mathematical questions left unsettled by ZFC. At least it isn't obvious that *all* these forms of scepticism are false. And I think that examining sceptical arguments might help us work out which of these forms of scepticism is true.

A second is that in some cases, it is not obvious which premise in a sceptical argument is false, and it is of philosophical interest to figure out which one it is. For instance, I think that the underdetermination argument fails because the supervenience premise is false. Timothy Williamson has argued at length that this premise is true, and argues that 'same evidence' premise is the false one. It's interesting to work out which of these positions is right, independent of whether we think the conclusion is plausible. We'll come back to this question in chapters 14 and 15.

A third is that thinking about the sceptical arguments gives us a way to respond to philosophical arguments that trade on the fear of scepticism. The philosophical literature is full of 'transcendental' arguments, that say that since we know a lot of things, such-and-such theory must be true. Another way to put this kind of argument is that the theory in question is the only thing that stands between us and scepticism. Famously, many forms of idealism in metaphysics have been supported on these grounds.

Once we present scepticism as an argument (or as a series of arguments), we can ask just which premise the target philosophical theory is supposed to help with. I think in many such cases the answer is that the theory doesn't help with any premise of any plausible sceptical argument. So thinking carefully about scepticism can help defuse a philosophical weapon that others may use against us.

And finally, once we have set out the sceptical arguments clearly, we can try running some transcendental arguments of our own. That is, we can see which philosophical positions carry with them commitments to each of the premises of a particular sceptical argument. A lot of philosophers have held, over the years, that necessary truths (of a certain kind) can be known without evidence, but any truth that is not necessary in this way requires evidence to be known. I think that position leads to scepticism, and we'll come back to why in later chapters. The arguments there will be complex. If we did just quickly dismiss scepticism, we wouldn't have the resources to evaluate whether they are true. So while scepticism itself is implausible, thinking through carefully why it fails can be relevant to debates where both sides (or many sides) have plausible positions.

Chapter 5

Scepticism and Intuitions

Let's step back a bit from the five sceptical arguments I started with. As Peter Klein (2013) notes, the canonical form of the academic sceptical argument has this structure.

1. *S* cannot know $\neg SH$.
2. If *S* knows *O*, then she can know $\neg SH$.

C. *S* does not know *O*.

What could motivate this argument? We might start with a raw intuition that both premises are true. After all, as soon as one raises the disaster scenarios, the threat of illusion seems immediate. And premise 2 is a very simple statement about how knowledge, if it can be obtained, can be extended by deduction. And the conclusion follows immediately.

It seems to me, though, that this is actually a fairly weak argument. The standard reason for thinking so traces back to some comments by G. E. Moore (1959). There may well be some intuitive support for the premises of this argument. But there is an even stronger intuition that the conclusion is false. I know that I have hands, that I live in Michigan, and so on. If I have to weigh this intuition against some fairly subtle intuitions about illusions, and about deduction, I feel I should hold onto the intuition that I know a lot.

Let's put this point another way. As you probably recall from other philosophical work, a valid argument is one with the following feature: *It's impossible for the premises to be true and the conclusion false*. That's equivalent to saying the following: *If an argument is valid, then the set consisting of all the premises, and the negation of the conclusion, cannot be all true*. And that suggests that any valid argument with intuitive premises, and an unintuitive conclusion, is really a *paradox*, a set of intuitively true claims that cannot be true together. It might be helpful to write out the sceptical argument as a paradox like this. The paradox is that all three of these claims are (allegedly) intuitive, but they cannot all be true.

1. *S* cannot know $\neg SH$.
2. If *S* knows *O*, then she can know $\neg SH$.
3. *S* knows *O*.

We can convert that impossibility claim into one of three different arguments, by taking two of the claims as premises, and the negation of the third as conclusion. We've seen one of these arguments already; here are the other two. First, an argument to a pro-knowledge conclusions.

1. If *S* knows *O*, then she can know $\neg SH$.
2. *S* knows *O*.

C. *S* can know $\neg SH$.

And an argument to a conclusion about closure.

1. *S* can know $\neg SH$.
2. *S* does not know *O*.

C. It's not the case that if *S* knows *O*, then she can know $\neg SH$.

If the sceptic is to rely on intuitions, she must not just show that the premises of her preferred argument are intuitive. She must show that this is a better argument than either of the two arguments we can get from the paradoxical set. And I just don't see how that could be so. We have to give up one claim out of the three. Why should it be the very obvious claim that we know we have hands? I don't see a reason for this yet.

This doesn't mean the sceptical argument is weak. It just means it needs something stronger than a raw intuition to back it up. As we saw in the last chapter, there are lots of ways to do that, so this isn't a problem for the sceptic. But the sceptic owes us an argument here, not just an intuition.

5.1 Scepticism and Closure

One of the possible moves to make in response to the paradox of the last section is to deny the conditional: *if S knows O, then she can know* $\neg SH$. In doing this, we are denying that knowledge is **closed** under competent logical deduction. Sometimes, an agent can know something, competently deduce something else from it, but not know the conclusion she deduces. That's not particularly intuitive, but perhaps there aren't any intuitive options around here.

This was the response that Robert Nozick (1981) recommended we take in response to the sceptic. He thought that we could, roughly, know all and only the things we sensitively believed. A belief is sensitive just in case you would not have it were it not true. Since we can sensitively believe that we have hands, but any belief that we're not in *DS* will be insensitive, there is something we can know (namely, that we have hands) even though we can't know one of its obvious consequences.

This has led some philosophers to suggest that the sceptic would be best off doing without this closure premise. If there is a sceptical argument that doesn't rely on the claim that knowledge can be extended by competent deduction, then this kind of Nozick-like response will be blocked. In general, it's a good idea to try and formulate one's arguments with as few premises as possible, since that leaves fewer points where one can go wrong. So in principle this could be a good idea.

Anthony Brueckner (1994) has argued that the sceptic really needs to do away with the closure premise. Summarising a lot, he argued that the sceptic should do this becase the following three claims are true:

- Sensitivity based motivations for the first premise of the sceptic's argument, that *S* cannot know $\neg SH$, are incompatible with the closure premise.
- Underdetermination based motivations for the first premise of the sceptic's argument work just as well to motivate the conclusion of the argument, without any appeal to a closure premise.
- There are no other good motivations for the first premise of the sceptic's argument.

We'll spend most of the rest of this chapter on the first point, then much of the rest of this part of the course on the second and third points. Note that we've already seen a sceptical argument that didn't rely on a closure premise: what we called the underdetermination argument. That argument directly concluded that *S* did not know that *O*, without any closure step. And that's what's relevant to Brueckner's argument.

5.2 Scepticism and Sensitivity

Can the sceptic motivate the first premise via sensitivity concerns? Brueckner notes one particularly important point here. Any motivation the sceptic gives for the first premise had better not undermine the second premise. And there is a danger here for the sceptic. After all, Nozick's analysis of knowledge, which relies on sensitivity, does end up rejecting the second premise.

But it isn't the case that the sensitivity motivation is inconsistent with the second premise. We need to distinguish these two claims:

- Having a sensitive belief is necessary and sufficient for knowledge.
- Having a sensitive belief is necessary for knowledge.

If you believe the first of these claims, then you'll accept the first premise of the sceptic's argument, but not the second. As we've noted many times now, the property of being a sensitively held belief is not preserved by competent deduction. If I start with the (sensitively held) belief that I have hands, and draw the conclusion that I'm not an immaterial being, then I'll derive an insensitive conclusion.

But the sceptic shouldn't say that. Rather, the sceptic should just say that sensitivity is one condition, among many, for knowledge. The sceptic need not be in the position of providing a full account of knowledge, just starting with some intuitive constraints on knowledge, and showing we cannot satisfy them. And sensitivity is (at least before we look through complaints like Kripke's), an intuitive constraint. And it is logically compatible with the idea that knowledge can be extended by competent deduction. So the sceptic isn't being inconsistent here.

Brueckner has another way of developing this point. He says, "If knowledge is closed under known entailment, then it seems quite plausible that each necessary condition for knowledge must also be so closed." But this is not plausible at all. It isn't true in general that if some property is closed under entailment, then all the necessary conditions of that property are so closed. Consider the property of being necessarily true. As a matter of logic, any proposition with that property has the property of being non-contingent, i.e., of being either necessarily true or necessarily false. So we might say that being non-contingent is a necessary property of being a necessary truth. But while being necessarily true is closed under entailment, being non-contingent is not. Let p be any contingent proposition. Then $p \wedge \neg p$ entails p. (That's hard to deny; any conjunction entails each conjunct.) But $p \wedge \neg p$ is non-contingent; it is necessarily false. On the other hand, p is contingent, by hypothesis. It could be that the same is true for knowledge; it is closed under known entailment, while one of its necessary conditions is not so closed.

So I don't think the sceptic's appeal to sensitivity is self-undermining, in the way that Brueckner worries that it is. On the other hand, I do think the appeal to sensitivity is particularly implausible. It seems to imply that we lose knowledge at exactly the wrong spot.

Assume, as we have on behalf of the sceptic, that sensitivity is necessary but not sufficient for knowledge. And assume (contra the sceptic) that we can know we have hands. The sceptic won't accept this, but what we're trying to show is that sensitivity is not intuitively a constraint *we* should accept. Now imagine the thinker going through the following four steps.

1. I have hands.
2. I'm not handless.

3. For all *X*, I'm not a handless *X*. That is, I'm not a handless human, and I'm not a handless jellyfish, and I'm not a handless ghost, and I'm not a handless dinosaur, and so on.
4. I'm not a handless brain in a vat.

Now her first belief, and her logically equivalent second belief, are sensitive. If she'd lost her hands, she'd know that. And her third belief is sensitive as well. If it were false, there would be some *X* such that she isn't a handless *X*. And if that were true, it would presumably be because she were a handless human. I'm assuming here that the following is true of human beings.

- If that human didn't have hands, they would be a handless human, not a handless ghost, or dinosaur, or anything else.

But that seems entirely right; losing one's hands would not make one change species. So if the thinker were handless in some way at all, she'd be a handless human. But if she were a handless human, she'd know that. So her third belief is sensitive too.

Insensitivity kicks in only at the very last step. If the thinker were a handless brain in a vat, she wouldn't know that. But this seems a very strange place to draw the boundary between knowledge and ignorance. The thinker can apparently know that for any *X* whatsoever, she isn't a handless *X*. But if she tries to apply that knowledge to infer she's not a handless brain in a vat, she won't acquire knowledge. This seems bizarre; if we know that something is true for any *X* whatsoever, then we can know that it is true for a particular *X*.

The key point here is to distinguish between the following two claims:

- Sometimes, competent logical deduction does not extend one's knowledge.
- Adding a sensitivity constraint to knowledge explains why sometimes, competent logical deduction does not extend one's knowledge.

I happen to think the first is false, so there's nothing to explain. But even if you liked the first claim, you shouldn't like the second. Sensitivity based explanations get the knowledge failures appearing in completely implausible points.

So I agree with Brueckner's conclusion that the sceptic shouldn't appeal to sensitivity, although not entirely with Brueckner's reasons. Next time, we'll look at three other ways the sceptic might motivate premise 1.

Chapter 6

Dialectic and Discrimination

6.1 Dialectic

Here's something that seems clearly true about the sceptical situation. We can't argue our way out of it without begging any questions. Conisder a debate between a sceptic, S, and an ordinary thinker, O.

> O: I know I have hands.
> S: How do you know that? You could be the victim of an evil demon.
> O: I can see my hands.
> S: That could be an illusion.

And so on. Anything O says will be dialectically ineffective. Anything she can say will be explained away by the sceptic. She can't win.

To convert this into an argument, we need some extra premise. It isn't immediately a sceptical conclusion to say that we can't convince the sceptic. The problem would be if we can't convince ourselves. The obvious way to generate an argument is by the following premise.

- To know that *p*, you have to be able to give an argument that is convincing to someone who rejects *p*, and this argument must not be question-begging.

One interesting feature of a premise like this is that it shows why the sceptic might want to reach straight for extreme scenarios, like that I'm the victim of an evil demon. Compare scepticism about whether I have hands. My knowledge that I have hands is consistent with this principle. After all, I can convince *someone* who doesn't believe I have hands. If a colleague believes I lack hands because she thinks I've had a nasty accident, I can convince her otherwise by simply showing my hands. What's distinctive about the proposition that I'm not the victim of a demon is that anyone who disbelieves that, who thinks I am the victim of a demon, will be beyond convincing.

But really there's no good reason to accept a principle as strong as this connecting knowledge and convincingness. Knowing something is one thing, being able to convince others of it is another. Here are two kinds of examples that make this point.

Jones is widely admired as the most upstanding, honest, member of his community. One day, Smith sees Jones stealing money from a charity box. It's clear as daylight that it is Smith; Jones has good vision, and sees Smith in the act clearly. But Jones can't convince anyone that Smith is guilty. This doesn't mean Jones lacks knowledge, she just lacks the ability to transmit that knowledge to others who are in a worse epistemic position.

The other kind of case we can use involves extreme philosophical positions. In any discussion about the nature of philosophical disputes, it's worth thinking how the dispute applies to disagreements with people holding extreme views. For instance, we might consider the *trivialist* who thinks that every proposition is true, or the *nihilist* who rejects every proposition. It is very hard to argue against those two. Any argument you offer, even an argument to the conclusion that trivialism is false, will be happily accepted by the trivialist, since she obviously accepts the conclusion. But she won't think it is a problem for trivialism. And any argument you offer will be rejected by the trivialist. Indeed, he'll say that by presenting premises, and presumably presenting them as true, you are simply assuming that nihilism is false. And that's to beg the question against nihilism.

So if our principle linking knowledge and convincingness were true, we would get the conclusion that we know nothing at all. That's because we have no way of arguing against the nihilist. Even most academic sceptics won't accept that. (Though some Pyrrhonian sceptics might accept it.) So this isn't a principle the academic sceptic can rely on. And of course it is self-defeating, if we have an argument from some premises to the conclusion we know nothing at all.

So I think these arguments involving philosophical dialectics are basically non-starters. It's sadly true that we can't argue the sceptic out of her stance, like we can't argue certain conspiracy theorists out of their stances. But nothing about our knowledge follows from that.

6.2 Discrimination

Let's consider a different way the sceptic might motivate the claim we don't know we're not in a disaster scenario. Intuitively, we can't discriminate between real and disaster scenarios. That's a general property of illusions; they are indiscriminable from some distinct real situation.

The short version of our response to the sceptic will be that once we understand discrimination correctly, we'll see that it is **asymmetric**. In particular, the first of these claims could be true, while the second is false.

- A normal human cannot discriminate a normal scenario from a disaster scenario.
- A normal human cannot discriminate a disaster scenario from a normal scenario.

These claims are in a slightly stilted English. We don't normally talk about discriminating *x* from *y*. Rather, we talk about discriminating between *x* and *y*. But it turns out to be really important to resist this (admittedly natural) terminology. Discriminating between *x* and *y* sounds just the same as discriminating between *y* and *x*. But it is important to the response I want to suggest on behalf of the anti-sceptic that we make this distinction.

Admittedly, what I'm saying here might sound absurd at first. Consider saying this about a person whose job it is to authenticate paintings.

> On the one hand, he's very good at telling real paintings from fakes; on the other hand, he's not very good at telling fake paintings from reals.

This sounds absurd, almost inconsistent. But it's just this kind of response that I'm going to endorse. Sometimes responding to the careful sceptic forces us to make hard choices!

Here's the more careful statement of the sceptic's argument.

1. For *S* to know that she is not in a sceptical scenario, she must be able to discriminate the actual situation from the sceptical scenario.
2. *S* cannot discriminate the actual scenario from the sceptical scenario.
3. Indiscriminability is symmetric, i.e., for any two situations o_1 and o_2, if *S* cannot discriminate o_1 from o_2, she cannot discriminate o_2 from o_1.

C. *S* does not know she is not in a sceptical scenario.

It turns out to be surprisingly difficult to state just what the notion of discrimination at play here is.

As a first pass, we might say that *S* can discriminate o_1 from o_2 iff she can know that she is in o_2 and not o_1. But if we individuate situations finely this will lead to some odd results. In particular, this will make us say that the agent **can't** make some discriminations she intuitively can make.

Assume that there is some *q* such that I cannot possibly know *q*. Let *q* for instance be the truth about whether there are an odd or even number of grains of lunar soil on the moon. And let o_2 be the actual situation, and o_1 be a situation where I've been in New York for the past 6 weeks. Intuitively I can discriminate o_1 from o_2; if I had been in New York I would know it! But I can't know that I'm in o_2, since o_2 settles the question of whether *q* is true, and I can't settle that. So this is too strong.

As a second pass, we might say that S can discrminate o_1 from o_2 iff there is some p such that she can know p and p is true in o_2 and not o_1. But this turns out to be surprisingly too weak, if properly interpreted.

Note here first that 'can know' here must be interpreted non-factively. One 'can know' p in this sense if one's cognitive capacities allow for coming to know p, even if the capacities can only do this if the facts are a little different. To see why this is important, see what happens if we make 'can know' factive. Consider the case where o_1 is a scenario where I've been in New York for the past 6 weeks, and o_2 a scenario where I've been in Los Angeles for the last 6 weeks. There's no p I can know (in the factive sense) that's true in one but not the other. But I can tell New York and Los Angeles apart, so we want a theory of discrimination that says I can.

We can get this if we allow for a non-factive sense of 'can know'. There's a good sense in which I can know I'm looking at the Statue of Liberty. It's fairly distinctive, and I know what it looks like. That's (one of) the ways in which I can discriminate being in Los Angeles from being in New York. So those situations are discriminable (at least in that direction).

The problem is that once we do this, we end up with too liberal a notion of discriminability. When I'm looking at very tall buildings from ground level, I can't discriminate a 77 story building from a 78 story building. I can't tell the difference by judging heights, and if I tried to count windows or something similar, I'd just lose count.[1] But imagine that my 'margin of error' in detecting building heights is 5 stories, and the building is 72 stories tall. When I say that's my 'margin of error', I mean that when faced with a building x stories tall, I know it is between $x - 5$ and $x + 5$ stories tall. So I know this building is between 67 and 77 stories tall, inclusive of the end points. And that proposition is true in the scenario in which the building is 77 stories tall, but not in the scenario in which it is 78 stories tall. So on this account of discrimination, it turns out that I can discriminate a scenario in which the building in front of me is 77 stories tall from one in which it is 78 stories tall. And that's a mistake; I can't make such fine grained discriminations.

Summing up, the proponent of this account of discrimination faces a dilemma. Either we understand discrimination factively, or not. If we understand it factively, the account denies us discriminatory capacities we intuitively have. If we understand it non-factively, the account says that we have discriminatory capacities we intuitively lack.

We get closer to the truth if we adopt a specifically counterfactual notion of discriminability. Say that S can discriminate o_1 from o_2 iff, were she to be in o_2, she

[1] This is a version of what's called the 'speckled hen problem'. When looking at a speckled hen, we typically don't know how many speckles it has, even though it would look different if it had a different number of speckles.

would know that she's not in o_1. This gets a lot more clean cases right, but it doesn't help the argument above, since it is clearly asymmetric. Let o_1 be a scenario in which S is dead, and o_2 a scenario in which S is alive. Then she can discriminate being alive from being alive, but she can't discriminate being dead from being alive.

But perhaps that's a feature of an oddly quirky example. Let's see what we can do about making the asymmetry claim more plausible here.

To be able to discriminate between two things is a certain ability. I have lousy depth perception, so I can't discriminate between situations that differ solely with respect to the location of an object a few hundred feet away. Other people, with different depth perception, could make this discrimination.

Abilities are not matters of necessity. I could have better depth perception than I do. Maybe I would have better depth perception with surgery, or the right glasses. I certainly could have worse depth perception, if, for example, I lost my sight. But the big lesson is that I could have had different abilities.

This makes for a challenge when we are asking about whether the agent can discriminate between two situations, and the situations are such that she would have different abilities in the two. What should we say when in o_2 she would the ability to tell o_1 and o_2 apart, but in o_1 she would not have this ability? Anything we say here is, I think, bound to sound somewhat unintuitive. But I think it's not terrible to say that she can discriminate o_1 from o_2, but not o_2 from o_1.

This way of thinking about it explains what's so unintuitive about the art evaluator sentence above,

> On the one hand, he's very good at telling real paintings from fakes; on the other hand, he's not very good at telling fake paintings from reals.

Whether the painting in front of the evaluator is real or fake doesn't seem to change the evaluator's abilities. And when abilities don't change between two situations, discrimination is symmetric. But that's not the situation that's relevant to thinking about scepticism.

There's another way we could describe the case though. Perhaps we should say that S can discriminate o_1 from o_2 iff her actual abilities suffice to tell o_1 and o_2 apart. That will make discrimination symmetric. But it will cut off the sceptic at a surprising point. We'll say we simply do have the capacity to discriminate sceptical from real scenarios. We would lack that capacity were sceptical scenarios realised, but actually we have it.

We'll see a similar dialectic play out in the next chapter, where we look at arguments for scepticism from the idea that we would have the same evidence in sceptical situations. The response we'll mostly look at simply denies the claim about sameness of evidence.

Chapter 7

Evidence

7.1 Williamson on Evidence and Knowability

Timothy Williamson (2000) is concerned to respond to sceptical arguments that go via the proposition that agents have the same evidence in the real world as in sceptical scenarios. He doesn't say just which argument he has in mind that takes this as an intermediate step, but it undoubtedly is a step in several sceptical arguments. It is, for instance, pretty essential to an underdetermination argument for scepticism that the agents have the same evidence in the good and bad cases.

Now why might we think that agents have the same evidence in normal and sceptical scenarios? Williamson offers the sceptic one such reason. If we suppose that agents always know what properties their evidence has, and make some other weak assumptions, it is easy to argue that agents have the same evidence in both cases. Or at least, it is easy to argue that the evidence the agents have in the two cases have the same relevant properties, which comes to the same thing for our purposes.

The argument Williamson offers the sceptic is quite simple. Assume that agents always know what properties their evidence has, and assume that agents in normal and sceptical scenarios have evidence with different properties. Since it is part of the definition of a 'normal' scenario that agents' evidence has the properties we associate with successful knowledge, it can be known, even in the sceptical scenario, what properties one's evidence would have if one were in a normal scenario. If we assume also that agents know what properties their own evidence has, then the agent in a sceptical scenario will be able to tell whether they are in a sceptical scenario or a normal scenario. They'll be able to do this by comparing the properties of their own evidence (which they know by hypothesis) and the properties their evidence would have if they were in a normal scenario. But this is absurd; the point of a sceptical scenario is that in it you can't tell you're in a sceptical scenario. (And it would take a brave anti-sceptic to deny the sceptic this step of their argument.) From this absurdity, we can conclude that either:

- Evidence has the same properties in normal and sceptical scenarios; or

- Agents in sceptical scenarios do not know which properties their evidence has.

Williamson wants to argue against arguments for the first disjunct that assume the falsity of the second disjunct.

The kind of argument Williamson offers is continuous with other arguments Williamson makes against the idea that certain conditions are **luminous**. A luminous condition (for a person S), is one that S knows obtains whenever it does. To take one prima facie plausible example, your pain is plausible to you iff whenever you are in pain, you know you are in pain. Williamson argues there are no non-trivial luminous conditions. (Later in the course we'll look at one very interesting example of this; Williamson argues that knowledge isn't luminous. So sometimes you know something but don't, and can't, know that you know it.)

Here's how the argument goes. Assume, for *reductio*, the general princple that agents know exactly which properties their evidence has.[1] And assume that if one knows one evidence has a property in a particular situation, then it also has the property in a very similar situation.[2] That is motivated by general principles about knowledge and indiscriminability, plus a general view that there's not much we can perfectly discriminate. Quite generally, if one knows p, then p must be true in all situations that are, for all one knows, actual. And since one can't tell the actual world apart from a lot of very similar situations, or at least ones that appear similar given your capacities, that means if you know p, then p must obtain throughout those similar situations. That's the key premise behind Williamson's argument.

Now consider a sequence of similar cases, starting with one where one's evidence is clearly F, and ending where it is clearly $\neg F$. (For example, let F be the property that the evidence includes a feeling of being cold, and imagine the cases are ones where the holder of the evidence feels ever so gradually warmer and warmer, step by step.) At the first step, the agent's evidence is F, so by the first assumption she knows it is F, so by the second assumption the next step is F. And this reasoning can repeat all the way until we derive the clearly false conclusion that the last step is F. Williamson concludes that at least at some point in this sequence, the agent does not know that their evidence is F, although it in fact is.

This, he argues, undermines the argument he offered the sceptic. Without a general principle saying that agents know what properties their evidence has, there is no reason to think that the agent in the sceptical scenario knows which properties her evidence has. And without that, the argument that the agent in the sceptical scenario could know that she's in the sceptical scenario doesn't get off the ground. So this particular

[1] When I say we're assuming something for *reductio*, I mean that we're going to prove that the thing assumed is false by showing that it leads to absurd consequences.

[2] Note that we're not assuming one knows it has the property in similar situations. That would lead to a nasty slippery slope.

argument for scepticism, from knowability of evidence to similarity of evidence across normal and sceptical scenarios, and from that to lack of external world knowledge in the real world via an underdetermination argument, fails. And, ironically, it fails because it supposes too much knowledge of one's own situation.

7.2 Perceptual and Phenomenal Models of Evidence

Williamson takes his primary opponent to be the person who holds a 'phenomenal conception of evidence'. This is the view that the evidence a person has supervenes on their phenomenal states, or perhaps just is their phenomenal states. This picture of evidence seems to be behind many sceptical arguments. One key reason that sceptical scenarios are troubling is because they are phenomenologically the same as actuality. And this would be really disturbing if, as most Western philosophers throughout history assumed, we access the world via our phenomenological states. (The 'Western' qualifier is needed here; (Nagel, 2014) notes that this assumption was rejected in some Indian traditions.)

John Hawthorne (2005) notes that there is another opponent Williamson needs to consider, the proponent of the 'perceptual conception of evidence'. This is the view that one's evidence supervenes on one's perceptual states. On this view, a person in the normal world and the person in a traditional sceptical scenario may have different evidence, but two people who see and hear the same things have the same evidence, even if this evidence is misleading for one and a good guide to the world for the other.

The person who has a perceptual conception of evidence won't be worried by evil demon scenarios. We see different things to the victim of an evil demon. We can see tables, chairs and coffee cups, and they can't.[3] But the perceptual conception of evidence won't help respond to the sceptic about the future. Scepticism about the future can be motivated by scenarios described by David Hume (1739/1978), where the future fails to resemble the past. So imagine a world much like ours in the past, but where from now on it is warm and sunny in Ann Arbor every January, but it snows every day in Miami. We have the same perceptions in the actual world as in this (presumably non-actual) world. So on a perceptual theory of evidence, we have the same evidence. So if sameness of evidence implies sameness of knowledge, we know the same things in the actual world as in that world. But that seems implausible; we know it will be colder next January in Ann Arbor than in Miami, and they don't.

Williamson's own view is that one's evidence is what one knows. So we really do have different evidence to the person in a world where Ann Arbor and Miami switch climates. We know that they won't switch climates, so we have some evidence they don't! Let's state the three theories of evidence we've looked at so far; noting that as

[3] At least, it seems they can't; we'll come back in the next chapter to whether they really can.

we go down this list it becomes harder and harder for two people to have the same evidence.

Phenomenological Theory of Evidence One's evidence is determined by one's phenomenological states. Two people have the same evidence iff they have the same phenomenology.

Perceptual Theory of Evidence One's evidence is determined by one's perceptual states. Two people have the same evidence iff they perceive the same things.

Knowledge Theory of Evidence One's evidence is determined by one's knowledge. Two people have the same evidence iff they know the same things.

Here are some pairs of cases that show how the last two theories come apart. The first pair of cases are Hawthorne's, the next two pairs are newish.

World 2A I see a gas gauge that reads "Full". The gauge is accurate, and so I come to know that the gas tank is full.

World 2B I see a gas gauge that reads "Full". The gauge is inaccurate, and so (since knowledge is factive) I do not come to know that the gas tank is full.

World 3A On Friday afternoon, *X* asks *Y* whether the nearby bank branch will be open tomorrow. *Y* knows that it is, so says "Yes", and *X* comes to believe that the branch will be open.

World 3B On Friday afternoon, *X* asks *Y* whether the nearby bank branch will be open tomorrow. *Y* believes that it is, so says "Yes". But the branch just this week decided to cease Saturday trading, so when *X* believes *Y*, she forms a false belief, and hence does not gain knowledge.

World 4A When walking down her building's hallway, through a door *Z* hears her neighbour say "We're moving to California this summer." The neighbour said this because she is in fact moving to California, and was telling a friend this. *Z* comes to believe, and know, that her neighbour is moving to California this summer.

World 4B When walking down her building's hallway, through a door *Z* hears her neighbour say "We're moving to California this summer." The neighbour said this because she is rehearsing for a play, and this is one of the lines. The neighbour is in fact moving to California, but this isn't why she said it. *Z* comes to believe that the neighbour is moving to California, but this does not constitute knowledge.

In each case, my instinctive reaction is that in the 'A' and 'B' worlds have agents with the same evidence. That is something that a perceptual model of evidence would suggest, but the knowledge theory of evidence would reject. If that's right, then we have some reason to believe the perceptual theory of evidence. And if that theory is right, then we have a simple response to the traditional sceptic - we have different evidence to people in sceptical scenarios - but not such a simple response to the sceptic about the future.

Hawthorne gives two arguments for treating the cases as being ones where the evidence is the same. One is a straight appeal to intuition, and to how we would usually talk about the case. Hawthorne's judgment, which I think I share, is that this kind of argument has some force for the 'same-evidence' position – they really are intuitively cases where the evidence is the same – but intuitions around here are weak and not obviously particularly useful evidence.

The other argument is an appeal to the kinds of inductive inferences the agents in the two cases can reasonably make. Here it seems plausible that (a) the agents in each case are warranted in drawing the same inductive conclusions, and (b) this is a sign that they have the same evidence. But to check this, we have to deal with one complication about the gas gauge case, and respond to what Williamson says about Hawthorne's argument.

When Hawthorne says that in **World 2B**, the gas gauge is inaccurate, there are two things this might mean. It might mean that for some reason the gauge just isn't working right now. Perhaps the gauge is generally reliable, but it doesn't work when the humidity is above 80\%, which it is now. Or it might mean that the gauge is broken, and is hopelessly unreliable. This might make a difference for how we consider the warrant the agent gets from looking at the gauge. It isn't obvious, to me at least, that we get the same inductive warrant from reading a defective machine as reading a working machine. We don't want to be infallibilists about how we think about measuring devices, so we should allow that we might be misled by a generally reliable machine. But I don't see why that should mean that we think a broken gauge can be evidence for anything. So let's either take Hawthorne's case to be a case where the gauge is generally reliable, or just focus on the other two cases.

Williamson's response to Hawthorne's case is a little odd, and worth quoting at some length. The response is in two parts, and we'll look at each.

> [S]uppose that for all the observer in World **2A** knows, he is in **World 2B**. Then we should retract the description of **2A** as a world in which the observer knows that the tank is full. Although the gauge is accurate, that fact is not epistemologically available in the way required for seeing that it reads "Full" to yield knowledge that the tank is full. (Williamson, 2005, 478)

This seems like a very strong constraint on what relationship we must have towards measuring devices. It isn't at all obvious that we must know that a device is reliable before we can get knowledge from it. Perhaps it is sufficient that it simply is reliable. Similarly, it is a very common view about testimony that we do not have to antecedently know that a source is reliable before we can get knowledge from their testimony. Having said this, if we interpret World **2B** in the way I suggested we should, then the person in World **2A** does know that they aren't in **World 2B**. That's something they can simply infer from the fact that their gas gauge says that the tank is full, and the tank is full, so the gauge must be working.

> Second, suppose that it is not the case that for all the observer in World **2A** knows, he is in **World 2B**. Thus, in **2A**, the danger of an inaccurate fuel gage is too remote for **2B** to be an epistemic possibility. The observer in **2A** is in better epistemic circumstances than is the observer in **2B**. Then it seems quite unsurprising that the observer in **2B** has better evidence than the observer in **2B**: for example, for the proposition that he will get home that night. Thus there is both asymmetry of evidence and asymmetry of knowledge; again, the equation E=K holds. (Williamson, 2005, 479)

There is a crucial move here which seems clearly mistaken to me. Williamson first states, correctly, that the observer in **2A** has **better** evidence than the observer in **2B**. He then infers that the observer in **2A** has **more** evidence than the observer in **2B**. And that doesn't follow. Or, at least, it doesn't follow if there is a distinction between the **quantity** of evidence one has and the **quality** of evidence one has. I think that the observers in the two cases have exactly the same pieces of evidence, but that evidence is a higher quality of evidence in world 2A.

7.3 Evidence, Metaphysics and Methodology

It is plausible that Descartes, and later Cartesians, believed the following two claims:

1. An agent can know which phenomenal states they are in.
2. What evidence an agent has supervenes on her phenomenal states.

Williamson wants to argue that the Cartesians are further committed to item 1 being a key reason for believing item 2. Or, perhaps more accurately, he presupposes this, for he thinks that an argument against item 1 is a strong argument against item 2.

But maybe that presupposition is false. It seems to me that there are two quite distinct ways to support item 2, even if we concede that Descartes was wrong about item 1.

One way is by an appeal to metaphysics. Towards the end of the *Meditations*, Descartes gives a story for why he thinks even a well-designed machine, like the human body, will make some mistakes.

> I observe, in addition, that the nature of the body is such that whenever any part of it is moved by another part which is some distance away, it can always be moved in the same fashion by any of the parts which lie in between, even if the more distant part does nothing. For example, in a cord ABCD, if one end D is pulled so the other end A moves, the exact same movement could have been brought about if one of the intermediate points B or C had been pulled, and D had not moved at all. In similar fashion, when I feel a pain in my foot, physiology tells me that this happens by means of nerves distributed throughout the foot, and that these nerves are like cords which go from the foot right up to the brain. When the nerves are pulled in the foot, they in turn pull on inner parts of the brain to which they are attached, and produced a certain motion in them; and nature has laid it down that this motion should produce in the mind a sensation of pain, as occurring in the foot. But since these nerves, in passing from the foot to the brain, must pass through the calf, the thigh, the lumbar region, the back and the neck, it can happen that, even if it is not the part in the foot but one of the intermediate parts which is being pulled, the same motion will occur in the brain as occurs when the foot is hurt, and so it will necessarily come about that the mind feels the same sensation of pain. And we must suppose the same thing happens with regard to any other sensation.
>
> My final observation is that any given movement occurring in the part of the brain that immediately affects the mind produces just one corresponding sensation; and hence the best system that could be devised is that it should produce the one sensation which, of all possible sensations, is most especially and most frequently conducive to the preservation of the healthy man. (Descartes, 1641/1996b, 61–2)

Descartes is primarily making a claim about physiology here. But we can see, especially in the note that this is the best possible physiology, he's not just describing humans but evaluating them. The key point is that distant causes are in a way screened off by nearby causes. In a very strong sense, things are just the same for us whether the distant causes are normal or deviant, as long as the nearby causes are the same. This suggests the following principle:

- Any two people such that the nearby causes of their beliefs are the same have the same evidence.

Now if you are a Cartesian, you think the nearby causes of one's beliefs are one's phenomenal states, so this gets you the phenomenal conception of evidence. But that's (probably) a false Cartesian view about how the mind works. So let's set it aside. The principle connecting evidence to nearby causal connections still seems independently plausible, and an independent challenge to Williamson's position on evidence.

Note that if you don't think phenomenal states have a crucial causal role in the formation of belief, then this view won't lead you to a phenomenal conception of evidence. But it might lead you to something very close, such as a perceptual, or physiological, view of evidence. In particular, any position in this vicinity will have the consequence that agents in a normal world, and agents in the kind of scenario the inductive sceptic brings up, will have the same evidence.

Indeed, once we start thinking of evidence as a causal notion, we can weaken this principle even further, and still retain the conclusion that in inductive sceptical scenarios, agents have the same evidence. All that we need for that conclusion is the following principle:

- Any two people such that all the causes of their beliefs are the same have the same evidence.

The upshot of this is that while it is a little tricky to find a principle that entails that agents in the brain in vat scenario have the same evidence as agents in the normal scenario, it is relatively easy to find plausible principles which entail that whether the future goes normally or abnormally makes no difference to your current evidence. And that's enough to motivate the relevant premise in an underdetermination argument for inductive scepticism.

But even that, I think, understates the strength of the sceptic's position here. It is far from obvious that the sceptic owes us a theory from which their views about sameness of evidence in normal and sceptical scenarios can be derived. In general, our judgments that a certain person lacks knowledge is more reliable, and more secure, than our theories about why they lack knowledge. This about the ancient cases Nagel (2014, 58) discusses; we can know that they are not cases of knowledge well before we figure out why they are not.

The exact same thing could be true of the sceptic's premise here. We have, in everyday life, to frequently judge which of two people has better evidence with respect to some question. This is part of how we judge which person is better placed to answer that question, which is often a crucial judgment. As a consequence of having that skill, we often are in a position to judge that two people have exactly the same evidence. And we can deploy that skill in counterfactual settings, or in fictional settings, just as we can in real life. When we deploy that skill in thinking about sceptical scenarios, we judge that the agents in the normal and sceptical scenarios have the same evidence. Perhaps one could undermine that judgment by showing that it is false. Indeed, I suspect reflection on the nature of perception suggests that it is indeed false in certain external world sceptical scenarios. But it isn't a kind of judgment that stands in need of support by philosophical argumentation. So it isn't something that can be undermined by showing that a particular philosophical argument for it does not work.

Chapter 8

The Semantic Response to Scepticism

So far we've assumed that the following argument is sound.

1. In a disaster scenario, most of our beliefs are false.
2. False beliefs do not amount to knowledge.

C. So, in a disaster scenario, we have very little knowledge.

Later in the course, we'll come back to premise two. Today's topic is whether premise one is true. We'll look at an interesting argument that in fact it is not true, that even in disaster scenarios, many of our beliefs are true. Then we'll look at reasons for thinking that this undermines sceptical arguments, followed by some arguments that in fact the sceptic can easily get around the problems here raised.

8.1 Semantic Externalism

Let's get away from scepticism and disasters for a little, and instead talk about water. Actually, we need to talk about both 'water', the word, and water, the substance, i.e.,H_2O. And actually even further, we need to think about what it means to say that water just is H_2O.

The word 'water' has been around for a long time, much longer than human's knowledge of oxygen. Shakespeare, for instance, uses the word over one hundred times, though we had no knowledge of oxygen then. I don't know how much Shakespeare thought about chemistry, but let's assume that he held onto the ancient idea that water was, alongside fire, earth and air, one of the four basic elements. Indeed, let's assume that he onetime, perhaps just in school, uttered the sentence "Water is an element". And assume that he also frequently said that there was water in rivers and oceans, and in cups, and so on.

Did the discovery of oxygen change the meaning of the word 'water'? Presumably not. Consider these two hypotheses about Shakespeare.

- When he said "Water is an element", he was making a true claim about a non-existent substance, not a false claim about the stuff all around us.

- When he said “Water is an element”, he was making a false claim about the stuff all around us.

It seems more plausible that the second is true. That makes sense of the idea that Shakespeare could truly say that there was water in rivers, lakes and oceans, that people drink water, and so on. And it means we don’t have to ‘translate’ Shakespeare’s use of ‘water’ into our language; the word hasn’t changed meanings so it doesn’t need translating.

There is a simple explanation for the second idea. It is that the word ‘water’ gets its meaning not from our theories about water, but from our interactions with it. When we use ‘water’, we intend to pick out some stuff in the world. At this level, referential intentions are usually successful, and in fact we do pick out that very stuff. And that stuff is, whether we know it or not, H_2O. So Shakespeare was able to talk about H_2O, i.e., water, whether he knew it or not.

There is a flip side to this. If water just is H_2O, then a world without H_2O is a world without water. This fact is often dramatised with a thought experiment referred to as Twin Earth. Imagine a world that is, on the surface, much like ours. In particular, there is a clear liquid that fills the rivers, lakes and oceans, that is important for sustaining life and so on. But this liquid is not H_2O; we’ll use XYZ for its chemical composition. The intuition many have is that this isn’t really water; that such a world lacks water altogether. The inhabitants of such a world must mean something different by ‘water’ to what we do; we mean H_2O, and they mean XYZ. This is quite striking if you think not of us and our modern-day counterparts on Twin Earth, who have presumably learned about the chemical difference, but about Shakespeare and his twin counterpart. They use the word ‘water’ in just the same way as each other, but mean different things by it. This puts a bit of a limit on how much the meanings of words are fixed by their use.

Getting back to epistemology, there is an important lesson from the Shakespeare example. It is that we can know rather a lot of facts involving a thing, while being seriously mistaken about its underlying nature. Even if Shakespeare thought that water was an element, he still knew that there was, for example, water in a cup in front of him. It turns out, perhaps surprisingly, you don’t need to know much about water to know where a lot of it is.

Now here’s the radical suggestion that’s relevant to epistemology. Perhaps the victim in a disaster scenario is just like Shakespeare. They are radically mistaken about the fundamental nature of things. But this does not mean that their everyday, quotidian, beliefs about things are wrong, or even that they fail to be knowledge. The brain in a vat who thinks she is in the cold, wearing gloves on her hands, is correct. Or, more precisely, the brain in a vat who thinks to herself the sentence “I am in the cold, wearing gloves on my hands” expresses a true belief, one that amounts to knowledge. Why the quotation marks in that last sentence? Because in an important sense, she speaks a foreign language, and quoting a foreign speaker without quotation marks is misleading.

Let's think through how this could be. When we use 'water', we talk about the stuff that is the causal origin of the use of that word. That's how we get to talk about H_2O, even if we don't know anything about oxygen. By extension, when the brain in a vat uses 'hand', she talks about the things that are the causal origin of her use of that word. As a matter of fact, those are very different things to the things that are the causal origin of our word 'hand'. They are images, or perhaps computational processes, not biological entities. (Well, not what we call 'biological'; the brain in vat can and should call them 'biological'.) But she speaks truly when she says she has hands; she just speaks a different truth to the one we speak when we say we have hands. Her sentence is true just in case she has some of the things that are related to her word 'hand' in just the right way; and she does. That's just like how Shakespeare speaks truly, even knowledgably, when he says "I have a cup of water". That's true just in case he has a cup of the stuff that's related in the right way to his use of 'water'. And, as long as he has H_2O in the cup, he does.

8.2 Externalism and Scepticism

If the reasoning of the previous section is right, then the victim in a disaster scenario does not have a lot of false beliefs. But there are obvious fixes for this. Indeed, Chalmers (2005) mentions two of them. Consider the following two hypotheses, both of them described by Chalmers. (He is talking about the film *The Matrix*, so he uses 'matrix' for a somewhat generic kind of disaster scenario.)

> **New Matrix Hypothesis**: I was recently created, along with all of my memories, and was put in a newly created matrix.
>
> **Recent Matrix Hypothesis**: For most of my life I have not been envatted, but I was recently hooked up to a matrix.

Each of these hypotheses, if true, makes various beliefs of mine false. On the first, my beliefs about the past are mostly false. I think I've lived in Michigan for several years, but in fact I've only lived there for a few minutes. (Strictly speaking, I don't believe that I've lived in *Michigan*, since my beliefs are about something other than the actual Michigan. But for ease of exposition, I'll speak a little loosely here.)

On the second, my belief that I'm currently in Michigan is false. After all, on that hypothesis, my words get hooked up to the world in just the way that we think that they do, and I mean *Michigan* by 'Michigan'. And, unless the vat is still in Michigan itself, I don't live in Michigan any more.

Chalmers notes that neither hypothesis supports a global scepticism. On both of them, any number of my beliefs about the external world are true. On the first, my beliefs about the present state of the external world are true. On the second, my beliefs about the past state of the external world are true. So both of these are only partial sceptical hypotheses.

But this shouldn't really bother the sceptic. There was never any particular rule that said that academic scepticism requires a single global sceptical hypothesis. The very general structure of a sceptical argument is as follows:

1. To know some ordinary proposition *O*, *S* needs to rule out any hypothesis inconsistent with *O*.
2. There are sceptical hypotheses inconsistent with *O*.
3. *S* cannot rule out any sceptical hypothesis.

C. So, *S* does not know *O*.

Assume for now that this argument basically works; in particular that premises 1 and 3 are correct. (Most of what we are discussing in this part of the course are challenges to those premises, but we'll grant them for now.) Now if this argument works, it doesn't show that we know nothing; it shows we don't know one very specific claim *O*. The sceptic obviously wants to generalise that. But let's look at what the generalisation should be. In fact, all the sceptic needs is the first claim here, not the second.

- For any ordinary proposition that an agent purportedly knows, there is a sceptical hypothesis such that the agent can't rule it out, and it is inconsistent with that proposition.
- There is a sceptical hypothesis that the agent can't rule out, and which is inconsistent with all ordinary propositions the agent purportedly knows.

The first claim is *much* weaker than the second. It allows the sceptic to proceed piecemeal, finding distinct sceptical doubts for each claim, rather than finding a single reason for global sceptical doubt. But the upshot is the same.

The logical point here is that the first claim is weaker than the second claim.

- $\forall x(Fx \rightarrow \exists y(Gy \wedge Rxy))$, i.e., for any F, there is a G that is R-related to it.
- $\exists y(Gy \wedge \forall x(Fx \rightarrow Rxy))$, i.e., there is a G that is R-related to all \$F\$s.

Now perhaps semantic externalism implies that the second claim is false; there isn't a global sceptical hypothesis. But it doesn't obviously undermine the first claim, i.e., that any purported knowledge can be subject to sceptical doubts.

8.3 False Beliefs and Inconsistent Hypotheses

There is another oddity of the semantic externalist response to scepticism. Think again about Twin Earth, the world where there is no H_2O, but there is watery stuff. It seems the first claim is false, but the second true.

- There is water on Twin Earth.
- If my counterpart on Twin Earth said "There is water", he would say something true.

Now consider two reasons why a sceptical hypothesis, or a disaster scenario, is relevant to scepticism.

- It is a scenario where what I actually believe is false, and so I need to be able to rule it out in order to have knowledge.
- It is a scenario where my belief forming mechanisms are unreliable, and hence it poses a risk to my claims to knowledge.

The semantic externalist says that we shouldn't be so worried about the second of those risks. Your belief forming mechanisms work fine whether you're a human, a brain in a vat, or whatever. But they agree that your actual beliefs are not made true by the disaster scenario. And that could be relevant to some kinds of sceptical argument.

So the externalist reply is a good one to this kind of argument.

1. To know *O*, I have to be such that I wouldn't falsely believe X were X not true, for any X that is entailed by O.
2. Letting *O* be *I have hands*, and X be *I am not a handless brain in a vat*, I believe X, but if it were not true, I would still believe it.

C. So, I don't know I have hands.

The semantic externalist says, wait a minute, premise 2 is false. If I were a handless brain in a vat, I would not have the false belief that I was not one. Rather, I would have some other, hard to express, true belief. Great, so premise 2 fails.

But it's harder to say what the externalist reply is supposed to be to this argument.

1. To know *O*, I need a reason to rule out any scenario inconsistent with *O*.
2. I have no reason to rule out the scenario where I'm a handless brain in a vat.
3. That I'm a handless brain in a vat is inconsistent with my having hands.

C. So, I don't know I have hands.

Note that semantic externalism does not seem to undermine the crucial premise 3 here. If we use the words 'hands' and 'vat', not the brain in the vat, then it is true that the brain in vat scenario is inconsistent with my having hands.

But maybe the semantic externalist does have something to add here. They give us a somewhat distinctive way of objecting to premise two. Let's think about this with a different example, and then come back to the brain in the vat.

Think again about Shakespeare and the water. He is holding a cup with water, i.e., H_2O in it. Intuitively, he knows it is water. Now think about his twin, who we'll imaginatively call Twin-Shakespeare. *He* is holding a cup with twin-water, i.e., XYZ, in it. It isn't water, so he doesn't know it is. But he knows it is what we call 'twin-water', and he calls 'water'. Moreover, there are any number of other Shakespeare twins, in words where the watery stuff has a yet different structure, who are also good at knowing when they can truly utter the sentence, "This is water".

So what's going on here? The following argument must fail.

1. To know that this is water in front of him, Shakespeare needs a reason to rule out any scenario inconsistent with it being water.
2. Shakespeare has no reason to rule out the scenario where it is XYZ, i.e., twin-water, rather than H_2O, i.e., water, in front of him.
3. That it is twin-water and not water in front of him is inconsistent with it being water in front of him.

C. So, Shakespeare does not know there is water in front of him.

Which premise is false? Perhaps it is premise 1. That would be very bad news for the sceptic, since it undermines a key part of their reasoning. Perhaps it is premise 2. Perhaps Shakespeare's perception of the water is a reason to believe it is water, not twin-water, though he would not, and could not, put it that way. And perhaps it is premise 3. Perhaps in the relevant sense, the twin-water hypothesis is not inconsistent with it being water. It's true that, as a matter of metaphysical necessity, something could not both be H_2O and XYZ. But maybe 'inconsistent' here means something stronger, such as that we can prove by logic that the two hypotheses are not true at once. And to prove nothing is both H_2O and XYZ, you don't just need logic, you need chemistry too.

But let's talk a bit more about this response to premise two. In a way, it is a version of the Williamsonian response of the previous chapter. Williamson said that the sceptic has us start in the wrong place. Where we can start in reasoning is dependent on features of the world that aren't in some sense visible to us. Shakespeare can start with the there is water in front of him, even though in some sense things would look the same if it was twin-water.

The same goes for the ordinary person, like you or me, replying to the sceptic. We do have a reason for thinking we are not brains in vats, we say. Look, here are some hands. But, says the sceptic, they might be mere images. No, we say, then we would still be speaking truthfully when we said they were hands, though we would be speaking a different truth. It's true that we don't know precisely what kind of hands we have, in a deep metaphysical sense. But we know we have hands. And perhaps that's enough of an anti-sceptical response.

Chapter 9

Humean Scepticism

9.1 Methods Argument

Our final sceptical argument is, I think, the most interesting. Here's a quick statement of it.

Methods Argument (Quantified)

1. There is no means by which *S* could know $\neg SH$.

C. *S* does not know $\neg SH$.

We could try to refine that argument in one of two ways. First, we could try to list the ways in which we could know $\neg SH$. So we could get an argument that looks like this.

Methods Argument (List)

1. *S* could not come to know $\neg SH$ by visual perception.
2. *S* could not come to know $\neg SH$ by tactile perception.
3. *S* could not come to know $\neg SH$ by memory.
4. *S* could not come to know $\neg SH$ by testimony
5. *S* could not come to know $\neg SH$ by reasoning.
6. Any knowledge is knowledge we acquired by visual perception, or tactile perception, or memory, or testimony or reasoning.

C. *S* does not know $\neg SH$.

Now there's obviously a bit of work to do to defend each of the first 5 premises. There's something odd about thinking you could come to know $\neg SH$ by testimony. After all, you don't know that there are even other testifiers before you know $\neg SH$. But perhaps we could overcome those challenges.

The biggest problem is that premise 6 is implausible. There's plenty of things we know by things other than those methods. Right now, I know there are birds outside my window because I can hear them. But there's nothing on the list for auditory perception. We could add that, but it still wouldn't be clear that we'd exhausted the list.

And it isn't clear what to say about things that things we know by many methods. I know my neighbor is home because I can *see* a late model blue Ford in the driveway, and I *remember* that my neighbor drives that car. Is my knowledge due to visual perception or to memory? The right answer feels to be that it's a bit of both. Perhaps I also know $\neg SH$ by some combination of methods, not any one method. So even if the sceptic could convince us that no one method could let us know $\neg SH$, there remains the possibility that we know it by a combination of methods.

9.2 A Priori and A Posteriori

There is a way to fix this argument, but it requires a detour through some philosophical terminology. Many philosophers, over many centuries, have thought that there is an important distinction between knowledge that relies, in whole or in part, on our sensory contact with the world, and knowledge which does not.

For instance, I know that it isn't raining right now in my corner of Ann Arbor. This knowledge relies on my sensations about the world. I couldn't just sit back and reason to the conclusion that it isn't raining; I have to look. Knowledge that requires looking at the world, we'll call *a posteriori* knowledge.

On the other hand, I know that there's no largest prime number without this knowledge depending on my sensations of the world. At one stage, I knew this by what looks like an a posteriori method. I looked at some ink marks on some pages, decoded them using my (a posteriori) knowledge of how ink marks and intended messages are correlated in the English speaking world, and inferred that smart people thought that there is no largest prime, so there is no largest prime. But now I know the proof that there is no largest prime. (It isn't a hard proof. Assume *n* is the largest prime. Given that assumption, you can prove both that $n!+1$ is prime, and that it is not prime, contradicting the assumption. So there must not be a largest prime.) So by pure reasoning, I can come to know that there is no largest prime.

When you know something due to pure reasoning abilities, we call this *a priori* knowledge. It is standard at this point to make an important clarification to the notion of the *a priori* that traces back to Kant. Another piece of *a priori* knowledge I have is that all tigers are tigers. That isn't a deep piece of knowledge, but it does seem to be true, and not something I need to observe the world to know. But note that if I had never had any sensory contact with the world, it isn't clear that I could even have thought about tigers, so I couldn't believe that all tigers are tigers, so I couldn't know that all tigers are tigers. Kant's idea, which I think is right, is that we should distinguish the class propositions that we need sensory contact with the world to be able to *think*, from the class of propositions where our *justification* for believing them comes from sensory contact with the world. It is the latter class that cannot be known *a priori*.

So that's our key distinction, between the *a posteriori* knowledge, knowledge that is justified in some part by our empirical, sensory evidence about the world, and the *a priori* knowledge, knowledge that is independent of this empirical evidence. Not all philosophers accept that this distinction is useful, or even coherent. Some philosophers reject the idea that there is any knowledge that is completely independent, even in its justification, from our contact with the world. And some philosophers say that the relevant notion of indepenence cannot be cashed out in a coherent way. But to keep the discussion manageable, I'll assume that we do have a useful distinction here.

Given that distinction, we can restate the methods argument in a tighter way.

Methods Argument (Types)

1. *S* could not know $\neg SH$ a priori.
2. *S* could not know $\neg SH$ a posteriori.
3. All knowledge is a priori or a posteriori.

C. *S* could not know $\neg SH$.

If we define a posteriori knowledge as simply knowledge that is not a priori, the third premise looks fairly secure. And the argument is valid. So the issue is how to defend the first two premises. We'll look at that after first looking at an important historical predecessor to this argument.

9.3 Hume on Induction

David Hume (1739/1978) presented an argument for inductive scepticism. There is some scholarly dispute about just how he intended that argument to be taken, but whatever Hume's intent, it is an interesting argument. Here's a somewhat simplified version of Hume's reasoning.[1]

1. We cannot know that the future will resemble the past by pure reasoning, since the only things we can know by reasoning are those whose falsity implies a contradiction, and there is no contradiction in the future failing to resemble the past.
2. We cannot know that the future will resemble the past by observation, since knowledge of the future by observation presupposes that the future will resemble the past, so to use observation to get this knowledge would be hopelessly circular.

[1] One important simplification is that I've changed Hume's talk about reason into talk about knowledge. This isn't the only simplification, but I think the argument I'm presenting here is representative of a common reading of Hume, and is important for at least that reason.

3. All knowledge is either by pure reasoning or by observation.

C. We cannot know the future will resemble the past.

This is scepticism about the future, since all knowledge of the future requires that there be some resemblence between the future and the past.

The biggest challenge, I think, to this Humean argument is to the second premise. As we saw in the discussion of the easy knowledge argument, it is hard to formulate a plausible general principle against 'circular' belief forming methods. Rather than rehash that, I'll discuss a more contemporary form of the argument, one that has come to prominence due to important work by James Pryor (2000) and Roger White (2006).

9.4 What is the Sceptical Hypothesis?

To make **Methods Argument (Types)** really stick, we need to clarify what it is that we mean by *SH*. We'll focus on the evidence that *S* actually has.

Let's say *E* is the sum total of all of *S*'s empirical evidence about the world. It is the conjunction of all of the things that form the basis for her a posteriori knowledge. If she knows anything a posteriori, it is justified by something in *E*. Let *O* be some ordinary proposition that intutively she knows, and is not in her evidence. (If you think, as we discussed in the previous chapter, that all knowledge is evidence, you'll think is is an impossible stipulation. That's one way to get out of the sceptical argument, but since we've already discussed it, we'll set it aside.) Now the sceptical hypothesis is simply that *E* obtains but *O* does not. In symbols.

- $SH = E \wedge \neg O$

So $\neg SH$ is simply the opposite of that, i.e., that it isn't true both that *E* is true and *O* false. Equivalently, either *E* fails to obtain, or *O* does obtain. Again in symbols:

- $\neg SH = \neg(E \wedge \neg O) = \neg E \vee O$

9.5 Against A Priori Knowledge

The argument that this can't be known a priori is rather simple, and very close to Hume's earlier argument. The things we can know a priori are things that, in an intuitive sense, couldn't have been false. If something could be false, then to know whether or not it is false, we have to go and look. But that means our knowledge that it is true rests in part on our looking, i.e., it is a posteriori.

This line of thought can be backed up by thinking through our paradigms of a priori knowledge. Recall that these were,

- There is no largest prime.

- All tigers are tigers.

These are not just a priori knowable, they could not be false. And for the reasons just given, this doesn't seem to be an accident. Since the sceptical hypothesis could be true, that's part of it's appeal, this seems to suffice to show that it can't be known to be false a priori.

There is one interesting class of exceptions to this general rule, but it isn't obviously relevant to scepticism. There are some names in natural language that function as 'descriptive names'. They both pick out a particular individual, but pick them out by means of a description. Perhaps the most famous such name in English is "Jack the Ripper". This is meant to be a name for the person, whoever it was, who committed certain murders in late 19th century London. Arguably, both of the following claims are true.

1. We can know a priori that if anyone committed those murders, it was Jack the Ripper.
2. There is a metaphysical possibility in which Jack the Ripper, whoever he is, dies in infancy, and so does not commit those murders.

When there are descriptive names around, it seems that we can know a priori things that could have been false. But it isn't obvious how relevant this is to the sceptical argument. After all, $\neg E \vee O$ doesn't have any descriptive names in it.

9.6 Against A Posteriori Knowledge

This side of the argument is trickier, but we'll start with a very intuitive motivation for the sceptic's view. Here's what the sceptic is going to want to argue: It is impossible to know a disjunction, like $\neg E \vee O$ is true, solely on the basis of knowing that one of the disjuncts is *false*. If you learn that one of the disjuncts is false, and then you know the disjunction, you must have known the disjunction all along. The alternative is that we could have situations like this.

> A: Who will get the best grade on the logic exam?
> B: No idea, it could be anyone.
> A: I think it will be Billy or Suzy.
> B: You could be right, but I think it might also be someone else.
> C: Hey, I just heard, Billy didn't get the best grade on the logic exam.
> B: Hey, A, you're right, either Billy or Suzy will get the best grade on the exam!

Ruling out Billy seems like it should make A's hypothesis *less* likely. (We could carefully show this in a probabilistic framework, and indeed Roger White (2006) makes that probabilistic argument central to his presentation. But I won't go over those details here.) In general, evidence *against* one of the disjuncts looks like a very poor candidate to be evidence *for* the whole disjunction.

And that's what the sceptic is trying to argue here. All they have to show, at this stage of the argument, is that if *S* didn't know $\neg E \vee O$ a priori, then getting the extra evidence *E* can't suffice for knowing that. And it can't suffice because it points in entirely the wrong direction; it makes the disjunction less plausible. But if *E* is all of *S*'s empirical evidence, and it doesn't help justify belief in $\neg E \vee O$, then that belief isn't a posteriori justified. So it can't be a piece of a posteriori knowledge.

Here's another way to make the point vivid. Assume that getting evidence *E* is a good way to justify belief $\neg E \vee O$. Presumably getting evidence $\neg E$ is an even better way to justify belief in $\neg E \vee O$, since $\neg E$ guarantees the truth of $\neg E \vee O$. So whether the evidence is *E* or $\neg E$, belief in $\neg E \vee O$ will be justified. So there's no need to wait to see what comes in; the belief is justified either way.

Actually, that's a bit quick. It's meant to be an argument by cases, but the two cases don't exhaust the possibilities. One case is that *S*'s evidence is summed up by *E*. The alternative is that *S*'s evidence is *not* summed up by *E*. That doesn't mean that *S*'s evidence is summed up by, or even includes $\neg E$. It might be that *S* gets no evidence whatsoever.

9.7 Summing Up

So what, if anything, is wrong with **Methods Argument (Types)**? My own view is that both of the first two premises are mistaken, but this is hardly an intuitive view. I think we should just accept, in the face of sceptical pressures, that we know but not on the basis of evidence that various disaster scenarios do not arise. And I think that when one of the disjuncts is as multifaceted and complicated as $\neg E$, it may well be that learning that the disjunction is false (i.e., that *E* is true) suffices for learning the disjunction.

But these are, to put it mildly, somewhat tentatively held views. What I really think is that **Methods Argument (Types)** is the best way to put the sceptical argument. That's where the sceptic's real challenge comes from, just as Hume said it did. Say we accept that we can know that we're not in a disaster scenario. The real challenge is in saying how we know that. And whatever answer is correct there will be both unintuitive, and deeply revealing about the nature of knowledge. This is a topic on which there is a lot of ongoing philosophical work - see for instance Ralph Wedgwood (2013)'s recent paper for one up-to-date take, and it isn't clear at all what the final verdict will be.

Part II

Analysis of Knowledge

Chapter 10

Introduction, and Challenges to Necessity

In this part we'll be looking at attempts to **analyse** knowledge. We'll often be following closely the Stanford Encyclopedia of Philosophy article on "The Analysis of Knowledge" by Ichikawa and Steup (2013), and it would be a good idea to read that article alongside these notes. First, a brief reminder of what it would be to analyse knowledge.

As we noted in the introduction, an analysis of a philosophically interesting concept would satisfy three conditions.

1. It would provide **necessary** conditions for the application of the concept.
2. It would provide **sufficient** conditions for the application of the concept.
3. It would in some way illuminate the concept.

Here is one example of an analysis that does all three of these tasks.

> *x* is *y*'s **sister** if and only if:
>
> - *x* is female; and
> - *x* and *y* have the same parents.

Both clauses are necessary. If *x* is not female, then *x* is not *y*'s sister. And if they don't have the same parents, then *x* is not *y*'s sister. And between those two clauses we get sufficient conditions for sisterhood. And the account is illuminating, indeed you could use it to explain what a sister is to someone who doesn't know it. Illuminating isn't the same as deciding; there are vague cases of being female, and vague cases of having the same parent. But that's all good, since there are vague cases of being a sister.

The orthodox view in contemporary philosophy is that analyses that are successful in this sense are very very rare. In contemporary philosophy, scepticism about analysis was given a huge boost by Wittgenstein (1953). He showed how hard it would be to give a successful analysis, in something like the above sense, of 'game'. And for good measure, he argued that the concept of a game is very important to a lot of philosophical projects, especially projects to do with communication. He was right about both of these points, and we should be hesitant about the prospects for success of other attempts at analysis. But it turns out that even if the project of analysis rarely succeeds, it can be worth doing. Plato's early Socratic dialogues are full of failed attempts at analysis, indeed attempts that are shown to be failures by the end of the dialogue. But we learn a lot from seeing why they fail. The same is true, I think, of knowledge.

10.1 The JTB Analysis

As Ichikawa and Steup (2013) note, there is a particular analysis of knowledge that is often called the 'traditional' analysis. Nagel (2014, Ch. 4) calls it the 'classical' analysis. I'm a little sceptical of those designations. The main citations that are usually given for the so-called traditional, or classical, account are to A. J. Ayer (1956) and Roderick Chisholm (1957). The main citation given for a refutation of it is Gettier (1963). It seems the reign of the traditional account was only 7 years!

What's true is that the account these philosophers are describing is simple and attractive, and it feels like the truth must be somewhere in its vicinity. Although very few, if any, philosophers endorse it these days, many philosophers think that the right account of knowledge is nearby to this account.

The account in question says that knowledge has three components. It says that the following account is a good analysis of knowledge.

S knows that *p* iff

1. *S* believes that *p*;
2. *S*'s belief that *p* is justified; and
3. *p* is true.

That is, *S* knows that *p* if and only if *S* has a **J**ustified **T**rue **B**elief that *p*. This is often called the JTB account of knowledge, and that's the designation that we'll frequently use.

There are three natural ways to challenge an analysis. We'll spend most of our time on the second of these challenges.

First, one could challenge the **necessity** of the conditions. That is, one could argue that the key term holds without the allegedly necessary conditions obtaining. In this case, this would require arguing that there are cases of knowledge either without justification, or without truth, or without belief. The bulk of this chapter will be on these challenges.

Second, one could challenge the **sufficiency** of the conditions. That is, one could argue that the key term does not hold although all the conditions are met. The last fifty years of work on analyses of knowledge has consisted, in large part, of people proposing analyses of knowledge, and other people showing that those conditions can obtain without *S* knowing that *p*. We'll spend the rest of this part of the course on reactions to the discovery that the JTB conditions are insufficient for knowledge.

Third, one could argue that the conditions are not suitably illuminating. In these notes we aren't going to spend much time on that challenge, though it is a very interesting idea. A number of philosophers, especially those working in the tradition established by Timothy Williamson (2000), think that the best way to understand justification is in terms of knowledge. If that's right, then the JTB account might be extensionally correct (it really might provide necessary and sufficient conditions) without being a good analysis. In particular, this would be the case if it turned out that *S* is justified in believing that *p* if and only if *S* knows that *p*, and moreover, this is something like a good analysis of justification. As I said, we're going to set that option aside here, though I want you to remember that it is an option as we progress through other alternatives.

10.2 Knowledge without Truth

The JTB account requires that knowledge is **factive**. That is, it requires that *p* is true for *S knows that p* to be true.

The orthodox view in semantics is that *S knows that p* doesn't just require *p*'s truth for its truth, it **presupposes** that *p*. The idea, somewhat roughly, is that using the clause *S knows that p* when *p* is not true is as mistaken as using the phrase *Smith's sister* when Smith does not have a sister. For some evidence for this, consider what happens when a knowledge ascription (a claim of the form *S knows that p*) is in the antecedent (the conditional clause) of a conditional. It seems to me that someone asserting (1) is committed to both the claim that Kanga has a child, and that it has been kidnapped.

> (1) If Kanga knows that her child has been kidnapped, she will be angry.

So the factivity claim looks, if anything, to understate how important *p*'s truth is to *S knows that p*. But this has been challenged. Allan Hazlett (2010) has argued that there are cases of true knowledge ascriptions of falsehoods. Consider these cases.

> (2) Many years ago, everyone knew the earth was flat.
> (3) She knew she could trust him, but as happened so many times before,
> she was destined to be let down.

In both cases, it seems we have an appropriate knowledge ascription of a falsehood. It's natural to think that the appropriateness here requires truth, so we have a true knowledge ascription of a falsehood, contrary to what the JTB analysis requires.

The standard response to these cases, which I agree with, is that they involve what Richard Holton (1997) called **protagonist projection**. Very often, in telling a story, we can say things that aren't really true, but which seem true from the perspective of a protagonist of the story. For instance, it is fine to say (4) in a story.

(4) He gave her a diamond ring, but it turned out to be a fake.

This is perfectly fine in a story, though read literally it looks like a contradiction. If he gave her a fake, he really didn't give her a diamond ring.

This kind of construction works best in contexts that are clearly marked as storytelling contexts. I've tried to indicate this here by using as cliched tropes as possible, which I think it easier to 'get' the intended reading. But it's even easier to indicate this within an actual story. Holton's paper includes a lot of cases from novels, and this nice example from Billy Bragg's 1983 song *A New England*.

I saw two shooting stars last night,
I wished on them, but they were only satellites.

Of course, satellites are really not shooting stars. But the story isn't blatantly contradictory; we know exactly what happened. The narrator wished on things he thought were shooting stars, but they were in fact satellites.

So this hypothesis I think neatly explains the cases like (2) and (3), while retaining the intuitive idea that knowledge is factive. If you need yet more evidence that something funny is happening with (2) and (3), note that they require a slightly abnormal pronunciation of *knew*. (At least in my idiolect, it's a slightly stressed, considerably elongated pronunciation.) A requirement of non-standard pronunciation is usually good evidence that something odd is happening, and what's happening here is protagonist projection.

10.3 Knowledge without Belief

Colin Radford (1966) suggested that the B condition of JTB was unnecessary. He argued that there could be cases of knowledge that *p* where the agent did not even believe that *p*. The most intuitive kind of case here, I think, involves quiz settings.

A student, call her Brown, is asked some questions about U.S. history. (In Radford's original example, it was British history.) For example, she is asked when Michigan became a state, and she insists that she has no idea. She is asked to guess, and she says 1837, the correct answer. And this is repeated for several questions. Each time she insists that she has no idea, but on being prodded to give an answer, gives the correct answer.

There is some intuitive support for the idea that the student (a) does not believe that, for example, Michigan became a state in 1837, but (b) does know that Michigan became a state in 1837. If that's right, then there can be knowledge without belief.

In support of (a), we can note that the agent would not bet very much on the proposition that Michigan became a state in 1837. Given a choice between $9 for sure, and a bet that pays $10 if Michigan became a state in 1837, and nothing otherwise, she would take the $9 for sure. Yet someone who believed that Michigan became a state in 1837 would believe that this was a choice between $9 and $10, and would take the $10. So she doesn't act like someone who believes that Michigan became a state in 1837 acts.

In support of (b), we can note that we would readily attribute her knowledge. We could easily imagine saying things like "Brown knew an incredible amount about U.S. history, though she wasn't aware she knew it." Or she might say "I had no idea how much I knew about U.S. history." Moreover, she clearly *possesses a lot of information* about U.S. history; else she would not be able to answer correctly.

But I think there are things to be set against that verdict. Consider again the bet we discussed two paragraphs ago. If Brown knows that Michigan became a state in 1837, she knows that taking the bet will return more money than taking the $9. And someone who knows that taking a bet will lead to more money than an alternative is justified in taking such a bet. But it seems it would be very strange, in Brown's position, to take this bet. Indeed, it would seem somewhat fortuitous that she came out ahead by taking the bet.

So I think we should conclude these are not really cases of knowledge, even though we may describe them that way at times. What then, should we say about the fact that we make these descriptions. There are a couple of options here.

One option is to adopt some kind of error theory, saying that these statements about Brown aren't literally true, but they are convenient falsehoods. (Why call this an error theory? Well, usually when we use convenient falsehoods, it is evidence we are doing this. If we don't realise that describing Brown as knowing things involves falsehoods, then we are making an error.)

Another option is to say that the English verb *know* is slightly ambiguous. Most of the time it picks out a state of mind that requires belief. But in special cases we can use it to mean something mere like information possession. This latter special sense might explain why it is a little more natural to attribute knowledge than belief to non-agents (like computers, cars, etc.) If by *knows* we sometimes mean *possesses the information*, then it is literally true that cars and computers know things. After all, they do possess a lot of information. But this isn't the sense of knowledge that most philosophers, and certainly most epistemologists have been most interested in. That sense, I think, requires belief. (If you're interested in more information on this option, chapter 7 of Nagel (2014) has a long and excellent discussion of philosophers who think the verb *know* is systematically extremely flexible, in much the way gradable adjectives like *tall* are flexible. We won't be discussing such moves in this course though.)

10.4 Justification

The last clause to consider is the J clause, that beliefs must be **justified** to constitute knowledge. The notion of justification here is itself a highly contested one. We could spend a semester (or more) just on it. Instead, I'll briefly note the two dominant themes in theories of justification.

One theme links justification closely to **evidence**. To have a justified belief is, on this picture, to have this belief on the basis of good evidence. The papers collected in Conee and Feldman (2004) provide a good guide to this approach to justification.

Another theme links justification closely to **reliability**. To have a justified belief is, on this picture, is to have formed the belief by a reliable process. Reliabilist theories of justification have been prominently defended in recent years by Alvin Goldman (1976).

The best way to see the difference between these two views is to look at the cases they treat differently. A victim of an evil demon has (in some sense) good evidence, but unreliable belief formation methods. A reliable clairvoyant has reliable belief formation methods, but (arguably) poor evidence. Let's look at those cases in more detail.

Consider the kind of person Descartes (1641/1996b) fears he is towards the end of the First Meditation. He has all sorts of (apparently) sensory images. And he has, all his life, come to believe things on the basis of these images. But this is a horribly unreliable method. When he has the image of hands in front of him, there is never a hand in front of him. Yet one might have the intuition that Descartes was never doing anything wrong. If we were the victim of a demon, then we would be blameless in acting as if our senses were reliable, although they are not. If that blamelessness implies justification, it follows that the demon victim has justified beliefs without reliable beliefs. And the natural explanation for that is that the victim has good evidence for her beliefs. (The argument in this paragraph is a version of the 'New Evil Demon problem', as formulated by Stewart Cohen (1984). There is a fascinating variant of that problem at the end of (Pryor, 2001).)

Alternatively, consider an example due to Laurence BonJour (1985), of Norman the clairvoyant. One morning, Norman wakes up with a reliable clairvoyant ability. He has no reason to believe he has such an ability, and indeed every reason to believe he lacks it. After all, he's human, and humans typically have no such ability. But as a matter of fact, Norman's ability is real and reliable. On the basis of this ability, Norman forms the belief that the President is in New York. (Imagine he had a vision as of the President in New York.) Is this belief justified? Bonjour argued that this was a problem case for reliabilist theories of justification, because intuitively the belief is not justified, yet it was formed by a reliable means. Whether that's true or not, what's interesting for us is the possibility of such a case. If the agent with the reliable ability does not know that the ability is reliable, and indeed has evidence against its reliability, a belief formed reliably may nevertheless be poorly supported by evidence.

Perhaps unsurprisingly, there have been attempts to capture what is attractive both in the evidentialist strand, and in the reliabilist strand, of theories of justification. One picture is that there isn't a disagreement between two theories here, but just two different concepts of good-making features of belief Alston (1985). Another picture, I think more attractive, is that both evidence and reliability are needed for justification. If it is a requirement on a source being a genuinely evidential source that it is a reliable guide to the truth, then the idea that beliefs must be based on evidence will imply that they must be reliably connected to the truth (in some way). One nice development of this idea is Timothy Williamson's contention that something is part of an agent's evidence if and only if she knows it, plus (as we'll see in the discussion of safety constraints on knowledge) a requirement that knowledge be reliable.

To avoid some of these complications, we'll focus in what follows on beliefs that are both formed by a reliable means, and based on excellent evidence. By doing this, we can make some progress on talking about the connection between justification and knowledge without having to take a stand on exactly what justification is. This is hardly a new move; most of the literature on the possibility of justification without knowledge involves examples where the beliefs in question are both reliably formed and based on good evidence. And that seems to me the sensible way to approach the matter.

But should we think that justification is necessary for knowledge? Here it seems the relevant argument is similar to the argument of the previous section. Knowledge can justify action. In particular, if an agent knows that *p*, the agent can use *p* as a reason in deciding what to do (Hawthorne and Stanley, 2008). And, provided that *p* is relevant to her decision, she can thereby come to a justified decision about what to do. But an unjustified belief can't justify action. This should be a fixed point, whatever theory of justification you have. Indeed, when trying to think through whether the demon victim, or Norman the clairvoyant, are justified in their beliefs, it is natural to think about whether they would be justified in acting on their beliefs. From this it follows that knowledge requires justification.

As Ichikawa and Steup (2013) note, there are cases where it seems intuitive to describe agents as having knowledge, but lacking justification. But again the comparison to the Radford cases is instructive. What's going on in these cases is that it is natural to describe people who possess the information that *p* as knowing that *p*. And that can be true even if they have very poor justification for the belief that *p*. As we noted above, this might be a misuse of the verb *knows*, or it might be an alternate meaning of that verb. But it can't be the philosophically interesting usage. That usage is closely tied to rational action, and unjustified beliefs are not closely tied to rational action.

Chapter 11

Counterexamples to Sufficiency

In the next two chapters we'll look through five counterexamples to the JTB analysis of knowledge. The first of these, due to Bertrand Russell (1948, 154) was not clearly directed against the JTB account, and indeed the famous JTB accounts due to Ayer (1956) and Chisholm (1957) postdate it. The second and third are from Dharmottara, as cited in Nagel (2014, 57). They date from around 770, but weren't properly considered by Anglophone philosophers until recently. And we'll look at the two cases due to Edmund Gettier (1963) are clearly directed against that analysis, and constituted one of the most successful attempts in the history of philosophy to convince the field that a prominent theory was mistaken.

11.1 Russell's Stopped Clock

A woman, call her Alice, looks at the clock on the tower she sees every day to check the time. The clock shows 4 o'clock, so Alice forms the belief that it is 4 o'clock. And it is indeed 4 o'clock. But the clock has stopped; it has been showing 4 o'clock for the last twelve hours. Intuitively, Alice has a well justified true belief that it is 4 o'clock. But she does not know this; you cannot come to know what time it is by looking at a stopped clock.

At first pass, this looks like a strong counterexample to JTB. But let's see if we can put any pressure on that first impression. Is Alice really justified? Well, Alice clearly has excellent evidence that it is 4 o'clock. After all, she has checked the time on the prominent clock tower. I said she had seen the clock many times before, and presumably if it were unreliable, she would have found out about this by now. The clock isn't permanently marred; it just hasn't been fixed since it stopped twelve hours ago. So if we understand justification in terms of evidence, then Alice's belief is justified.

Matters are a little trickier if we think of justification in terms of reliability. On the one hand, the clock is generally reliable. Indeed, I said in the previous paragraph that it wasn't unreliable, and in context that sounded like the right thing to say. On the other hand, it's a stopped clock! It's just about the paradigm of an unreliable measuring instrument. So there's a very good sense in which it isn't reliable.

We've run into a version of what's known as the **generality problem** for reliabilist theories of justification. Reliability is, in simple cases, a matter of having a high ratio of successes to failures. But that assumes that we can identify which cases are the relevant successes and failures. Imagine that we have a drug designed to cure disease X. It successfully cures X in 90% of women, but only 10% of men. And it cures X in 90% of old people, but only 10% of young people. If a 20 year old woman has disease X, and we give her the drug, are we giving her a reliable cure? Perhaps yes - she's a woman and the drug cures X in 90% of women. Perhaps no - she's young, and the drug cures X in only 10% of young people. The technical term here is that we need to determine the relevant **reference class** before we can say how effective the drug is for people like this patient. If the reference class is all women, the drug is very effective, very reliable. If the reference class is all young people, the drug is almost useless, highly unreliable. In a case like this we would hopefully have more data involving other young women, though in practice solving variants of this problem can pose some complications.

To bring this back to our case, imagine that the clock had not been stopped for 12 hours - it just stopped right as Alice looked at it. Is it a reliable clock? If the reference class includes all and only times up to when Alice looked, it is very reliable. If it includes all and only times starting when Alice looks, it is pretty unreliable. If there's no fact of the matter about what's the right reference class, there arguably is no fact of the matter about whether the clock is reliable.

In Russell's original case, where the clock has been stopped for 12 hours, there is both a reference class relative to which the clock is unreliable, the times around when Alice looks, and a reference class relative to which the clock is very reliable, namely the lifetime of the clock. It feels like the smaller reference class is the relevant one, but it is rather tricky to try to argue for this feeling, or to come up with a broader theory from which it falls out as a consequence. (The best such arguments, I think, involve considerations about **safety**, which we'll return to in a few chapters.)

11.2 Dharmottara's Examples

Here is how Nagel presents the examples.

> A fire has just been lit to roast some meat. The fire hasn't started sending up any smoke, but the smell of the meat has attracted a cloud of insects. From a distance, an observer sees the dark swarm above the horizon and mistakes it for smoke. "There's a fire burning at that spot," the distant observer says. Does the observer know that there is a fire burning in the distance?
>
> A desert traveller is searching for water. He sees, in the valley ahead, a shimmering blue expanse. Unfortunately, it's a mirage. But fortunately, when he reaches the spot where there appeared to be water, there actually

> is water, hidden under a rock. Did the traveller know, as he stood on the hilltop hallucinating, that there was water ahead? (Nagel, 2014, 57)

Now it is a little anachronistic to describe these as counterexamples to the JTB theory of knowledge, since the Indian and Tibetan philosophers who discussed them didn't have much use for the concept of justification as we've been using it. More precisely, they didn't think there was much epistemological interest in non-factive epistemic or doxastic concepts. And both the ideas of justification we started with - support by the evidence, and generation by a reliable process - are non-factive, in the sense that they are consistent with falsity. But set aside the issue of how to best make sense of these examples within Indian and Tibetan philosophy, and just look at them from our viewpoint.[1]

For it does look like in both cases the agent has a belief that is true, and well based on the evidence, and formed by a reliable process. We could quibble about just what exactly the evidence is in these cases. Does the traveller have the same evidence as the person who actually sees smoke from a fire? (I gather this point is one that Indian philosophers took to be central.) And, as with the clock, we could ask whether there is some sense in which the traveller is using an unreliable process. But it does seem pretty plausible here that the believers are responding sensibly to their evidence, and are using generally reliable methods, and yet they don't get knowledge.

If these aren't cases of knowledge, then why aren't they? At least five answers seem initially plausible.

1. The believers use a false assumption in getting to their conclusion. The observer thinks, "That's smoke, where there's smoke there's fire, so there's fire." The first step is wrong.
2. Although beliefs are true, they aren't true in the way the believers thought. The desert traveller thinks there is a large amount of water there, not a small amount under a rock.
3. The beliefs aren't caused by the facts they are about, or at least they aren't caused in the right way.
4. There are very similar situations to the actual world in which the beliefs are false. There easily might have been a swarm of insects over a dead animal, or a mirage without a rock hiding water.

[1] Jonathan Stoltz (2007) notes that the focus on factive states and concepts is very foreign to the Western tradition throughout the second millenium, but things might be changing in more recent times. He also suggests, intriguingly, that some of the gap between the Indo-Tibetan traditions and Western traditions comes from differences in thinking about whether dispositional or occurent mental states are the primary object of evaluation. These notes are very much written from the perspective that dispositional states are what we evaluate in the first instance; seeing what hapens if we drop that presupposition is a very interesting project, but not one we'll attempt to tackle.

5. The beliefs are insensitive; in the most realistic situations in which they are false, the beliefs are still held.

Just looking at any one example, it isn't easy to say which of these is the right explanation of any one case. We may need more examples to test various explanatory theories. And we shouldn't be certain before doing an investigation that there is one unified explanation of all the cases like the three we've seen so far. So let's look at some more cases. We'll spend the rest of this paper on the famous, though somewhat convoluted, counterexamples to JTB proposed by Edmund Gettier (1963).

11.3 Gettier's Target

Gettier starts his paper by describing analyses of knowledge due to A. J. Ayer (1956) and Roderick Chisholm (1957). Neither of these are explicitly of the JTB form. They are both tripartite analyses, and in both cases truth and belief are two of the parts. But the third parts are a little more precise. In particular, they are

- *S* has adequate evidence for *p*. (Chisholm)
- *S* has the right to be sure that *p*. (Ayer)

The idea that these could be grouped into a broader category of JTB theories is Gettier's. And while Gettier is correct in this, it is worth noting that both these accounts feel much more like evidence-based accounts than reliability-based accounts. I'm not quite sure how to understand Ayer's notion of a right to be sure; it's hard to make sense of the political notion of a right carrying over to epistemology. (For example, political rights usually imply restrictions on what others can do, but there aren't any restrictions on others that follow from my having great evidence that *p*.) But it seems that Alice does have a right to be sure that it is 4 o'clock, even if she is blamelessly unaware that the clock she is using has stopped. But let's not worry about Ayer and Chisholm too much; Gettier's target is really much broader than that.

Indeed, after giving his examples, Gettier goes on to note that the problems will arise for any theory that analyses knowledge as truth plus belief plus X, where X does not entail truth, and is closed under logical entailment. This is an important point, but one that was not sufficiently appreciated in the literature until it was expanded upon by Linda Zagzebski (1994). So the targets here are really a large group.

11.4 Gettier's First Case: Ten Coins

Smith and Jones are both up for a promotion. Smith, somehow, gets excellent evidence for these two propositions:

- Jones will get the promotion.

- Jones has ten coins in his pocket.

And Smith believes both of these. From these two beliefs, Smith infers that the man (the characters in these stories are all male) who will get the job has ten coins in his pocket. Call that proposition *p*.

Now it turns out that *p* is true. But it isn't true for the reason that Smith thinks it is true. In fact, Smith himself will get the job. And, unbeknownst to anyone, Smith has ten coins in his pocket. So Smith's belief that *p* is true, and presumably well justified – he has excellent evidence for it. But it doesn't seem that it is something Smith knows. Just like with Dharomattara's examples, the intuition that Smith doesn't know *p* seems stronger than any explanation we might offer for that intuition. But it isn't like we have no idea what could explain it; the five options I suggested above all look like they carry over.

11.5 Case Two: Brown in Barcelona

This time, Smith has some different evidence about Jones. He has evidence that Jones owns a Ford. This isn't actually true, but Smith has excellent evidence for it. Jones has always owned a Ford in the past, and has just offered Smith a ride while driving a Ford. It turns out that Smith just sold his old Ford, and doesn't own any car right now; he is renting the one that he is driving. But Smith couldn't possibly have known this.

Now Smith has some quirky interest in propositional logic. He finds it interesting that once you have *p*, you can infer $p \vee q$ for any *q* whatsoever. And he's been wondering where in the world Brown is, not having any evidence at all about where Brown is. So he forms all of the following beliefs.

- Either Jones owns a Ford, or Brown is in Boston.
- Either Jones owns a Ford, or Brown is in Barcelona.
- Either Jones owns a Ford, or Brown is in Brest-Litovsk.

Now as it turns out, Brown is actually in Barcelona. So the middle one is true. And all of them are justified. But Smith, says Gettier, doesn't know any of these three. Again, we have a justified true belief without knowledge.

There is something a little surprising about this example given a reliabilist conception of justification. How, exactly, should we describe the exercise Smith is going through? Here are two competing descriptions.

- Drawing logical entailments from a justified belief.
- Forming beliefs of the form *Jones owns a Ford, or Brown is in X* for arbitrary cities *X*.

The first method is very reliable, at least if we understand justification in terms of reliability. The second is extremely unreliable; it works in just one case and fails in thousands. Again, we have an instance of the generality problem. Smith can be described in two ways, and one description picks out a very reliable method, while the other picks out a very unreliable method. It seems that the first description, drawing logical entailments from a justified belief, is the most epistemologically significant one. In general, a person who doesn't do anything epistemologically wrong doesn't start going wrong by drawing further logical entailments. But part of why we want to say that Smith doesn't have knowledge is that the second description is also relevant. We'll come back to this point when we discuss resolutions of the puzzle cases in terms of safety in later chapters.

11.6 Knowledge from Falsehood

Gettier hints at a particular argument for the claim that Smith lacks knowledge in these two cases. It is that ascribing knowledge in either case would be a violation of the following principle.

- If *q* is false, and *S*'s basis for believing that *p* is that she deduced it from *q*, then *S* does not know that *p*.

In short, the principle says that you can't get knowledge from falsehood. It's certainly true that in the bulk of cases, a belief derived from a falsehood is not going to be a piece of knowledge. But whether this is a universal claim is a little trickier. Recent papers by Ted Warfield (2005) and Federico Luzzi (2010) have argued that it is not. Here is an example of Warfield's that makes the point.

Dora is watching the news, and she sees a report from a reporter on the Presidential press corp. The reporter starts by saying "I'm here in Wyoming with the President". Dora forms the belief that the President is in Wyoming, and infers from that that the President is not in Washington DC. Now as a matter of fact, the President is not in fact in Wyoming. They had been in Wyoming, but the President, and the travelling media, had just crossed into Utah before they went on air. So Dora's belief that the President in Wyoming clearly isn't knowledge; it isn't true. But her inferred belief that the President is not in Washington DC is true, and feels like a piece of knowledge.

Now it is true in this case, unlike in the Gettier case, that there are plenty of other things Dora does know, from which she could have inferred that the President is not in Washington DC. She knows that the President is near Wyoming. She knows that he is out west somewhere. She knows, from the backdrop to the reporter's story, that he's in a sparsely populated area. But we can assume that that's not the actual route she takes to her belief. She actually uses the false premise that he's in Wyoming.

In any case, the mere availability of alternative, truthful evidence doesn't seem to be enough to rescue a claim to knowledge. Imagine that Smith really had evidence that Brown was in Barcelona. Brown had told him that was where he was going a week ago. But Smith had simply forgotten it, and was aimlessly coming up with city names starting with 'B' when he made his momentous inference. The availability of a route through true beliefs to his conclusion doesn't seem to make that conclusion a piece of knowledge.

But this example does put paid to a suggestion that we complicate the JTB account by adding a fourth clause, namely that *S* did not infer *p* from a false lemma (Clark, 1963). That isn't necessary for knowledge. And it often isn't sufficient either. If Smith does not believe that Jones owns a Ford, but merely bases his belief that Jones owns a Ford or Brown is in Barcelona on evidence of Jones's Ford-owning habits, it still isn't knowledge. But he need not have actually formed a false belief along the way.

11.7 True in the Wrong Way

There's something else funny about the two cases that we might hope to turn into a way to rescue the JTB analysis. In both cases, Smith believes something true, but the thing he believes isn't true for the reason that he thinks it is true. The idea behind that last clause is (somewhat deliberately) inchoate, but maybe via recent work on truthmakers (Rodriguez-Pereyra, 2006) we can turn it into a workable theory.

But in general it isn't true that in order to know for *S* to know that *p*, it must be that *p* is true in the way *S* thinks it is. Albert is looking for a place to put down his coffee cup. He sees a nearby table, and thinks it will support his coffee cup, so he puts it there. It's a perfectly normal table, and a perfectly normal coffee cup, so this is not surprising. Indeed, we might say that Albert knows the table will support his coffee cup. But here's the thing about Albert. He wasn't paying very close attention in high school physics or chemistry. So he thinks that a solid object is solid all the way through. He thinks that's why tables support cups - there is no space inside a table-surface, so the cup doesn't fall. If he were to be told the real story about how solid objects are constituted at the atomic and molecular level, he'd be stunned. But he gets through life fine without knowing this, and his coffee cups of course never fall through the 'space' inside table-surfaces.

Albert's belief that the table will support his cup is true. But it isn't true in anything like the way Albert thinks it is. Does that mean he doesn't know the table will support his cup? I doubt it. After all, views like Albert's are very common in the broader community, and I'm sure were even more common before the development of modern atomic theory. We don't want to say all these people simply do not know their cups will be stable when placed on tables. That would be conceding too much to the sceptic.

So while it's true that Smith might be very surprised to find out how his belief is true, that alone can't be what prevents him from having knowledge. There are lots of mundane cases where someone's true belief does amount to knowledge, though they would be very surprised to find out exactly how the world has cooperated in making their belief true.

11.8 Summing Up

Gettier offers us two cases where an agent has a justified false belief, and then infers something else from that belief which is true, but which the agent has no more reason to believe than the falsehood with which they started. These seem to be cases of justified true belief without knowledge.

But the simplest ways of explaining what goes on in these cases do not work. It's not always true that having a false lemma in one's inferences defeats knowledge. And it's not true that knowledge requires the belief being true in something like the way one thinks it is true.

Next, we'll look at further potential counterexamples to the JTB theory, and then look at alternatives to JTB, and see how they measure up against these counterexamples.

Chapter 12

More Counterexamples to JTB

12.1 Zagzebski's Generalisation

Assume that we try to fix the JTB analysis by somehow strengthening the J condition. Call this new condition J+. One of the distinctive things about the J condition is that it is not **factive**. That is, it is possible to satisfy the condition without p being true. This is, pretty clearly, central to the original Gettier examples. Linda Zagzebski (1994) argues that if J+ is similarly not factive, then the problem will recur.

Here's the basic idea. Assume a philosopher says that knowledge requires J+,T, and B, where J+ is our new non-factive condition, and T is truth and B is belief as before. Call this the J+TB theory. Now consider a case where q is a false belief of the subject, but one which is J+. By hypothesis, such cases are possible. Let r be some random truth that the subject does not believe, and which they have no reason to believe. Now if J+ is closed under entailment, and q is J+, then $q \lor r$ will also be J+. And assuming that the agent considers $q \lor r$, they will presumably form the belief that it is true. After all, it follows trivially from something that they do believe. So they will have a true, J+ belief that $q \lor r$.

But in any such case, it seems very implausible that the agent would have knowledge. One doesn't get knowledge by adding a random disjunct to a belief that does not amount to knowledge, and being lucky that the added disjunct turns out to be true. So the J+TB theory will be subject to counterexample in just the way the JTB theory was.

There is one loophole in this argument, but it is hard to see how a good analysis could exploit it. It could be that J+ is not closed under logical entailment. If that were the case, then q could be J+, while $q \lor r$ was not J+, and hence $q \lor r$ would not be a J+TB without being knowledge. But could that be plausible?

It's certainly conceptually possible. Here's one very boring way to get such a J+. Say that S's belief that p is J+ if either S knows that p, or p is false. It's pretty easy to show that there are no counterexamples to *this* version of the J+TB theory of knowledge. So how does it escape Zagzebski's argument? Well, since q is false, it satisfies the second clause of J+. (Remember, a belief is J+ if it is known or false.) But $q \vee r$ doesn't satisfy either clause, so it can't be a counterexample.

Still, this is a perfectly useless analysis of belief. It certainly isn't illuminating. We can't come to understand anything about knowledge by thinking about the property of being known or false. And there's no independent philosophical interest in the property of being known or false.

Perhaps a better analysis is the 'no false lemmas' analysis. Say a belief is J+ if it is justified and not derived from any false lemmas. That is, for a belief to be J+, the thinker must not have gone via any false steps along the way. We can already see how this is not closed under entailment. In Zagzebski's own example, q is potentially J+, but $q \vee r$ is definitely not J+. After all, the thinker derived $q \vee r$ from q, which is false. So we can't use Zagzebski's argument to show that this analysis won't work.

But this analysis looks implausible on independent grounds. Someone who had evidence for q, and uses that evidence to derive $q \vee r$ without going through q, doesn't thereby come to know q. And as we'll see below, there are all sorts of problem cases for JTB that don't involve false steps.

A very common, and I think correct, reaction to Zagzebski's argument has been to infer that the condition that must be added to belief to get knowledge is **factive**. That is, there can't be an analysis of knowledge that has T as an independent condition. Rather, there must be some philosophically interesting constraint on belief, and that constraint itself must entail truth, if we are going to get a theory of knowledge. We'll see a few such accounts over the next few chapters, some of them directly motivated by (and in one case produced by) Zagzebski.

12.2 Double Luck

One point that is central to Zagzebski's positive thory, as well as the method she uses to generate counterexamples to more theories, is that the prominent puzzle cases for JTB have a *double luck* structure. Think back to the five cases we discussed in the previous chapter. Each of them starts with something unlucky happening to the protagonist. A clock turns out to be stopped; a black swarm is insects, not smoke; an unlikely candidate will get a promotion, etc. And then each of them continues with something that is, from the point of view of getting true beliefs, incredibly lucky. The clock is stopped at *just the right time*, for example.

These kinds of cases, where one bit of bad luck is 'balanced out' by another bit of good luck, we'll call double luck cases. If JTB only fails in double luck cases, that would give us a pretty big hint as to how to fix it. And double luck cases are, by far, the easiest way to generate counterexamples to a JTB-like theory. But perhaps they aren't the only way.

12.3 Fake Barns

Alvin Goldman (1976), working on a suggestion by Carl Ginet, introduced 'fake barn' cases to the epistemological literature. These quickly becamse the most common kind of problem case for JTB that didn't involve a 'double luck' structure.

A man, call him Barney, is driving through a strange and unfamiliar part of the countryside. He looks out the window and sees a barn. He thinks to himself, *That's a barn*. Call this proposition *p*. As it turns out, *p* is true. So far, nothing odd has happened. Barney has seen a barn, and come to believe that it is a barn on the basis of his visual perception. And there's nothing particularly defective about Barney's vision.

What is strange is the land that Barney is in. It is the land which has become known as Fake Barn Country. For some reason, the inhabitants of Fake Barn Country have engaged in an elaborate game of building barn façades that are not genuine barns. If a driver of usual visual acuity, like Barney, saw one of these façades, he would believe that it was a barn. It is a surprising coincidence that Barney has seen the one real barn in all of Fake Barn Country.

So, does Barney know that *p* is true?

On the one hand, Barney's belief that *p* is produced in a very straightforward way. He sees a barn and, being a normal thinker blessed with reasonably good vision, believes that it is a barn on the basis of what he sees. Unlike in the Gettier cases, nothing goes wrong on the path between the world and Barney's belief.

On the other hand, there is a good sense in which Barney's belief is unreliable. Had Barney deployed his barn-recognition capacities anywhere else in Fake Barn Country, he would have been led into error. It seems like sheer good luck that Barney gets it right in this case. But this kind of good luck seems incompatible with knowledge. As we have noted a few times already, it seems key to our idea of knowledge that lucky guesses do not amount to knowledge.

My own view is that we should lean towards the view that Barney does know that *p*, while being extremely sceptical of our ability to decide on these cases in advance of having a well worked out theory of knowledge. In both cases, I'm moved largely by considerations due to Tamar Gendler and John Hawthorne.

In a paper that's considerably more amusing than the average philosophy paper, Gendler and Hawthorne (2005) suggest that intuitions about these 'fake barn' cases are extremely sensitive to the details of the case. And what intuitions we have about the cases are not obviously coherent, or the kind of intuitions we'd like to endorse after careful reflection.

Here's a simple variant of the original fake barn case that makes both points. Say that Barney doesn't see the *only* real barn in Fake Barn Country. There are actually half a dozen real barns, all congregated in the area right around the barn Barney saw.

Driving at some speed, as Barney is, it will only be a minute or two until he's back among the fakes.

But Barney isn't the only person looking at this barn. A hiker, call her Walker, is also looking at this barn. She has a very similar visual impression to Barney. And she also believes that *p* is true; she thinks it is a barn. And like Barney, she has no idea about all the fake barn façades in the surrounding area. But unlike Barney, she is in no immediate danger of being fooled. She is walking slowly to a campsite a mile and a half away, where she will make camp for a couple of days. It will be a long time until she sees any fakes again.

Arguably, Walker's belief that *p* is not true due to any luck, and is not in any way unreliable. Any beliefs she forms about barns in the next couple of days will be true. It's true that in some larger area around her, her barn-detection facilities are unreliable. But in an even larger area around that, Walker's barn-detection facilities go back to being highly reliable. It isn't clear, to put it mildly, why the existence of a mid-sized range of unreliability should override the fact that in the immediate environment, and in the world as a whole, Walker is reliable. But if Walker thereby knows that *p*, it is more than strange to think that the difference between a knower like Walker and an ignoramus like Barney is due simply to their mode of transport.

Even if Barney does not know that *p*, it isn't clear that this is a counterexample to JTB. It does seem like a counterexample to JTB plus the idea that justification is primarily a matter of having good evidence. If one thinks that justification is a matter of reliability, presumably whether Barney knows that *p*, and whether his belief that *p* is justified stand and fall together. After all, the only reason for saying that Barney doesn't know is that he is, in some sense, unreliable. So that's also a reason, on some views about justification, for saying he isn't justified.

But there's still some interest in the case of Barney. The case poses a particular problem for what we'll call virtue-theoretic accounts of knowledge, to be introduced in several chapters' time. Or, at least, it does pose a problem if the existence of the fakes blocks Barney's claim to knowledge. So it's an interesting test case to keep around.

12.4 Lotteries

The slogan for the New York State lottery is "You never know". That is, if you buy a ticket in the lottery, you never know if you'll win. The lottery makers are obviously talking their book here, but it's very tempting to agree with them. That is, it's tempting to agree that we can't know that tickets in a (fair) lottery will lose. A lot of philosophers report the raw intuition that we can't know these tickets will lose. But a few others report the intuition that we can know this. The case strikes me as too hard to just so quickly. But there are a couple of considerations in favor of the view that you can't know of any given ticket that it will lose, in advance of the drawing.

Some ticket, or at least some combination of numbers, will win. And each combination is as likely as any other. So the winning combination is just as likely as yours. So the only grounds you have for ruling out your ticket winning are equally good grounds for ruling out the actually winning ticket winning. And obviously you can't rule that out; it's actually true. So you can't know that your ticket loses.

Alternatively, if you could know that your ticket will lose, you'd have reason to do things that in fact are irrational. Assume that the prize in the lottery is $10,000,000, and there are a million tickets. So the expected value in dollars of any given ticket, its actuarial value, is $10. That doesn't mean that a ticket is worth $10 - the vagaries of marginal utility theory suggest it will be worth a bit less - but it is likely of some substantial value. Now consider what you do once you read that your ticket has lost, i.e., what you do with tickets that you know have lost. You throw them out. And this follows from a more general principle. If you know something is worthless, it is fine to throw it out. So if you know your ticket will lose in advance of the drawing, it is fine to throw it out. But this is crazy; you shouldn't throw out a ticket with an actuarial value of $10. So the hypothesis that you can know it will lose must be false.

On the other hand, there are some considerations in favor of the view that we can know that lottery tickets will lose. We can certainly be extremely confident that they will lose. The chance that any ticket will win is extremely low. If we form the belief that we'll lose the lottery, we'll almost certainly be right. So a belief that you'll lose the lottery is supported by very good evidence; evidence that make it almost certainly correct.

Assume that isn't enough for knowledge. Then it seems that we will know very little. Jonathan Vogel (1990) and John Hawthorne (2004) have argued that if we don't know that we'll lose lotteries, we stand to lose a lot of everyday knowledge. Consider a driver who parks her car on the 4th floor of her regular parking lot one morning. That afternoon, she is trying to remember where her car is, recalls parking on the 4th floor of that lot, and infers that's where the car is. She's right, but car thefts do happen, and she had a one in ten million chance of having her car stolen during the day. Does that chance mean she didn't really know where her car was when she left the office? If we want to say she knew where her car was, what's the difference between that case and the person who believes, on probabilistic grounds, that they'll lose the lottery?

If we take the orthodox line that we can't know we'll lose lotteries, then it seems there's a new kind of counterexample to the JTB analysis. Consider again the person who believes, truly, that she'll lose, and who bases this belief on the ground that there are so many tickets in the lottery that there's no realistic chance of winning. It seems to me her belief is both true and based on very good grounds. So it satisfies the JTB conditions. But orthodoxy says that it is not knowledge. So we have a new counterexample to JTB.

Notably, this is not a counterexample that has anything like a double luck structure. As we'll see in later weeks, some explanations of why we don't have knowledge in Gettier cases easily explain why we don't have knowledge in these cases, but other explanations do not generalise so easily.

12.5 Margins

Our final kind of problem case is from recent work by Timothy Williamson (2013). Again, it helps to illustrate the case with an example. Esther's job is to estimate crowd sizes at major events. One day she is at an event at Michigan Stadium, and has to estimate how many people are there. It isn't quite full, but there aren't many empty seats, and after calculating how many people are likely to not be visible from where she is, she estimates there are 94,500 people there.

Now Esther is a professional estimator, and not stupid. She knows that her estimates are not exact. In fact, she knows that it's hard to do better than get within 2,000 people when estimating a crowd of this size. But she also knows that she is very good at her job, and usually (not always, but usually) does get within 2,000. So she forms the belief that there are more than 92,000 people in the stadium. Call this proposition *p*.

Now as it turns out, *p* is true. There are 93,000 people there. But here's the strange thing. Imagine that Esther had done a better job with her initial estimation, and her estimate had been correct: she had estimated the crowd size at 93,000, not 94,500. Then she would have thought, well since I can't be confident in estimations being more accurate than plus or minus 2,000, the strongest claim I should believe is that the crowd size is between 91,000 and 95,000. That is, had she done a better job at estimating, she would not have believed *p*. She wouldn't have believed *p* to be false, but she would have thought that it was an open possibility that *p* is not true.

So in a good sense, Esther only got to the true belief in *p* because she was less than perfect in her initial estimation. Had she done better, she would not have believed *p*. In that sense, her (true) belief that *p* seems to be due in some sense to good luck. And this is the kind of luck that we usually take to be incompatible with knowledge.

On the other hand, Esther's belief that *p* is, by hypothesis, true, and well justified. She is good at her job, she knows that she is almost always within 2,000 of the correct number, and so she has excellent evidence that *p* is true. It seems like this is a justified true belief. If so, it looks like another kind of counterexample to the JTB theory. And, like the case of Barney, it is a counterexample that doesn't rely on any funny business going on between Esther's evidence and her belief.

So I think this is an interesting, new counterexample to JTB. But to close with an argument for the other side, consider a small variant on the case. Esther's friend, Fred, is also interested in how many people are at the stadium. He asks Esther what her best guess of the crowd size is. She says, 94,500. Fred knows that Esther is good at this, and usually is right to within 2,000. So Fred concludes that there are more than 92,000 people there. That is, Fred concludes that *p*. Do we also want to say that Fred lacks knowledge in this case? I suspect we do; if Fred uses Esther to measure the world, then what's a lucky break for Esther is also a lucky break for Fred. But I can easily imagine someone arguing that this is heading too far down the path to scepticism, and that we should be able to rely on our measuring devices being roughly accurate, at least in the cases where they are roughly accurate.

12.6 Moving Forward

Starting next chapter, we'll look at attempts to 'solve' the problems with the JTB account. We've now got six kinds of tests to use on theories of knowledge, though as noted several times above, a few of them are less than clear tests. The tests are:

1. Russell's stopped clock.
2. Dharmottara's observer who believes there is a fire.
3. Dharmottara's traveller who thinks there is water in fact there is a mirage.
4. Gettier's two examples.
5. Zagzebski's generalisation of Gettier's examples.
6. Barney in Fake Barn Country.
7. Lottery ticket holders who believe they will lose.
8. Williamson's example of good but imperfect estimators.

We'll see whether we can come up with a theory that handles all six cases.

Chapter 13

Sensitivity

In this chapter and the next, we will focus on an idea usually associated with Robert Nozick (1981) – a belief only amounts to knowledge if it is **sensitive**. *S*'s belief that *p* is sensitive just in case it satisfies this condition:

- If *p* had not been true, then *S* would not have believed that *p*.

We'll start with saying a bit about conditionals like this one, which are called **counterfactual** conditionals. Then we'll look at some advantages of adding the sensitivity condition to the analysis of knowledge. Finally, we'll look at how the sensitivity theory provides a novel response to a certain kind of scepticism.

13.1 Counterfactual Conditionals

A counterfactual conditional is a conditional of the following form: *If p had not been the case, then it would have been that q*. We use counterfactuals a lot in everyday life in assigning responsibility, and in planning.

If you want to know whether a person's actions are responsible for a bad outcome, it matters a lot whether the outcome would not have occurred but for their actions. This isn't the only thing that matters. Perhaps if two people do something wrong, and either one's bad actions would have led to the bad outcome, then it can be true that each is responsible even though the outcome would have happened even if they had not performed the bad action. But in general, whether we treat them as responsible, and whether the outcome would have happened if they had not performed the action, go together.

And we use counterfactuals in general in learning from past events. Consider a general case where we've performed some action, and a positive event happens. We want the positive event to happen again. So should we perform the same action? That depends in part on whether, in the earlier case, the positive event would have happened whether or not we had performed the action. We've all heard stories about people who are suspicious in the sense that they keep doing things that correlated with good results in the past. Think of the cargo cults in the Pacific islands after WWII, or the sports fan who insists on sitting on the same spot on the couch that he sat on during a team's big win. What strikes us as crazy about these practices is, in part, that we don't have evidence for a counterfactual connection between the activity and the desired end. We don't know, for example, that the conditional *If I hadn't sat on that spot on the couch, my team wouldn't have won the Superbowl* is true. In fact, we have good reason to think it is false.

So counterfactuals are important for responsibility, and important for planning. They have been used extensively in philosophy as well, including in important contributions to debates on free will (Ayer, 1954), causation (Lewis, 1973a) and mental content (Fodor, 1990). The purpose of this chapter and the next is to look at their relevance to epistemology.

Over the last few decades an extensive literature on the logic of counterfactual conditionals has developed. We're not going to look into that in any detail, though interested readers can see some of the available options outlined in work by David Lewis (1973b) and by Dorothy Edgington (1995). But we will look at one influential idea, due to Lewis, and one complication to it that is relevant to epistemology.

Lewis suggested the following very intuitive way to understand counterfactuals. If you want to know what would have been the case had *p* not been true, look to the most similar possible world to our own where *p* is not true, and see what's true there. That gives the right result in a lot of cases, though as Lewis (1979) later conceded, in many cases you need to clarify the notion of 'similarity' a lot to get the right results.

But there is one kind of case that goes wrong. Imagine that Smith punches Jones, and now Jones is in pain. Intuitively, Smith is responsible for Jones's pain. And, intuitively, that's because the counterfactual *If Smith hadn't punched Jones, Jones would not be in pain* is true. But maybe Smith was so mad at Jones that the most similar world to ours where Smith doesn't *punch* Jones is one where Smith *kicks* Jones. This fact doesn't seem relevant to whether the counterfactual is true, or to whether Smith is responsible for the pain. When thinking about responsibility, we don't consider counterfactuals where Smith harms Jones in some other way.

The key point here is that counterfactuals are context-sensitive. In some contexts, we are happy to accept conditional 1, below. In other contexts, we are happy to accept 2. In no context do we accept both, so in no context do we accept conditional 3.

1. If Smith had not punched Jones, Jones would not be in pain.
2. If Smith had not punched Jones, Smith would have kicked Jones instead.
3. If Smith had not punched Jones, Smith would have kicked Jones instead, and Jones would not be in pain.

Now as ordinary English speakers, we can usually work out what kind of variation from reality is relevant to assessing a particular counterfactual. But this kind of disambiguation isn't always easy, especially when counterfactuals are being quantified over, or used schematically. We'll have to be careful of this point in many of our uses of counterfactuals below.

13.2 Sensitivity

Let's recall again the definition of a sensitive belief.

- *S*'s belief that *p* is **sensitive** just in case this conditional is true: If *p* were not true, *S* would not have believed that *p*.

Using this, we can formulate the Simple Sensitivity Analysis (SSA) of knowledge.

- *S* knows that *p* if and only if *S* has a sensitive belief that *p*.

A sensitive belief is a belief, so the SSA entails the B part of the JTB analysis. It also implies the T part. Assume *p* is false. And assume that the belief is sensitive, so that if *p* were false, *S* would not believe it. Well, those two assumptions imply (by the rule of modus ponens) that *S* does not believe that *p*. But *S* does believe that *p*. So our first assumption, that *p* is false, must itself be false. And that seems to imply that *p* is true.

That's a good feature of the analysis. What we looked at Zagzebski's generalisation of the Gettier case, we saw that it is impossible to understand knowledge as true belief + X, where X is not a feature that implies truth. The SSA avoids that feature. It puts a new requirement on belief, and that requirement both rules out lucky guesses and rules out false beliefs. It's essential to a good theory of knowledge that you rule both of these out in one step; otherwise the original problems will recur.

The SSA is not Nozick's own theory. His theory had one complication that we'll discuss in the next section, and one extra clause. The extra clause was confusingly stated in the form of a 'counterfactual' with a true antecedent. Nozick's presentation of it was like this:

- If *p* were true, *S* would believe it.

It isn't clear how to make sense of that, since *p* really is true. What Nozick meant is that it isn't an accident that *S* belives that *p*; in all the realistic scenarios where *p* is true, *S* still believes it. It isn't clear why this is a good thing to add to knowledge. Imagine that in the course of randomly looking up some other facts, I see that Queen Elizabeth was born in 1926. That seems like something I could easily know, even though I could easily have had no beliefs about it. So this extra condition in Nozick's account seems like a mistake. We'll focus on the simpler, and more plausible, SSA, even though it isn't Nozick's own account.

13.3 Methods

Nozick worries about the following case. Young Tom visits his grandmother. He is healthy, and his grandmother can see he is healthy. Moreover, his grandmother is rather good at detecting Tom's health. From his appearance, she can tell whether he is ill. But Tom has a plan. If he were ill, he wouldn't visit, but send a trusted friend to say he was well. And the friend would be able to convince his grandmother.

Nozick says that despite this, Tom's grandmother can still know he is healthy. After all, her belief that he is healthy is based on a well-functioning, and highly discriminatory, visual capacity. So he complicates the SSA in this way.

First, say that *S*'s belief that *p* obtained via method M is method-sensitive just in case were *p* false, *S* would not have believed that *p* via method M. Second, say that a belief amounts to knowledge just in case it is method-sensitive.

That makes Tom's grandmother's belief into knowledge. Although her belief is not sensitive, it is method-sensitive. There is no realistic way she could have come to a false belief about Tom's health by using her visual capacities. She could have come to a false belief about Tom's health using some other method, e.g., trusting testimony. But that doesn't defeat knowledge; only the fallibility of the method she uses defeats it.

Having noted this complication, we will mostly proceed to ignore it. We'll mainly look at cases where the method is reasonably fixed, so we can just investigate whether the SSA gets the case right, and the verdict should carry over to the more complicated theory.

13.4 Advantages of the Sensitivity Account

Sensitity accounts of knowledge, either the SSA or the more complicated account in terms of method-sensitivity, get a lot our troubling cases right.

It gets the **stopped clock** example right. Someone who forms a true belief on the basis of a stopped clock will form a false belief if the facts are different. That's the paradigm of an insensitive belief. So the SSA rules, rightly, that their belief is not knowledge.

It gets Dharomattara's **desert traveller** case right. Even if there was no water under the rock, the traveller would have still seen the mirage, and believed there was water. So his belief was insensitive. It is a little trickier to say whether it gets the fire case right; let's set that aside for now.

It gets Gettier's **coins example** right. If it were not true that the man who will get the job has ten coins in his pocket, Smith would still believe that. That is, even if Smith had nine or eleven coins in his pocket, he would still believe both that Jones had ten coins in his pocket, and that Jones would get the job, so the man who will get the job has ten coins in his pocket. So his belief is insensitive.

And it gets Gettier's **Barcelona example** right. If the disjunction *Either Jones owns a Ford, or Brown is in Barcelona* were false, then Jones would still not own a Ford, and Brown would be elsewhere. So that doesn't require changing any part of the world concerning Jones, but does require moving Brown somewhere else. Still, Smith's belief was grounded in any fact at all about Brown, so he would still believe *Either Jones owns a Ford, or Brown is in Barcelona*. So this belief is still insensitive, and is not knowledge.

In each of these three cases, the SSA gets the right verdict. But more importantly than that, there is something that feels right about how it gets this verdict. It doesn't seem crazy to think that our judgment that these are not cases of knowledge is related to the fact that these beliefs are insensitive. The insensitivity comes from the fact that the beliefs are somehow disconnected from the world, and that's why they aren't knowledge. We're going to spend some time looking at cases like these, but I don't want us to just count up how many cases the theory gets right or wrong. We want to look at whether we get a better explanation of the cases from the theories, and plausibly the SSA does give us a good explanation of what's gone wrong in these cases. (Spoiler alert: What we see in the next chapter will make us reconsider this. What I'm stressing here is that these are plausible explanations of what's going wrong; we aren't simply curve-fitting.)

We already discussed how the sensitivity account avoids Zagzebski's generalisation of the Gettier case. We're going to spend a lot of time on the fake barns case in the next chapter. And the margins case has a lot of variants which are best discussed when we later get to virtue-based approaches to knoweldge. So let's just look at one more case, the lottery.

Assume a holder of a lottery ticket, in a fair lottery, thinks that they will lose. This belief could be very probably true. If the lottery is big enough, it could be much more probable than most things we believe. But it won't be sensitive. Were the lottery to turn out differently, the person would still believe they will lose. Indeed, even if the lottery has happened, but the person has not been informed of its result, their belief is insensitive. So it isn't knowledge.

As I said in the previous chapter, I'm not sure what the right answer is here. But there's a lot to like about the way the SSA handles the case. Consider these two cases.

1. Alice has a ticket in a 100 ticket lottery. She reads in the newspaper (which occasionally makes mistakes) that she has lost.
2. Bruce has a ticket in a quadrillion ticket lottery. The winner has been drawn, but Bruce has not seen any reports. Still, based on the long odds, Bruce believes that he lost. And he did in fact lose.

Many people report feeling differently about Alice and Bruce's beliefs. It is more comfortable to say that Alice knows she lost than that Bruce knows he lost. This can't be explained probabilistically. Given the frequency of misprints and typos in newspapers, the probability that Alice won given the newspaper report could easily be higher than the probability that Bruce won. But Alice's report is grounded in the result of the lottery in a way that Bruce's isn't. If Alice had won, she would (probably) have believed something different. If Bruce had won, he would have the same beliefs he actually has. This way of explaining the felt difference between the cases feels very elegant to me.

13.5 Scepticism and Closure

Let's refresh a little bit what we said in part one about sensitivity and closure. Descartes worried about the following possibility. We do not have physical bodies, as we think we do. Rather, we are immaterial souls being deceived by a demon into thinking we live in a physical world. Call a person who is so deceived a **Cartesian victim**. Can you know you're not a Cartesian victim?

Here's an argument that you can know this. You know you have hands. (If you don't have hands, substitute some other body part that you do have for hands in this argument.) You can hold them up and see them. If you didn't have hands, if you lost them in an industrial accident or something similar, that would be quickly apparent to you. Cartesian victims do not have hands. And you know this too, by quick reflection on the nature of hands, and the nature of Cartesian victims. Since you have hands, and you know this general fact about Cartesian victims, you can infer you're not a Cartesian victim.

Here's an argument that you can't know this. Imagine you were a Cartesian victim. What would you think? Well, you'd think the very same things you actually think! You would have no idea about what was happening. So it isn't at all clear how you can know you're not a Cartesian victim.

There are three incompatible intuitions being appealed to here:

1. You can know you have hands, and you can know that no Cartesian victims have hands.
2. You can extend your knowledge by competent logical deduction from things you already know.
3. You can't know you're not a Cartesian victim.

If 1 and 2 are true, then you can know you're not a Cartesian victim by a simple deduction from the two things that 1 says you do know. So 1, 2 and 3 are inconsistent. One of them must go.

Sceptics say that 1 is false. Since 2 and 3 are true, and 1 is incompatible with 1, it follows that 1 must be false. This is the kind of scepticism Descartes worries about. (His own response to it involves arguing that we can know that God exists and would not permit us to be deceived, so 3 is false. This is not particularly popular with contemporary philosophers.)

Anti-sceptics say that 3 is false, perhaps because 1 and 2 are true. They concede that our belief that we are not Cartesian victims is insensitive, but think this is a case where the SSA goes wrong.

Sensitivity Theorists say we can keep the intuitions about knowledge; 1 and 3 are both true. The problem is 2. We can't always know the logical consequences of our knowledge. If you think our intuitions about particular items of knowledge are stronger and more reliable than our intuitions about theoretical claims like 2, this might look more attractive than either of the other options.

Principles like 2 are called **closure principles**. The SSA implies that a lot of closure principles are false. We can know things without knowing, or even being in a position to know, their logical implications. Nozick thought this was an advantage of the view, since he thought both the sceptical and anti-sceptical positions discussed here were implausible. But others have argued that this is a fatal weakness of the theory. We'll look at the problems that arise for the SSA in virtue of violating closure principles in the next chapter.

Chapter 14

Kripke's Response to Nozick

Saul Kripke wrote an influential critique of Nozick's theory that was widely circulated, and indeed presented at many universities, shortly after Nozick's theory appeared in *Philosophical Explorations*. This paper was finally published in the volume of his collected papers that was recently published (Kripke, 2011). Some of the criticisms Kripke raised appeared in print well before this, in some cases (perhaps many cases) under Kripke's influence. But it's easier for our purpose to simply focus on Kripke, rather than work through the various places that Nozick's view has been criticised over the past 30 years.

We'll look at six of the very many criticisms of Nozick's view that are in Kripke's paper. These six have the feature that they generalise to other theories of knowledge that use counterfactuals to get around Gettier cases.

14.1 Barns, Real and Fake

We skipped the fake barn case in the previous chapter because it turns out that how well Nozick's theory does on this score is a matter of some contention. At first, you might think that the theory does rather well, and indeed Nozick presented the case as one of the advantages of his theory.

Recall Barney driving through Fake Barn Country. He sees one of the real barns, but has no way of distinguishing it from the fakes all around him. He thinks, *That's a barn*. Nozick says this isn't knowledge for the following reason. If he hadn't seen this barn, he would have seen one of the nearby fakes. And then he would have formed a false belief. So his belief is insensitive to the truth. So it isn't knowledge.

But, says Kripke, this conflates two kinds of alternatives. It's true that in nearby physical space, there are a lot of fakes. But that isn't obviously relevant to ways in which Barney's situation could have been different. Perhaps there is no physical way to build a fake barn on the farm Barney is actually looking at. (Maybe the soil is weak enough that you need four walls to hold each other up, for instance.) Or maybe the owner of the farm thinks these fakes are a terrible idea, and refuses to countenance one. In that case, had there not been a barn here, there would have been a vacant field, not a fake barn. So Barney's belief is sensitive after all.

Philosophers will often describe ways the world could easily have been as *nearby parts of modal space*. This is often a useful metaphor. Think of the ways the world could have been as arranged as small planets in some giant multi-dimensional space. Worlds where things are only a little different, say the wind is blowing just a little less strongly, are nearby. Worlds where the sky is green and the grass is several shades of blue are far away. Intuitively, the counterfactual *Had it been that p, it would have been that q* require the nearby, in this sense, worlds in which *p* to be worlds in which *q*. It's a useful enough metaphor that we'll use it often in these notes.

But it can be a dangerous metaphor. The fact that there are fake barns near to Barney in physical space should not be taken to imply that there are worlds where this very field contains a fake barn nearby in modal space. The two 'spaces' may have very different distributions of facts, and indeed of barns, across them. We only get the view that the sensitivity view handles the fake barn case easily if we conflate these two senses of space.

14.2 Placebos and Rationality

Nozick's view of knowledge doesn't supplement the JTB condition with a sensitivity condition, it replaces the J condition instead. It could also have replaced the T condition, though Nozick did not. (Kripke opens his paper by chiding Nozick for including this redundant condition, but as we saw in the discussion of Zagzebski, the redundancy of the T condition is a virtue of Nozick's theory, if not perhaps of his presentation.) I argued above that we really should think knowledge entails justification. So obviously I don't think the lack of a J condition is a virtue of Nozick's theory. But it is worth noting how dramatically unjustified one can be while retaining knowledge in Nozick's sense.

A scientist is testing a new drug for disease X. The scientist knows that in previous trials of drugs for X, there was a large placebo effect. Some drugs that looked like they were making a real difference, were not performing notably better than placebo. The scientist has a gut feeling though that in the trial he is running, there would not be a placebo effect. He has no evidence for this gut feeling, and his gut feelings have usually been wrong in the past. But it is, somewhat surprisingly, true. (To make this plausible, assume the mechanism for the placebo effect in previous trials was that patients in the trial, both getting the drug and getting the placebo, got more medical attention, and this helped prevent other illnesses popping up. But this trial won't involve any more attention than patients otherwise would get.)

Since the scientist doesn't think there will be a placebo effect, he runs the trial without a placebo. The drug does well, in the sense that a higher than expected number of people were cured. But as noted, other drugs have had that property, but failed trials because they didn't beat the placebo. Still, the scientist believes that this drug does help cure X. And he's right. Moreover, if it didn't cure X, the trial would have turned out differently, and he wouldn't have believed it. So his belief is sensitive. But, says Kripke, we don't want to say it is knowledge. After all, his belief is based on a terribly designed trial.

The general point is that a belief can be extremely unjustified, and extremely irrational, but still count as knowledge for Nozick. If this happens, it seems that knowledge isn't doing the work we want it to do in justifying action and practical reasoning.

14.3 Red and Green Barns

As we saw, it was advertised as a virtue of Nozick's theory that it did not imply that knowledge was closed under logical entailment. That was how Nozick explained the attraction of scepticism; we really don't know we're not being deceived.

But the view has some downsides as well. And Kripke exploits one of those downsides rather effectively. Change the fake barn case in the following way. Barney really is looking at a red barn. And he thinks *That's a red barn*. The farm owner toyed with the idea of putting up a green fake barn there, but decided against it. There was no thought to the idea of putting up a red fake barn. Apparently in the local community, that sort of thing is Just Not Done. So all the nearby worlds in which there is not a real red barn, are worlds in which there is a green fake barn.

Barney knows none of this. When he drove past a green fake barn a few minutes ago, he thought *There's a green barn in that field*. He was, of course, wrong. It was a green fake barn. But he can't tell real from fake barns, and doesn't know that their color is a significant clue.

Now focus on Barney's belief that there's a red barn in the field. If that had not been true, if there had not been a red barn in the field, Barney would not have believed that there was one. In fact, he would have believed there was a green barn in the field. So Barney's belief that it's a red barn is sensitive. And it is true. So, by Nozick's standards, it is knowledge. (Again, Nozick's theory is a little more complicated, but the complications don't affect this case.)

Turn your attention to Barney's related belief that there's a barn (of some color or other) in the field. If that had not been true, if there had not been a barn in the field, Barney would still have believed that there was one. In fact, he would have believed it was a green barn. So this belief is insensitive. So it is not knowledge.

It seems, on Nozick's view, that the following two claims are both correct:

- Barney knows that there is a red barn in the field.
- Barney believes, but does not know, that there is a barn in the field.

And this, says Kripke, is absurd. It's one thing to say that knowledge isn't always closed under entailment. This is a striking view, but it might be the best way to save ourselves from scepticism. But to say someone can know a conjunction $p \wedge q$, yet not know *p*, seems absurd. And to say someone can know of a thing that it is a red barn, yet not even be in position to know it is a barn, also seems absurd. Something has gone badly wrong.

14.4 Negative Knowledge

The problem in the last section comes from something that was meant to be a strength of the sensitivity account; it violates closure. We saw in the previous chapter that this was meant to be how the appeal of a certain kind of scepticism was explained. According to the sensitivity theorist, for all we know we really could be victims of Cartesian demons. By definition, victims of Cartesian demons don't have hands; they are immaterial. So if you knew you had hands, which most readers do, and could come to know things by deduction from what you know, you could deduce that you're not the victim of a Cartesian demon. Nozick wants to block the last step.

In doing so he helps rescue some of the intuitions around scepticism. But he does so at a terrible cost. It becomes possible to know something is a red barn while not even being in a position to know that it is a barn. And we lose a lot of what we might call negative knowledge.

I know that I'm not exactly 1500 days old. If I were 1500 days old, I might easily have made a mistake about how old I am. Let *p* be the proposition that I'm not 1500 days old. If that were false, i.e., if I were 1500 days old, what would I have believed about *p*. Well, I might easily have made a mistake, so I might have believed *p* were true. So my belief that I'm not exactly 1500 days old isn't clearly sensitive. But it is a clear case of knowledge.

Or imagine that I'm introduced to a new student, and I come to believe *The person I'm being introduced to is not exactly 1500 days old.* If that were false, if the person I was being introduced to was exactly 1500 days old, I could easily have made a mistake about their age in days. (It isn't that easy to always tell by looking how old in days someone is!) So it isn't obvious that my belief that they are not exactly 1500 days old is sensitive. But this is something I can come to know by looking; my age-judgment isn't that bad.

To use a final example of Kripke's, say that someone is **supercredulous** if they believe everything. They believe that grass is green and that it is blue, and that it is not both green and blue, and so on. I'm not supercredulous. I don't believe grass is blue. Now let *p* be the proposition *I'm not supercredulous.* I think I know that. But what would happen if *p* were false. I would be supercredulous. So I would believe everything. So I would believe *p*. Hence my belief that *p* is not sensitive. But again, it does seem like something I can know.

There is a general pattern here. Often when it comes to 'negative' propositions, there is some 'positive' proposition that we both know, and are sensitive with respect to. The 'positive' proposition clearly and obviously entails the 'negative' proposition. But our belief in the 'negative' proposition is not sensitive. Now if we had a clear metaphysical division between negative and positive propositions, maybe we could impose sensitivty as a requirement only on positive propositions. But this division is not, to put it mildly, at all clear. (Hence the scare quotes earlier in this paragraph.) So we're left with the view that these cases look like knowledge without sensitivity.

14.5 Knowledge of the Future

There is one last big set of problems for the sensitivity account of knowledge. Let *p* be the proposition that the average temperature in Miami next summer will be above freezing. I know that *p* is true; it's the kind of knowledge I use in planning where to travel. But is my belief that *p* sensitive?

This turns on hard questions about how to interpret counterfactuals about the far future. We are familiar with counterfactuals about the past. If someone asks "What would have happened if such-and-such event had turned out differently?", we have a rough idea of how to figure this out. We go back to the event, or a time very closely before it, and mentally simulate how the world would have evolved had the event turned out in the stipulated way. The details are as always tricky, but it's a relatively well understood procedure.

It's a little harder to think about counterfactuals about the future. What we want to know is how to think about counterfactuals like:

- If the average temperature in Miami next summer were below freezing, things would be such that…

There are a couple of ways to interpret sentences like this. We can imagine continuing it by saying *People would be really surprised that Miami was so much colder than it had ever been*. Or we can imagine continuing it by saying *The climactic patterns of the world would have been very different to what they actually are for a long time*. The first kind of continuation is a 'forward-looking' reading. We imagine the world as it is, and continue it in some way where Miami freezes over in July. The second kind of continuation is a 'backtracking' reading. We try to imagine a world where Miami freezes over in July, then try to figure out what the present (and indeed past) would have to be like to have such a continuation.

I think the 'forward-looking' readings are the normal ones for counterfactuals. But if that's true, then basically none of our beliefs about the future are sensitive. Imagine that something I believe about the future were to turn out false. What would the world be like? Well, on the forward-looking reading, we keep the present fixed, and change the future. But the present includes my current belief about the future. So if the future were different, I'd still have the beliefs I actually have.

So we get a very odd result. I don't have any sensitive beliefs about the future. So if sensitivity is necessary for knowledge, I don't know anything about the future either. But surely I do; I know it will be warmer in Miami next July than it will be in Ann Arbor next February.

Now maybe we could try to understand 'back-tracking' counterfactuals a little better, and say that beliefs about the future have to be sensitive according to a back-tracking view of counterfactuals. This feels like changing the subject to me. We'd still be using words like "Would I have still believed this if it weren't true", but we wouldn't be using them with their normal meaning.

14.6 Summary

There are cases where sensitivity seems clearly insufficient for knowledge, as in the bad scientist case, and cases where sensitivity seems clearly unnecessary for knowledge, as in the negative knowledge, and knowledge of the future, cases.

Moreover, the big selling point of the sensitivity account, that it doesn't require closure of knowledge under logical deduction, is also a source of serious problems, as in the colored barn case, and again in the negative knowledge case. So we should not include sensitivity as an extra condition on knowledge, or as an alternative condition on knowledge.

Next time we'll look at a different condition, safety, that to some extent grew out of the failures of the sensitivity account.

Chapter 15

Safety

Through the late 1990s, three different epistemologists stressed the idea that for a belief to count as knowledge, it should not only be true, but be **safely** true. The idea is that if you could easily have gotten things wrong, then you don't count as knowing, but rather as having made a lucky guess.

15.1 Williamson

Timothy Williamson (1994) introduced a version of this idea in his work on knowledge of word meanings. But the idea behind his work on meaning could easily be generalised, and over a series of papers that eventually were collected into a book (Williamson, 2000), he showed how powerful the idea could be. The easiest way to set out his idea involves cases of estimating quantities. We've already used an example of this, when we looked at cases of estimating crowd sizes. But let's start here with Williamson's own example, of estimating tree heights.

Smith looks at a tree, and estimates how tall it is. The tree is in fact 50 feet tall, and Smith correctky estimates it is 50 feet tall. But Smith is not usually this accurate; she is often off by 10 feet or so in either direction. So while she gets some knowledge from the appearance and her estimation, she can't know it is exactly 50 feet tall. If she believed that, she would be right, but it would be a lucky guess. What she can know, presumably, is that the tree is between 40 and 60 feet tall.

Let's think of this in terms of what possibilities Smith is in a position to rule out. She can rule out that the tree is less than 40 feet tall. And she can rule out that it is more than 60 feet tall. But she can't rule out, or at least she can't rule out with knowledge, possibilities where the tree is 41 feet tall, or 42 feet tall or … Or 58 feet tall, or 59 feet tall. For all she knows, one of those possibilities is actual. And the reason those are open possibilities is because they are sufficiently similar to the actual world. If Smith tried to rule them out, she would be doing something unsafe.

Williamson uses these considerations to reject a principle on knowledge that had been (and to some extent still is) widely accepted. Assume Smith is reflective, and thinks carefully about what she believes and knows. Then you might think that this principle is plausible.

- If Smith knows that *p* , then Smith knows that she knows that *p*.

If we shorten 'Smith knows that' to K, and use → for if then, then we can state this as K*p* → KK*p*. For this reason, the principle is sometimes known as the KK-principle. The idea is that a reflective agent knows about their own mind, including their own knowledge. Williamson argued that cases like the one we have been discussing about the tree height show that it is wrong. Here's why.

Smith looks at the tree, and estimates it is about 50 feet tall. She comes to know that it is less than 60 feet tall. But it could be, for all she knows, that the tree is 59 feet tall. Imagine that were the case, and that Smith still believed the tree were less than 60 feet tall. Then her belief would not, in the imagined world, be safe. She wouldn't know that the tree is less than 60 feet tall, although she would truly believe this. So in that world, Smith doesn't know the tree is less than 60 feet tall.

Let *p* be the proposition that the tree is less than 60 feet tall. Smith knows *p*. But there's a world consistent with everything she knows, the world where the tree is 59 feet tall, where she doesn't know *p*. That is, although she knows *p*, for all she knows, she doesn't know it. She knows *p* without knowing she knows it. The KK-principle is false.

15.2 Lewis

David Lewis (1996) developed an important version of **contextualism** about knowledge. His idea was that to know that *p* was to be able to rule out all of the salient possibilities in which *p* is not true. What's a salient possibility? Lewis argued that this turned in part on what we were interested in when we were talking about what a person does or doesn't know.

In this course we are not going to examine contextualism, so we are going to set that part of Lewis's theory aside. But we are going to look at some of his other criteria for salience, two of which combine to form something like a safety criteria on knowledge.

Lewis put forward a number of rules in virtue of which possibilities could be salient. One of these is the Rule of Actuality; the actual world is always salient. This was what made it the case on Lewis's theory that one can't know a falsehood. Another is the Rule of Resemblence. Any possibility that is sufficiently like a possibility that is otherwise salient is salient. In particular, any possibility that sufficiently resembles the actual world is salient.

The effect of this is that to know that *p*, you have to rule out all of the similar worlds in which *p* is false. In practice, this works much like a safety condition on knowledge.

And Lewis points out that his condition, and by similar reasoning all safety conditions, solves two hard problems about knowledge.

One of these is explaining why Gettier cases are not cases of knowledge. They are not safe beliefs. It is easiest to see this using Lewis's approach. Consider the story that ends with Brown being in Barcelona. Smith believes that either Jones owns a Ford, or Brown is in Barcelona. That's right, since Brown is in Barcelona, although Jones doesn't own a Ford. But since Smith was just making up random place names to end sentences with, there's a very similar (from Smith's perspective) world where Jones still doesn't own a Ford, and Brown is in La Mancha. That world is salient, by the Rule of Resemblence. And Smith has no evidence that rules it out. So for all he knows, it is true. So Smith doesn't know that either Jones owns a Ford, or Brown is in Barcelona, since there is a world which for all he knows is actual, in which that proposition is false.

This is a very nice explanation of what is going on. Note what it isn't. It isn't an alternative to the JTB analysis of knowledge, one on which it falls out that the Gettier cases are not cases of knowledge. But it doesn't need to be that in order to be a satisfying explanation of the cases. And that kind of explanation is real philosophical progress.

Lewis makes another point in passing that seems significant about the role of the safety condition. He notes that the condition can also explain why lottery beliefs are not cases of knowledge. Imagine that Smith believes that his lottery ticket will lose, on the grounds that it has a terrible chance of winning, and he is right. Still, the world in which it does win is very similar to the actual world. Some ticket has to be drawn at random, and there's no deep reason why it should be the actual one rather than Smith's. So a belief that a lottery ticket will lose is not safe.

Now we have an explanation of why these beliefs about lottery tickets losing do not amount to knowledge, and a theory that links the lottery case to the Gettier cases. That's a reason to like the safety-based theory.

15.3 Sosa

Ernest Sosa (1999) argued that a safety condition on knowledge had all of the advantages of Nozick's sensitivity condition on knowledge, but without the downsides. We've already looked at two of the advantages: explaining why Gettier cases and lottery beliefs are not instances of knowledge.

Sosa discusses another significant advantage, that a safety condition on knowledge is compatible with the idea that competent deduction always extends knowledge. Remember one of the big problems for the sensitivity account: it is possible for a person to know that there's a red barn in the field, but if they try to deduce from that that there is a barn in the field, they will form a belief that does not amount to knowledge. The safety account does not suffer this problem, at least when the safety condition is properly formulated. We'll see this by working through the particular case of the red barn, but the underlying idea easily generalises.

Assume a person, call him Fred, knows there is a red barn in the field. So his belief that there is a red barn in the field is safe. Now what precisely might we mean by this? Sosa suggests a formulation in terms of subjunctive conditionals:

- If Fred were to believe that there was a red barn in the field, there would be a red barn in the field.

This isn't particularly easy to understand, in the way that subjunctive conditionals with true antecedents are generally not easy to understand.[1]. It's possibly a little easier to understand if we paraphrase this conditional as:

- Fred would believe that there is a red barn in the field only if there were a red barn in the field.

But even this isn't particularly easy to understand, I think. Perhaps it is easier to think directly about ways the world could be.

- In all possible worlds similar to the actual world in which Fred believes there is a red barn in the field, there is a red barn in the field.

But the problem with this way of thinking about safety is that it doesn't imply Fred can know there is a barn there. Change the case a little as follows.

The fake barns have all been removed, being abominations. Now if Fred sees a barn, he can easily recognise it as a barn. But Fred has the unfortunate habit of sometimes inferring that there's a barn in the field from the sound the wind makes. Fred believes that there is a distinctive wind sound that occurs only when near a field with a barn. Unfortunately, this is a complete mistake on Fred's part. It is easy to have this wind sound come up in a field with no barns at all. Fortunately, Fred isn't foolish enough to form beliefs about the color of barns from the sound of the wind. So while Fred often forms the mistaken belief that there is a barn in a field, he never (or at least never in realistic cases) forms the mistaken belief that there is a red barn in a field. He only forms the belief that there is a red barn in cases where he can see the barn, and his eyes are extremely reliable. So a belief that there's a red barn in the field might satisfy this safety condition, while if Fred were to infer that there's a barn, this belief wouldn't satisfy the safety condition. After all, Fred often believes there's a barn without there being one.

[1] A subjunctive conditional is a conditional of the form *If it were that p, it would be that q*. The antecedent of the conditional is the condition, i.e., the bit after the if. In this general schema, it is *p*. In the conditional immediately above, the antecedent is *Fred believes there is a red barn in the field*.

Sosa acknowledges that problems like this imply that we need to tighten up the formulation of the safety condition. In the next section, we'll look at two ways to do this.

15.4 Method Safety and Content Safety

Williamson's first pass account of safety looks like this:

- S's belief that p is safe just in case there is no possible world similar to the actual one in which S believes that p, but p is not true. (Williamson, 2000, 128)

As it stands, this clearly won't do, for reasons related to the barn example above. If Fred uses a very reliable method to form the belief that p, it doesn't seem to matter that he could easily have come to falsely believe p via a different route. So let's explicitly put a clause into the definition that rules out those cases.

- S's belief that p on basis B is safe just in case there is no possible world similar to the actual one in which S believes that p on basis B, but p is not true.

By **basis** here, we mean what supports S's belief. So believing something because you saw it involves a different basis to believing something because you heard it. And this restriction to a basis means that if a person tries to extend their knowledge by competent deduction, they will succeed. Assume that S knows that $p \wedge q$, and infers p from that. By assumption, their belief that $p \wedge q$ is safe, so there is no similar world in which they falsely believe $p \wedge q$ on their actual basis. Now when they infer p, the basis for this will be the basis for the belief that $p \wedge q$, plus the fact that $p \wedge q$ entails p. If that wasn't safe, there would be a similar world in which p was false, despite being believed on that very basis. But that's implausible, since if p is false, then $p \wedge q$ is false, and we said it couldn't be while being believed on that basis.[2]

But there is a problem for Williamson's view that Mark Sainsbury (1995) points out. Imagine that Smith is working on a tricky arithmetic problem, and getting tired, decides to just guess at the answer. Smith isn't particularly good at arithmetic guessing, but on this occasion he gets it right. This isn't knowledge; lucky guesses aren't knowledge. And it feels like it isn't knowledge for the same reason that Gettier cases and lottery cases are not knowledge; it's basically a guess. But note that it satisfies Williamson's safety condition. Mathematical truths are true as a matter of necessity. If Smith

[2] There is a wrinkle here that we haven't entirely closed off. What if there were similar worlds where p was falsely believed on basis B, but any world where $p \wedge q$ was believed on basis B was one where $p \wedge q$, and hence p, was true? That would imply that adding a belief that q to the world where p is falsely believed on basis B would make it much less similar to actuality. And since q is actually believed, it is hard to see how that could be the case. This isn't a proof that no such case is impossible, but it does suggest that closure violations will not easily crop up around here.

guesses that 578 times 613 equals 354,314, there's no similar possibility in which he falsely believes that. That's because there's no possibility at all where the proposition that 578 times 613 equals 354,314 is false. Sainsbury's discussion points at a refined version of the safety condition that fixes that problem.

- S's belief that *p*, formed via method M on basis B is safe just in case there is no possible world similar to the actual one in which S forms a belief via method M on basis B, and this belief is not true.

Note that Sainsbury actually talks about 'mechanisms' rather than 'methods', but the talk of methods has been picked up a lot in recent discussions of safety, especially in work by Duncan Pritchard (2009), and we'll work with it here. Thinking about safety in terms of methods can help illuminate some otherwise tricky cases.

Suzy goes to a new town, and sees two newspapers on a stand: the *Star* and the *Globe*. She knows nothing about either of them, so she buys the *Star*, and starts reading about her new town. Unfortunately, the *Globe* is a satirical newspaper, which mostly prints things that are false. And it is a good enough satire that Suzy would have been fooled had she read it. But fortunately, the *Star* is a good reliable newspaper, and she gets lots of true beliefs from it.

Does Suzy know a lot about her new town? I think it's easy to be of two minds about this. On the one hand, reading things in a reliable newspaper is a paradigm way of acquiring knowledge. On the other hand, Suzy could so easily have been reading the *Globe*, and been sucked into a morass of false beliefs.

The safety-based approach has the virtue, I think, of reflecting our indecision here. What's the method of belief formation that Suzy uses? Is it *Believe what's in the newspaper I just bought*, or *Believe what's in the Star*? In some sense, she uses both methods. The first method is extremely unsafe; it many similar possibilities it leads to false beliefs via *Globe*-reading. The second method is extremely safe; the *Star* is an excellent newspaper. Picking which is the real method Suzy uses feels like making a metaphysical distinction where there is no actual difference. So it's better to simply say that these are hard cases, and it may in some deep sense be indeterminate what Suzy knows. That, I think, is a perfectly fine outcome. It's better for a theory to stay silent, or even to suggest there is no determinately correct answer, than to commit to the wrong answer.

15.5 A Safety-Based Analysis

If we can argue that safety is a necessary condition on knowledge, as I've been suggesting we can, then could we possibly extend this to an analysis of knowledge? Perhaps we could analyse this the following way:

> *S* knows that *p* iff

1. *S* believes that *p*; and
2. *S*'s belief that *p* is safely true

There are two things to consider here. First, is this analysis extensionally adequate? That is, does it classify all cases of knowledge as knowledge, and of non-knowledge as non-knowledge? Second, is it informative or illuminating? In particular, can we come to know what knowledge is via using this analysis?

Even proponents of the safety condition on knowledge tend to think the answer to the second question is *No*. Here is a representative quote from Timothy Williamson:

> On my view, one knows in a given case only if one avoids error in cases similar to that one, but we cannot specify in a non-circular way how similar a case must be to fall under that condition, or what relative weights should be given to different respects of similarity. On the contrary … we should expect to have to use our understanding of knowledge to determine whether the similarity to a case of error is great enough in a given case to exclude knowledge. (Williamson, 2009, 305)

As Williamson goes on to say in the same article, we might also want to use our knowledge of which cases are and aren't cases of knowledge to work out what the *basis* or *method* behind a particular belief is. So Williamson thinks we shouldn't work out whether a person in a fake barn case has or lacks knowledge by first figuring out whether she makes a mistake in a similar world, and then using the safety condition. Rather, if we decide that she does have knowledge, we thereby get reason to think worlds where she makes a mistake are not particularly similar. Or if we think someone does know something on the basis of a particular perceptual experience, it follows that beliefs in similar worlds where she makes a mistake on the basis of similar perceptual experiences must in the relevant sense have a different basis. So if you think Suzy gets knowledge by reading the *Star*, you'll have to think that reading the *Star* is a different method, or perhaps provides a different basis, to reading the *Globe*.

In other words, any analysis of knowledge in terms of safety would be **circular**. We would be analysing knowledge in terms of something that can only be understood in terms of knowledge. But that doesn't mean that the safety-based approach to epistemology is useless. On the contrary, we have already seen two important insights it yields. One is an argument that the KK-principle is false. Another is an argument that there is a unified explanation of why Gettier cases and lottery cases are not pieces of knowledge. But to look for an analysis, we have to look elsewhere. We'll resume that search by looking at recent work by virtue epistemologists.

Chapter 16

Virtue Epistemology

We saw last time that one plausible lesson of the Gettier cases is that knowledge should, in some sense, be safe. If the agent could easily have gone wrong, then their belief is more like a guess than it is like knowledge. But we also saw that this isn't much help, at least on its own, in providing an analysis of knowledge. What counts as 'easily' going wrong will be difficult to specify without presupposing an account of knowledge.

One of the aims of virtue epistemology, certainly not the only one, is to try to say something illuminating about this point. Perhaps we can specify what it is for the agent to safely succeed by talking about the epistemic virtues of the agent, and pick out those virtues without antecedently having an account of knowledge. My own view is that while this project ultimately does not work, we learn an incredible amount about knowledge by thinking of it in these terms. Our focus here is on putative analyses of knowledge, and I'm not very sympathetic to the virtue epistemology story on this score. But the overall program of virtue epistemology is much richer than that, and it has taught us a lot about the value of knowledge, about the relationship between epistemic and practical skills, and about the relationship between evaluating inquiry and evaluating other projects. (For much more on these topics, see Ernest Sosa's three excellent recent books, *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I* (Sosa, 2007), *Reflective Knowledge: Apt Belief and Reflective Knowledge, Volume II* (Sosa, 2009) and *Knowing Full Well* (Sosa, 2010).)

Recall that Linda Zagzebski argued strongly that we could not understand knowledge as the conjunction of three independent conditions. If knowledge is true belief plus X, and X does not entail truth, then (for most plausible values of X), the disjunction of a true belief and an X belief will be a true X belief, but will not thereby be knowledge. Zagzebski's conclusion, which seems just right to me, is that if knowledge is true belief plus X, then the truth condition at least must be redundant. X must be a condition that entails truth. That way it is impossible for a belief to satisfy truth in virtue of one disjunct, and X in virtue of another part.

Here is one simple way to make sure the conditions are dependent. Add in a condition that the truth condition is satisfied because the X condition is satisfied. A very simple theory of this form says that *S* knows that *p* just in case *S*'s belief that *p* is true because it is justified. The basic virtue epistemolgy account of knowledge adds to that simple theory the idea that justification should be understood in terms of epistemic virtues. So the core idea is that the agent has some skills in inquiry, and she uses those skills to complete an inquiry, and she ends up getting the right answer because she used her skills. That's how we get to knowledge.

This handles the original Gettier cases rather nicely I think. The agent in question does have skills in inquiry, and uses them effectively. But he doesn't get the right answer because he uses his skills. Rather, his answer ends up being correct due to a sheer fluke, one that offsets the bad luck of doing the right thing but getting an unluckily bad outcome. The point here is not that virtue theories get the right answer, though that's obviously better than getting the wrong answer. It's that they offer something that feels like a plausible explanation of why these are not cases of knowledge.

16.1 Sosa's Virtue Epistemology

So the basic idea behind the virtue epistemology account of knowledge is that we understand justification in terms of epistemic virtues, and we say knowledge is belief that is not just true and justified, but that somehow the truth and the justification of the belief are closely related. In some sense, the justification of the belief explains its truth.

There have been a number of variants of this basic idea in the recent history of virtue epistemology. You can see the versions of idea already present in early work by Ernest Sosa (1991) and Linda Zagzebski (1996). It's obviously not a coincidence that Zagzebski both developed a theory of knowledge that adds this causal-explanatory connection between the belief and justification conditions, and showed that any theory which didn't posit such a connection would be bound to fail. But the most worked out version of virtue epistemology is due to Ernest Sosa, and we'll spend a bit of time looking at his theory. (Most of this section is from (Sosa, 2007).)

Sosa thinks that it is actually something of a mistake to talk about the theory of knowledge, as if there is one particular thing there to theorise about. Rather, he thinks we need to distinguish 'animal knowledge' from 'reflective knowledge'. The distinction he is drawing here traces back to themes Sosa finds in Descartes, which though fascinating are sadly outside our scope. But the main thing to note is that animal knowledge will turn out to be a considerably more substantial achievement than you might suspect from the name. Indeed, much of what we might ordinarily take to be knowledge is animal knowledge in Sosa's sense.[1]

[1] A brief historical digression. One of the early objections to Descartes's theory of knowledge was that it made knowledge much too hard. In particular, it required that the knower had a justification for believing in the reliability of their faculties, and Descartes thought the only possibly justifications went via arguments that we were endowed with those faculties by a loving God. But, critics alleged, an atheist mathematician seems to be in a position to know some basic facts about, say geometry. Descartes's

We'll start with Sosa's account of animal knowledge. We get to it via thinking through three ways in which a performance might be praiseworthy. We're going to eventually apply the insights here to the perforance a thinker goes through in coming to a belief, but the idea is to start with a quite general case. So we can follow Sosa, for instance, in starting with the case of an archer trying to hit a target.

The simplest way to evaluate the archer's performance is by asking whether it is **accurate**. That is, does the archer actually hit the target or not. A good archer may have a bad day, or might just have an arrow caught by an unlucky gust of wind, but still it is better to be accurate than not.

Another way we might evaluate performances is by asking whether they are **adroit**. Sosa says that a performance is adroit just in case it is a manifestation of competence. So the archer herself must be a competent archer for her performance to be adroit. (This connection to evaluating the archer herself is part of what makes this a virtue theory.) But that's not enough for the performance to be adroit. The archer may be slacking off, and may not have put any particular effort into this shot. Or she may have had a brain-freeze and not accounted for some obvious facts about the situation she is in. Either way, her performance will not be adroit, even though she is competent. It will be a performance by a competent agent, but it won't be a manifestation of that competence.

Finally, we can ask whether the performance is **apt**. For Sosa, an apt performance is one that is accurate because it is adroit. As above, this use of 'because' is factive; being accurate because adroit entails being both accurate and adroit. But it requires more than this too; it requires that the adroitness of the performance be at least a large part of the explanation for why it is accurate.

It's easiest to see how this might matter by considering cases where it fails. A competent archer goes through her usual pre-shot routine, and fires at the target. Unluckily for her, a sudden gust of wind pushes the arrow to the right. Luckily for her, the now offline arrow grazes a passing bird and is deflected back onto the target. It is an accurate shot eventually; there's the arrow in the bullseye. And it was an adroit shot; there's the competent archer doing everything she is supposed to do. But it wasn't accurate because it was adroit. It was accurate because of a bird that deflected it back on target.

response is to distinguish between two relationships we might stand in towards mathematical facts. Translating his terminology here is not without pitfalls, but roughly we might say that the atheist mathematician can, according to Descartes, be aware of mathematical facts, but cannot have true knowledge of them. That is meant to explain the intuition about the case, that there are after all better and worse, more and less knowledgeable, atheistic mathematicians, without conceding the distinctive Cartesian claim about what atheists can really know. Sosa's animal knowledge is a descendant of the state Descartes thinks the atheist mathematician can be in, while his reflective knowledge is a descendant of the state Descartes thinks the atheist mathematician can't be in. Sosa, like most people, doesn't agree with Descartes's broader views about what atheists can and can't do, but he does think Descartes had his finger on an important distinction here.

The analogy to Gettier cases should now seem clear. An agent making an inquiry is engaged in a kind of performance. The inquiry might or might not be accurate. That is, the agent might or might not get it right. And the inquiry might or might not be adroit. That is, it might or might not manifest the agent's competency in inquiry. Finally, this manifestation of competency might or might not be the explanation for the success of the inquiry. This is what fails in Gettier's examples. The inquirer is both accurate and adroit, but the adroitness does not explain the accuracy.

Sosa thinks that performances in general can be judged by this AAA standard. Inquiries that meet this standard are animal knowledge. The more demanding standard of reflective knowledge requires that the agent not just have animal knowledge (i.e., apt belief), but the agent have an apt belief in the aptness of her belief. To a first approximation, to reflectively know something is to have animal knowledge that you have animal knowledge of it. We'll confine most of our attention in these notes to animal knowledge, since it is the notion that best matches up with the cases that we're considering.

Sosa's theory has a lot to like about it. It offers an illuminating account of several puzzle cases, including the original Gettier cases. But, like any good philosophical theory, it faces challenges too.

16.2 Sosa's Theory and Gettier-Like Cases

John Turri (2011), in the course of developing a refinement to Sosa's theory, offers this case as a puzzle for it.

> A competent, though not materful, inspection of the crime scene would yield the conclusion that a man with a limp murdered Miss Woodbury. Holmes saw through it and had already deduced that Dr. Hubble poisoned the victim under pretense of treating her.
>
> Holmes also recognized that the scene would fool Watson, whose own inspection of the scene was proceeding admirably competently, though not masterfully. Watson had, after years of tutelage, achieved competence in applying Holmes's methods, and while Holmes was no sentimentalist, he didn't want Watson to be discouraged.
>
> "Look at him," Holmes thought, "measuring the distance between footprints, noting their comparative depth, and a half dozen other things, just as he ought to. There's no doubt where this will lead him – think how discouraged he will be." Holmes then resolved, "Because he's proceeding so competently, I'll see to it he gets it right!"

> Holmes sprang into action. Leaving Watson, he hastily disguised himself as a porter, strode across the street to where Hubble was, and kicked him so hard that Hubble was thereafter permanently hobbled with a limp. Holmes then quickly returned to find Watson wrapping up his investigation.
>
> "I say, Holmes," Watson concluded triumphantly, "whoever committed this brutal crime has a limp."
>
> "Capital, Watson!" Holmes grinned. "I'm sure he does." (Turri, 2011, 5)

Turri's thought is that Watson gets a true belief, and he gets a true belief in large part because of his intellectual virtues. If Watson had been investigating the case in a shambolic manner, Holmes would not have taken pity on him, and would not have made Hubble have a limp. So Watson's belief only ends up being true because he was such a good inquirer.

Still, says Turri, this doesn't feel at all like knowledge. Watson made a mistake, and Holmes adjusted the world to fit Watson's mistake. That seems incompatible with knowledge. Although Watson's intellectual virtue led to his success, it seems it led to success for the wrong reasons.

Turri suggests a refinement to get out of this problem. Say that knowledge is not just belief that is true because of virtue, but belief whose truth **manifests** competence. The distinction Turri is relying on here may be illustrated with an example involving some other kind of disposition. We'll use fragility.

Some things are fragile. That is, they are disposed to break on being struck, even when they are struck lightly. If I take a wine glass, and drop it onto a concrete floor from waist height, it will break. And it will break because it is fragile; lots of things do not break when they hit that kind of floor with that impact. Moreover, the breaking will manifest the fragility of the wine glass. When we say the wine glass is fragile, what we mean is that it breaks in situations just like that, where many other things would not break.

Now change the situation, so the wine glass is sitting comfortably on a table. Unfortunately, there is a mad dinosaur, armed with a sledgehammer, roving the corridors. The dinosaur is both incredibly strong, and obsesses with a hatred for all fragile things. When he sees something fragile, he smashes it (and everything around it) with his sledgehammer. He sees the wine glass, pulls out the sledgehammer, and smashes the glass, the table it is sitting on, and the floor the table is standing on. The wine glass broke in part because it is fragile. It is the fragility, as opposed to say the transparency, of the wine glass that set off the dinosaur's rage. But the fragility of the glass is not manifest in its breaking. It was hit with a force that breaks non-fragile things too, like tables and floors. So there is a difference between something breaking because it is fragile, and something whose breaking manifests its fragility.

Turri thinks that by adjusting Sosa's theory to account for this difference, we can avoid the problem his Holmes-Watson case raises. So he proposes that knowledge is true belief whose truth is a manifestation, not just a result, of the believer's intellectual virtues.

There is still a problem as Ian Church (2013) notes. The problem concerns cases where the believer's competence get her most, but not all, of the way to the correct answer. Here is the case that Church uses to illuminate the problem.

> David is an expert botanist, able to competently distinguish between the over 20,000 different species of orchid. David is presented with an orchid and asked to identify its species. Using his amazing skill, he can clearly tell that this particular orchid is either going to be a *Platanthera tescamnis* or a *Platanthera sparsiflora* (which look quite similar), and upon even further expert analysis he comes to the conclusion that it is a *Platanthera tescamnis* orchid, which it is. However, Kevin, David's nemesis and an expert botanist in his own right, decided the night before to disguise the *Platanthera tescamnis* orchid to look like a Platanthera sparsiflora orchid. Thankfully, however, Alvin, David's other expert botanist nemesis (who is conveniently not on speaking terms with Kevin), decided to try to trick David in the same way – arriving shortly after Kevin left, perceiving that the orchid was a *Platanthera sparsiflora*, and disguising Kevin's disguise to look like a *Platanthera tescamnis*, which it happens to actually be. (Church, 2013, 174)

Church's case is fairly contrived, but there is, as he points out, a fairly general point here. Any successful inquiry will owe its success to a number of factors. These will include, but rarely be limited to, the intellectual virtues of the inquirer.

If I'm trying to figure out who the murderer is, and you notice the missing vase in the dining room, which gives me the vital clue I need, then the success in the inquiry is partially due to my deduction, and partially to your observation. Or if I stumble across a vital clue while looking for something else, my good luck in stumbling across the clue is part of the explanation for the successful inquiry.

Those two cases are, at least when filled out in obvious ways, cases of knowledge. It isn't required for knowledge that the **only** explanation of the inquiry's success is the inquirer's intellectual virtue. But if allow that intellectual virtue only be a part of the explanation for success, then we open the door to cases like Church's, where Gettier-style luck is the explanation for the rest.

In short, the virtue epistemologist faces a dilemma. If they say that knowledge is true belief that is **solely** in virtue of intellectual virtue, they get the absurd result that we can never gain knowledge through co-operative ventures, or through fortuitous discoveries. But if they say that knowledge is true belief that is **largely** in virtue of intellectual virtue, they get the undesired result that cases where a part of the explanation is Gettier-style luck are still knowledge. It seems we haven't yet figured out how to deal with all Gettier-style cases via consideration of epistemic virtues.

16.3 Other Puzzles for Sosa's Theory

We'll close with three puzzle cases for Sosa's approach, and for refinements of it like Turri's.

Harry walks out of his front door and sees a squirrel run up a tree. He comes to believe that there's a squirrel up the tree. And there is. This seems like a paradigm case of knowledge. But it isn't obvious that we can fit it into Sosa's AAA (Accuracy-Adroitness-Aptness) structure. After all, it isn't clear that the belief is in any sense the result of an inquiry, so it isn't clear how we can ask whether the inquiry was accurate, adroit, or apt. And even if there was an inquiry, it feels so automatic, so unlike the intentional actions like shooting an arrow that Sosa takes as his paradigms, that it feels odd to evaluate it in the same way that we evaluate intentional actions.

But it isn't clear that these are deep problems for Sosa's theory. For one thing, the AAA model is strikingly general, and can apply to paradigm instances of unintentional behavior. We can apply it, for instance, to evaluations of my kidneys. Do the liquids that go into my kidneys get sorted into the places they should? That's a question of accuracy. Is the sorting a manifestation of the kidney's competence, or good functioning? That's a question of adroitness, though admittedly the term feels a little odd. And is the sorting accurate because my kidneys are functioning well? That's a question of aptness. (Exercise for the reader: Construct a case, preferably realistic, where the kidneys are accurate and adroit but not apt, i.e., not accurate because adroit.)

In any case, the things we do when looking at a tree are more intentional than the things we do when our kidneys filter liquids. It's true that we don't decide, "Ah, now is the time to believe a squirrel is up the tree", in the way we decide "Ah, now is the time to shoot". Beliefs are not volitional – they are not the consequence of volitions – in the way that shootings are. But we do have a standing power to refuse to take some appearances at face value, and to withhold judgment until we have more evidence. Relatedly, we have a standing power to continue an investigation, or to regard it as settled, even in clear cases like this. In Harry's case it would be a mistake to exercise those powers. It's clear what the facts are, and further inquiry would be pointless. Harry manifests a competence by taking things at face value in appropriate circumstances. And because he manifests that competence, he concludes an inquiry successfully. So the AAA theory says he really does get animal knowledge here, as it should.

Tom is driving through Fake Barn Country. As usual, most of the things around here that look like barns are in fact fake. Tom is, in general, very good at telling real from fake barns and more generally barns from non-barns. He looks at a barn, and concludes *That's a barn*, thereby manifesting these competencies. And it is a barn, the only real barn in a sea of fakes. Duncan Pritchard (2009) has argued that this implies that, according to Sosa, Tom knows that it's a barn. But, he says, this is implausible.

As we noted in our earlier discussion of the fake barn case, it is easy to be pulled in either direction by this example. And we've now got some more evidence as to why we might be pulled to say it is knowledge. Tom really is good at telling barns from non-barns. He uses this skill, and gets the right answer, for completely the right reasons. That looks like knowledge. He's lucky to know; he could easily have been looking at a fake, but a lucky success is still a success. Sosa offers a version of this reply, although he also offers an argument that Tom can't have reflective knowledge that it's a barn, and suggests this might explain some of the intuitions.

John Greco (2009) suggests an alternative response. It is plausible that competence in general is environment relative. For example, someone who is a great race car driver on one kind of track will not always do well on a different kind of track. Perhaps we can say that Tom isn't really competent at detecting barns, at least in this environment. That way we can respect the intuition that it isn't knowledge, while holding onto the AAA account of knowledge.

Finally, consider the margin of error case. This one is, I've argued elsewhere, more of a problem for the virtue theorist (Weatherson, 2013). To slightly alter the example, Richard is trying to figure out how many people are in Michigan stadium. He has decided, after careful examination of his record, that he shouldn't trust his initial judgments to be accurate to within any finer margin than 2000. And this is a good thing to conclude; he's rarely off by more than 2000, but he does get the value a little wrong from time to time. One day, there are 83,100 people at the stadium. Richard looks around, and estimates that there are about 85,000. Knowing what he does about his own guesses, he concludes that there are more than 83,000 people there. Is this knowledge?

I think there's a reasonable case to be made that Richard's performance satisfies the AAA-criteria, but he does not get knowledge. He is accurate; there are more than 83,000 there. And he maniests competence; he's good (to within 2000) at guessing crowd sizes, and he uses his skills appropriately here. And he gets the right answer because he manifests his competencies. If he wasn't so good at estimation, or at considering his own track record, he would have got the wrong answer, or more likely no answer. So his performance is apt. But it seems ever so lucky! Had there been a few more people at the concessions, or in queue for the bathroom, or in any of the other places that Richard can't see directly and has to really guess about, he would have got the wrong answer. His answer is, in the terminology we've used already, extremely unsafe.

So what these margin of error cases show is that an AAA-performance need not be safe. It could be that a competent agent does everything right, but has nearly enough bad luck that things nearly go wrong. So if you think safety is a requirement on knowledge, you should conclude that the AAA-account, as illuminating as it is in many cases, isn't the full story.

16.4 Conclusion

We're going to leave the analysis of knowledge on this somewhat inconclusive note. As Ichikawa and Steup (2013) suggest, it does seem like if there is an analysis of knowledge, we haven't found it yet. But this does not mean that the history of attempts to analyse knowledge is a history of failure. Thinking through the variety of counterexamples to the JTB condition teaches us a lot about the contours of the knowledge relation. And thinking through the sensitivity, safety and virtue accounts teaches us a lot more about the theory of knowledge, even if we have no way of compressing our learning into a straightforward analysis. There's no shame in that. A civil war scholar can't compress everything she knows about the war into a one paragraph analysis of the war. We shouldn't expect a scholar of knowledge to do better.

Part III

Social Epistemology

Chapter 17

Testimony

There are some things that you know because you are experiencing them right now. There are other things you know because you have experienced them, and have retained a memory of them. And there are yet more things you know by reasoning from those pieces of experiential knowledge.[1]

But, at least at first glance, it seems this story so far leaves out the vast bulk of what you know. For it leaves out what you know because you were told it. In a world as 'connected' as this one is, one very central way we get to know about the world is by being told facts by other people who are, or at least were, better positioned to know them.

Historically, most work in epistemology did not may much attention to the role of testimonial knowledge. Our main examples have involved knowledge by perception, or by inference. To the extent that testimony has come in, it has been thought to be just another means of getting knowledge indirectly. Our implicit practice has been to treat the following two scenarios as equivalent.

> **Learning from Machines**
> Annelies wants to know how many people are in the coffee room. So she turns on a machine that does a heat scan of the coffee room, and increments a counter every time in finds a roughly person shaped, roughly person temperature, object. The machine is actually quite good, unless one goes out of one's way to trick it. (Humans do have relatively distinctive heat profiles, relative to most things you'd find in an office setting.) The machine says there are five people in the room, and it says this because there actually are five people in the room, and Annelies believes there are five people in the room as a consequence.

[1] As well as the chapter on testimony in Nagel (2014), the next four chapters draw heavily on the survey by Jonathan Adler (2015).

> **Learning from People**
> Bridget wants to know how many people are in the coffee room. So she asks Claudia to go and look in the coffee room and tell her. Claudia is perfectly competent at counting how many people are in a room, though she does occasionally make mistakes. And she does, like everyone, occasionally lie. But she doesn't have any particular reason to do so now. She sees there are five people in the room, and reports to Bridget that there are five people in the room, and Bridget believes there are five people in the room as a consequence.

A lot of work in epistemology has taken the two cases to really be on a par. Both Annalies and Bridget form a belief about something they cannot directly perceive by using an external device that is both reliable and known to be reliable. Annalies uses a purpose-specific device; Bridget uses one that has a lot of other functions as well. But that doesn't seem to make a big difference. The epistemology of scales doesn't really change if the scale only measures weight, or if it is one of the fancy new scales that measures other attributes.

There is a big moral difference between the cases. Bridget and Claudia, both being human, stand in a moral relationship that Annalies does not stand in to her machine. Those moral questions will become central to our interests in a few chapters. If Annalies simply decides to not believe the machine, that doesn't seem to have moral consequences. If Bridget simply doesn't believe Claudia, then it seems appropriate for Claudia to feel upset, or offended, at her. These reactive attitudes on Claudia's part are arguably tracking something that is morally rather significant, and that's already a hint that the two cases are a bit different.

When Bridget does believe Claudia, there is a sense that they are acting as a team, in a way that Annalies and her machine are not really a team. And that may matter too. Perhaps the fact that they are a team means that justifications can be shared amongst them. If we view Bridget primarily as an individual, and Claudia as a machine for helping her, it seems sensible, perhaps obligatory, to ask what justification Bridget has for relying upon this particular machine. But if we view the two of them as a team, then perhaps those questions become less urgent; Claudia's justification for believing there are five people in the coffee room is sufficient for the team to have similar justification.

So one of our big questions will be whether it is right for humans to treat their fellow humans as 'truth-o-meters'; machines for detecting truth. (Compare the term 'multimeter', which I guess could in theory be used for anything that detects multiple properties, but in practice is used for machines that are capable of measuring a number of properties of electric circuits.) If it is, then the epistemology of testimony will just be a special case of the epistemology of measuring devices. That's an interesting area of epistemology, I think, but it's one we've already said a lot about. But if it is not, then testimony becomes an independently interesting subject.

At some level here, our primary concern will be the meta-question: *Is testimony an independently interesting subject?* I hope at least that question is interesting!

17.1 Testimonial Scepticism

One tradition within work on scepticism, a tradition Nagel sees best exemplified in John Locke, holds that we never really know things on the basis of testimony. Of course, sceptics of all sorts of varieties will think we don't know things on the basis of testimony, but only because of a more general consideration. The question at hand is, if you're not a sceptic about the external world generally, should you be a sceptic about testimonial knowledge. And it does seem that in general the answer should be no.

It's true that other people can be wrong about the world. And even when they are right, they might go out of their way to deceive us. So testimony is no foolproof method. But if fallibility is to rule out knowledge, then we don't have knowledge of the external world either, since our senses are fallible. That can't be enough reason.

One possible worry is that our senses never have reason to lie to us, so there is a difference in kind between the ways senses can fail, and the way that testimony can 'fail'. But it isn't clear why this should be a reason for doubting others in general and not our senses. What matters is whether they fail, not necessarily why. And in any case, it isn't even clear there is a real disanalogy here. Perhaps our senses do deceive us, by design, in certain cases. I gather that the apparent brightness of very faint objects is much greater than it would be were our eyes were more faithfully representing brightness. This is for a good reason - it gives us better awareness of hard to detect objects. But it can be misleading. (I gather this was an early objection to Copernican theories of planetary motion; Mars and Venus are apparently brighter than you'd expect them to be given how far away they get from the earth. But this appearance goes away if you use a light-meter and not just human eyes.)

In practice we have two big tests for knowledge. One is verbal; we ask *Does the sentence "She knows that thus-and-so" seem true?*. The other is in terms of action; we ask *Is p something she can take to be a reason to act?*. It is plausible that the verb "know" picks out knowledge, and it is plausible that our factual reasons to act include all and only things we know. (Not that neither claim has been denied, but they are both plausible.) And by either test, we have lots of propositional knowledge. We say things like, "Bridget knows that there are five people in the coffee room." And if there is something that Bridget is supposed to do when there are five people in the coffee room, we say that she now has a reason to do it, a reason she lacked before Claudia spoke.

So Lockean scepticism seems misplaced. There isn't a good reason to be a testimonial sceptic and not a general sceptic. And our general pratice is to act as if testimony provides knowledge. So that's what we'll assume from now on.

17.2 Reductionism

So we'll assume from now on that we do have a lot of testimonial knowledge. The big question then becomes how we have so much of it. And there are two prominent schools of thought:

- **Reductionist** theories say that we have testimonial knowledge in virtue of other epistemic skills and capacities that we have. Our capacity for learning from testimony is not, they say, basic in the way that perception or inference is a basic skill.
- **Anti-Reductionist** theories say that testimonial knowledge. We can learn that *p* because we are told that *p* without having non-testimonial reasons for thinking that the speaker is reliable, or a truth-indicator.

The simplest way to be a reductionist is to think that the following process is typical (if mostly sub-conscious). We'll assume that there is a speaker *S*, a hearer *H*, an utterance *U*, and a proposition *p* that *H* comes to know after hearing *U*. The reasoning may go like this.

1. *S* uttered *U* (by perception)
2. *U* means that *p* (by background knowledge of language)
3. If *S* is sincere, then *S* believes that *p* (inferred from 1 and 2)
4. If *S* is sincere and reliable, then *p* is true (inferred from 3)
5. *S* is sincere and reliable (key background knowledge to testimonial learning)
6. So, *p* is true (from 4 and 5)

On this model, 1, 2 and 5 must be known, or at least justified, prior to knowledge of *p*. And the way they are known is through perception and reasoning. But if they are known, then we don't need anything extra. After all, the inference from 1, 2 and 5 to 6 is something that the faculty of reasoning can underwrite. So testimony is not in any way basic.

There are a couple of interesting variants of this basic reductionist position. Frank Jackson thinks that our testimonial reasoning often goes via a consideration of the evidence of the speaker. Here is how he describes the way testimony works.

> Why should you ever accept what I say, unless you already did so before I spoke – in which case speech is a luxury? What is wrong with your always either saying ‘I already believe that’ or saying ‘How interesting, here’s a point on which we disagree’. The answer cannot be that you are taking me to be sincere. Sincerity relates to whether my words mirror my beliefs, whereas we are wondering why your discovery of my beliefs, perhaps your presuming sincerity on my part, should ever make you alter your beliefs. Sincerity relates to whether you should infer prior agreement or disagreement in beliefs, not to whether posterior adjustment of belief is in order.
>
> The reason posterior adjustment in belief may be in order is that hearers (readers) sometimes have justified opinions about the evidence that lies behind speakers’ (writers’) assertions. You assert that P. I know enough about you, and your recent situation, to know (i) that you have evidence for P, for you would not otherwise have said it, and (ii) that your evidence is such that had I had it, I would have believed P. I borrow your evidence, so to speak. Typically, I won’t know exactly what your evidence is. Perhaps you visited a factory and came back and said ‘The factory is well run’. I don’t know just what experiences you had there – just what you saw, heard, smelt and so on – but I know enough to know that had I had these experiences – whatever exactly they were – I too would have come to believe the factory well run. So I do.
>
> This is what goes on when we acquire beliefs from books. A history book says that Captain Cook landed first at Botainy Bay. I may not know just what evidence the writer had for writing this, but I believe that had I had it – whatever exactly it is – I too would have believed that Cook landed first at Botany Bay. I borrow this evidence, without knowing exactly what it is, and in this way an epistemological division of labour is achieved. Imagine the work (and invasion of privacy) involved if we all had to duplicate each other’s evidence.
>
> Of course, I may not come to believe exactly what the speaker or writer believes. A friend returning from overseas may say to me of a certain country ‘It is very well run’. I may know enough of my friend to know that experiences that would make him say that, are the kind that would make me say ‘Dissent is suppressed’. In this case, I will borrow his evidence to arrive, not at what he believes, but at what I would have, had I had his experiences. (Jackson, 1987, 92–3)

So Jackson would modify the inference above in the following way.

1. *S* uttered U (by perception)
2. U means that *p* (by background knowledge of language)
3. If *S* is sincere, then *S* believes that *p* (inferred from 1 and 2)
4. *S* is sincere, and her 'evidence function' is *E*. (key background knowledge to testimonial learning)
5. So *S*'s evidence includes $E(p)$.
6. So $E(p)$ is true.
7. So, if my evidence function is a lot like *S*'s, *p* is true.

The 'evidence function' here is the function from a belief to the kind of evidence we can reasonably assume that the speaker would have if they believe *p*. Part of Jackson's argument for this being the way testimony works is that there are some surprising features of language that seem very sensitive to evidence. Jackson is arguing in English, though I gather this argument would be even more compelling in many other languages.

Imagine that you go to a UM basketball game, and you see UM win. The next day a friend, who doesn't know you were at the game, asks if UM won last night. You say "ESPN said that they did". You know that ESPN (at least if you include its website, seventeen TV stations or whatever it is these days) will report every result and get it right. So you know that what you say is true, and good evidence for what your friend is interested in. But it's a weird thing to say. Why is that? Jackson says that it is because it misleads your friend into thinking that your evidence is that you saw the result on ESPN. And that makes the utterance very weird because it really matters what your evidence is, in the way this theory of testimony suggests it is.

I think this is a very interesting theory of testimony, but it has not been particularly central to recent work in epistemology. Another theory that I think of as broadly reductionist (though perhaps not everyone agrees) is Jennifer Lackey's "Learning from Words" theory (Lackey, 2008). Where Jackson adds a complication to the 'mind-reading' stage of testimonial learning, Lackey takes one way. She thinks that hearers typically learn directly from the words speakers utter, without having to go via reading into the beliefs of the speakers. The reason for this is not a generalised scepticism about mind-reading. It is rather that she thinks mistakes about mind-reading do not typically undermine testimonial knowledge, but they would if this was a crucial step in the process of testimonial learning. So we can represent Lackey's process as follows.

1. *S* uttered U (by perception)
2. U means that *p* (by background knowledge of language)
3. If *S*'s words are reliable, then *p* is true (inferred from 1 and 2)
4. *S*'s words are reliable (key background knowledge to testimonial learning)

5. So, *p* is true (from 3 and 4)

A key motivation for Lackey's theory comes from her important objection to anti-reductionist theories. So we'll come back to it after discussing those theories.

Chapter 18

Transfer and Its Discontents

As Jennifer Lackey (2006) notes, many contemporary theories of testimony are built around the idea that testimony is a means of transferring something. What's transferred may differ from theory to theory: some say information, others belief, others knowledge, others justification, others responsibility, and so on. But the core idea is that one effect of testimony is to put the hearer, who we'll call *H*, in the same position that the speaker, who we'll call *S*, is in, in some crucial respect.

18.1 Features of the Transfer Model

The transfer model has a few features that separate it from any of the reductionist models of testimony that we've seen above.

First, it makes much testimonial knowledge **non-inferential** in a key way. *H* doesn't have to hear that *S* said that *p*, and then infer from facts about *S*'s reliability to the truth of *p*. He[1] simply adopts *p* as true. The way I like to conceptualise this is that in these cases, *p* isn't part of *H*'s conclusions, it goes directly into his evidence. (This is controversial; some anti-reductionists do not think that explaining testimony in terms of evidence at all is helpful.)

This has two consequences for testimonial knowldge. It means that mistakes about the speaker are not things that defeat knowledge, in the way that they might be for a reductionist. On Jackson's model of testimony, for example, beliefs about how a speaker came to her knowledge is an input to testimonial conclusions. Typically, we think mistakes in premises mean that the conclusion is not known. The reductionist thinks that this isn't important. And, perhaps more significantly, it means that ignorance about the speaker does not defeat knowledge. The hearer does not have to know whether the speaker is reliable in order to get testimonial knowledge.

Second, it makes the epistemology of testimony **social** in a much thicker sense than the reductionist allows. Consider the following two-part case.

[1] For ease of reference, we'll make the *h*earer be *h*e, and the *s*peaker be *s*he.

> *S* is an excellent detective, and *H* knows her to be an excellent detective. *S* investigates the theft of the cookies and the cake, and concludes that the butler stole the cookies, and the gardener stole the cake. Both conclusions are true, though only the first is well arrived at. The conclusion about the gardener involved atypical sloppiness on *S*'s part; jumping to a conclusion where she should have, and normally would have, paused to collect more evidence. *S* tells *H* both of her conclusions, and *H* believes both conclusions. What should we say about the quality of *H*'s two beliefs?

For the reductionist, *H*'s two beliefs (that the butler stole the cookies, and that the gardener stole the cake) are on a par. They are arrived at through the same source, namely testimony from *S*. In terms of what *H* does, from the perspective of *H* as an individual, the two cases are exactly the same. But the anti-reductionist is not committed to this conclusion. The anti-reductionist can say that the two beliefs differ in virtue of the fact that *S*'s beliefs differ. If *S* knows that the butler is guilty, but does not know that the gardener is guilty, then in virtue of that fact alone, perhaps *H* knows that the butler is guilty, but does not know the gardener is. If *S* does not have knowledge about the gardener to transfer, then *H* does not receive this knowledge. (Though, of course, he may think he does get knowledge.)

Third, anti-reductionists can say that transfers are to some extent voluntary and intentional. Not all anti-reductionists say this, but some take it to be an important part of the theory.

> *S* is an excellent detective, and *H* knows her to be an excellent detective, as does *H2*. Indeed, *H* and *H2* know the same facts about *S*'s background. *S* investigates the theft of the cookies, and comes to know that the butler did it. *S* then tells this to *H*, and she is overheard by *H2*, who she did not intend to tell. Both *H* and *H2* come to believe that the butler stole the cookies.

The reductionist is committed to the view that we should evaluate *H* and *H2*'s beliefs the same way. They have the same evidence; they both heard *S* say the butler did it, and they both have the same background knowledge of *S*'s reliability. The anti-reductionist is, at least, not committed to this conclusion. They may say that *H* has non-inferential knowledge of the butler's guilt, but *H2* has to infer it.

Perhaps the most dramatic suggestion of how an anti-reductionist model of testimony can make the epistemology of testimony social comes from Tyler Burge (1993). He considers the following kind of case.

> *S* is a great mathematician. She discovers that *p* by purely armchair, a priori methods. She then tells *H* that *p* is true. He knows that *S* is a great mathematician, and can understand the statement of *p*, but he has no way to follow the proof. Still, he comes to believe *p* on the strength of *S*'s claim.

Question: Is *H*'s knowledge that *p* a bit of a priori knowledge, or a bit of a posteriori knowledge? Burge argues that a non-reductionist theory of testimony should, correctly, say that it is a priori. *S* has a priori knowledge that *p*, and she transfers that a priori knowledge to *H*. On a reductionist theory of testimonial knowledge, then a crucial input to *H*'s knowledge that *p* will be the a posteriori knowledge that *S* is a great mathematician. So the knowledge that *p* will be a posteriori. Burge's claim has not been widely accepted, even among non-reductionists - see Anna-Sara Malmgren (2006) for some critical comments.

One way to think about all these features is in terms of responsibility. There is, arguably, a sense in which agents are typically responsible for their mistaken beliefs. Whether 'mistake' here means that the belief is false, or that it is unjustified, is a matter of some contention, as is the general notion that beliefs are the kind of thing we can be responsible for. But assume for now that you like the general idea of responsibility for belief. On a reductionist picture of testimony, the believer is always responsible for their own beliefs; if a mistake is made, it is their belief. One way to motivate anti-reductionism is by thinking that sometimes, the person responsible for the mistake is not the hearer, but the speaker. That is, there might be times when a mistaken belief is held, and there is responsibility for the mistake, but the responsibility falls on someone other than the hearer. This isn't a wholly unnatural view; if someone you reasonably trust tells you something which turns out wrong, you may feel excused for the mistake by the fact that someone else, namely the speaker, is responsible. Conversely, you may feel responsible for the mistakes of others than flow from mistaken information that you send to them.

18.2 Qualifications to the Transfer Model

To say that knowledge, or other epistemic properties, can be transferred by testimony is not to say that they always are transferred. There are three prominent ways in which theorists say that the transfer can be blocked.

First, the hearer might simply reject the testimony. If *H* doesn't believe what *S* says, then he doesn't acquire knowledge from *S*. Perhaps he acquires a reason to believe the truth of what *S* says, but it doesn't really matter because he doesn't act on that reason.

Second, perhaps the hearer should reject the testimony. If *H* has reason to think that *S* is unreliable, then even if *S* is telling the truth on this occasion, he should reject what she says. Since he should reject what she says, even if he does the wrong thing and believes it, he cannot get knowledge that way. So *H*'s reasons to believe that *S* is unreliable block the transmission of knowledge from *S* to *H*.

Note that this is different to the reductionist requirement that *H* have independent reason to think that *S* is reliable. The two requirements come apart in the case where *H* has no reason one way or the other to form a judgment about *S*'s reliability. The anti-reductionist position is that *H* may take a default stance of trust towards *S*, and this is only shaken by positive reason for doubt. The reductionist position is that *H* should not, by default, trust *S*, and that this trust can only be properly grounded in evidence of reliability.

Finally, as noted above, some anti-reductionists think that transfer only happens between speakers and intended recipients of the transfer. So there is no transfer of knowledge between a speaker and, for example, an eavesdropper. The eavesdropper may come to get knowledge by what they do, but the explanation for that will be the reductionist one. So there is an important difference, on these views, beteween tellings and eavesdroppings.

We could imagine an even tighter restriction on transfer principles, so that they can only take place in the context of an established relationship. Most non-reductionists would not want to put such a restriction on their theory, because it would undermine the motivations for it. (In particular, it would complicate the story about testimony from strangers, and testimony received by infants, and as we'll note later on, those are two of the big objections to reductionism.) So we'll set such possibilities aside, simply noting that there are very interesting, and very hard, questions about the relationship between friendship and testimony. Recent work by Sarah Stroud (2006) and by Simon Keller (2004) provide worthwhile starting points for investigating these issues.

18.3 First Objection to Transfer: Creationist Teacher

Much discussion of anti-reductionism in recent years has focussed on Jennifer Lackey's examples that purport to show that the transfer model is false. We'll start with her example showing that you can have testimonial knowledge without transfer.

> Clarissa is a devoutly Christian fourth-grade teacher whose faith includes a firm belief in the truth of creationism and an equally firm belief in the falsity of evolutionary theory. Nevertheless, Clarissa recognizes that there is an overwhelming amount of scientific evidence against both of these beliefs. Indeed, she readily admits that she is not basing her own commitment to creationism on evidence at all but, rather, on the personal faith that she has in an all-powerful Creator. Because of this, Clarissa does not think that she should impose her religious convictions on her fourth-grade students. Instead, she regards her duty as a teacher to include presenting material that is best supported by the available evidence, which

> clearly includes the truth of evolutionary theory. As a result, while presenting her biology lesson today, she asserts to her students "Modern-day *Homo sapiens* evolved from *Homo erectus*." [Call this proposition *p*]. Although Clarissa neither believes nor knows this proposition, her students form the corresponding true belief on the basis of her reliable testimony. (Lackey, 2006, 434–5)

The thought here is that knowledge that *p* can't be transferred from Clarissa to her students because Clarissa doesn't even have the knowledge. She doesn't even believe *p*, so she can't know it. Yet her students do come to know it, and come to know it via testimony.

There are a number of replies that the anti-reductionist can make to this case, and which indeed have been made. Let's start with the question of whether this is really testimony. We usually think of testimony involving an assertion by *S*, that is directed at *H*. If *S* is just rehearsing lines from a play, and says "Something is rotten in the state of Denmark", it doesn't seem like a case of testimony, even if *H* misunderstands the kind of speech act that *S* is performing, and comes to believe that something is indeed rotten in the state of Denmark. Now there is something similar between the performance of an actor on a stage and a teacher teaching from a mandatory curriculum; they are both performing off a script. Don't stress too much about the fact that Clarissa isn't given a literal script; an improv actor isn't making assertions any more than an actor with a script is. The general pattern that people whose speech acts are directed by others are not making assertions, in the sense relevant to the theory of testimony, may make us doubt that this is really testimonial knowledge that the students get.

If it is testimonial knowledge, it isn't obvious that it is testimonial knowledge from Clarissa herself. There are many ways that *S* can testify to *H*. She can speak to him. She can send him a written communication. And, mixing the two, she can write something down, and pay a third party to read it to him. This is testimony all right, but it is testimony from *S*, not from the herald who reads it out. Perhaps the students are receiving testimony from the state board of education, who does know *p*, and not from Clarissa, who is merely their mouthpiece.

Finally, it is worth attending to a subtlety in Lackey's example. It is stressed that Clarissa believes that *p* is false. Lackey concludes from that that Clarissa does not know that *p*. How is the argument meant to go? In the quote we saw, the most obvious way of filling in the gap is to argue that if Clarissa believes *p* is false, she doesn't also believe it is true, so doesn't know it. But this isn't clearly right; perhaps Clarissa both believes *p* is true, and believes it is false. That would make her incoherent, but she clearly is incoherent to some extent. By stipulation she believes *p* is best supported by the evidence, and is false, which is pretty incoherent. So maybe she does believe *p*. Now perhaps although she believes it, and for good reasons, the fact that she also believes it is false undermines her claim to knowledge. I'm personally sympathetic to that line of reasoning, but it relies on a very strong principle, roughly that incoherence undermines all that it touches. So it isn't totally out of bounds to say that the anti-reductionist could argue that Clarissa really does know *p*.

Finally, note Lackey's own point that this is just an argument against the claim that all tetsimonial knowledge comes from transmission. As we've already seen, not even all anti-reductionists endorse this; some say that testimonial knowledge of eavesdroppers works the way that reductionists say it does. To complete the case against reductionism, Lackey needs another case.

18.4 Second Objection to Transfer Model: Whale Spotting

> While drinking a latte at Starbucks yesterday,Larry ran into his childhood friend, Mary, and she told him that she had seen an orca whale while boating earlier that day. [Call this q.] Having acquired very good reasons for trusting Mary over the fifteen years he has known her, Larry readily accepted her testimony. It turns out that Mary did in fact see an orca whale on the boat trip in question, that she is very reliable with respect to her epistemic practices, both in general and in this particular instance, that she is generally a very reliable testifier, and that Larry has no reason to doubt the proffered testimony. However, in order to promote a whale watching business she is in the process of starting, she would have reported to Larry – in precisely the same manner – that she had seen an orca whale even if she hadn't. (Of course, she wouldn't have believed that she had seen an orca whale if she hadn't.) Moreover, given the pattern of the whales' travel combined with the particular time of year it is, it is in fact quite surprising that Mary saw an orca whale when and where she did. (Lackey, 2006, 436–7)

This case is designed to show that testimony can fail to yield knowledge even when all the conditions the anti-reductionist usually likes are satisfied. In this case we have all these conditions met.

1. Mary knows that q.
2. Mary tells Larry that q, and Larry knows that Mary is telling her that q.
3. Larry and Mary have a long-standing relationship, that suffices to provide a proper ground for mutual trust.
4. Larry has no reason to believe that Mary is being dishonest.

Yet, says Lackey, Larry does not come to know that q. And that's because his belief that q is neither safe nor sensitive. The fact that Mary would have told him that q even in situations where it was false, and that it could easily have been false, undermines his claim to knowledge.

I don't have any clear intuitions about this case, to be honest. It is, in effect, a fake barn case. Larry is relying on a usually reliable process, in this case trusting a trusted friend, and getting the right answer, but the process goes wrong in some very similar possibilities. Does that mean it is knowledge or not? I don't know, which means I don't know whether it is a problem for the anti-reductionist.

Chapter 19

Reductionism, Children and Strangers

Last time we looked at puzzles for the anti-reductionist approach that says that testimony is basically an act of transfer. Today we'll look at potential problems for the reductionist alternative. The core worry is that we simply don't have the kind of knowledge that reductionists think that we need in order to learn by testimony.

19.1 Global and Local Justifications

Let's start with a distinction Lackey (2006) makes between global and local justifications. We could imagine that the way people justiy their practices of learning from others is that they first reason that testimony is as a general rule reliable, and then infer that the particular person speaking to them is reliable. Alternatively, it could be that they primarily get information about the particular reliability of the person speaking (who we're calling *S*), and that's all they need.

Most reductionists will go for the local version, because there are some tricky problems facing the global version. But we might note that the same choice confronts the anti-reductionist. The anti-reductionist denies that the hearer (who we're calling *H*) needs a justification for believing that the person speaking to them is reliable. They merely need to lack reasons to doubt *S*'s reliability. But the theorist who says that testimony is a basic source of justification needs some reason to defend this claim. What could it be? It often is a defence of the global reliability of testimony. Perhaps it is in the nature of speech that most speech must be true. An interpretation of our language that made it mostly come out false would, arguably, be a misinterpretation. Or perhaps it is a presupposition of a functioning community that people mostly speak truths. So the global approach is not without advocates.

But it does seem like a hard route for the reductionist to take. Even if there are subtle philosophical arguments that testimony as a practice must be generally reliable (and note that Lackey brings up doubts against this), these arguments are not known by, or even available to, most agents. It doesn't seem that the ability for typical humans to learn things through testimony relies on their ability to reason through these complex arguments; testimony is too central to human learning for that to be plausible.

So the reductionist should say that agents typically have a local justification. But this runs into some problems as well, as we'll now see.

19.2 Strangers

Consider one familiar everyday case of testimony. *H* is lost in an unfamiliar town, and asks *S*, who is passing by, for directions. *S* tells *H* where he needs to get to, and *H* comes to know the location, and direction, of his intended destination. This little, very common, interaction suggests an argument against reductionism.

1. *H* can come to know which way to go from *S*'s testimony.
2. *H* has insufficient background reason to believe that *S* is reliable.
3. If reductionism is true, and *H* has insufficient background reason to believe that *S* is reliable, then *H* cannot come to know which way to go from *S*'s testimony.
4. So reductionism is false.

In terms of the previous section, the worry here is that the only way 2 could be false is if there is a global justification for believing testimony. But even reductionists typically deny that; they mostly plump for local justifications. Yet by definition we have no local justification for believing a stranger. So *H* has no justification for believing *S*, as 2 says.

Still, there are a number of things reductionists can say, and have said, about this case. There is a very nice treatment of the case in the survey article by Adler (2015), and I'll follow largely what he says. The general trick is to see how there are grounds for a local justification of *S*'s reliability, even though *S* is a stranger. In particular, the argument is that *H* can rely on three kinds of considerations particular to the case to justify confidence in this particular thing *S* says, even if *H* lacks a global justification for believing testimony.[1] These considerations are:

- The kind of question that *H* asks;
- The manner of *S*'s answer;
- The content of *S*'s answer

The example, of *H* asking stranger *S* for directions, is a commonplace example in the literature on testimony. And it is a commonplace in everyday life. But note that it is, in ordinary life, a fairly special case. Directions are, to a first approximation, the only thing I ask strangers about, or that I'm asked about by strangers. The reductionist has to defend the claim that *H* can get knowledge from *S* in this case. They do not have to defend the claim that *H* can get knowledge from reports about arbitrary topics. It isn't part of reductionism that *H* can come to know where the best place to invest his

[1] Our focus here on whether *H* can provide a background justification for believing *S*. We will look in subsequent notes at the empirical evidence that typical *H* in fact does do this. (Sperber et al., 2010)

money is from asking stranger *S*. Indeed, it is implausible that *H* can come to know this. (Why should this be, according to anti-reductionism?) For that matter, it isn't part of reductionism that *H* can get knowledge from arbitrary reports of stranger *S* about directions: the intuition is that *H* can come to know the true answer to his own question. If *S* simply walks up to *H* on the street and says "There is a Starbucks around the corner" for no apparent reason, then it isn't clear *H* can know this. So let's just notice one class of facts about *S*'s utterance: it is an answer to a question that *H* asked about directions. As long as *H* is sensible, he can restrict the questions he asks. And we don't have to defend the global reliability of testimony to defend the general reliability of answers to sensible questions.

In a normal human interaction, there are all sorts of cues that go along with *S*'s utterance. She speaks in a certain accent, that may or may not seem typical for the area. In some places (college towns, high turnover areas like Manhattan) that won't be much of a signal of reliability, but in some places it will be. It will be possible to tell from *S*'s reply whether she is drunk, or not paying attention, or a tree, or something else that will lower confidence in *S*. Further, she will project a certain level of confidence. At least to my ears, the most trustworthy informants are those with a high but not maximal such level. Someone who is hestitating so much they don't trust their own answers need not be trusted. And someone who answers what strike you as hard questions without a moment's hestiation, and with an eerie over-confidence, probably shouldn't be trusted either. In the very abstract example we started with, we just said that *S* told *H* the directions. In reality, how she tells him the directions matters too.

Finally, there is the content of *S*'s answer. The fact that *H* accepted *S*'s answer doesn't mean that he would accept any answer. Adler gives the example of someone who asks for directions to the nearest gas station, and is told it is 300 miles away. They wouldn't believe that. And the fact they wouldn't believe that means they are sensitive to the content of *S*'s utterance; they don't just believe anything *S* says. That is, they are distinguishing between whether *S* is in general reliable, and whether this particular utterance of *S* is believable. Or, to put the point in more reductionist-friendly terms, they are distinguishing between whether *S* is reliable full stop, and whether *S* is reliable when she says plausible-sounding things, about a question *H* asked, which was a sensible question, and which was answered in a manner typically correlated with accurate answers. It doesn't seem implausible that *H* could know that generally the latter kind of answer is true, even if he knows nothing about *S*.

19.3 Children

As well as stranger testimony, the other kind of case that has long been thought to pose problems for reductionists concerns children. One of the central arguments in the important book defending anti-reductionism by C. A. J. Coady (1992) was that a reductionist theory of testimony could not explain how children can learn as much as they actually do by testimony. There are, at least, three different kinds of puzzles for the reductionist posed by children. They are:

1. Children do not have enough evidence to support either a global or a local justification of testimony.
2. Children do not have the intellectual capacity to process the evidence in such a way that it would support a belief in testifier reliability.
3. Children do not have a sufficiently sophisticated theory of mind that is needed to form beliefs about the reliability of others minds, in a way that most reductionist theories of testimony require.

We'll address the first two of these here, and the third next time.

19.4 Children - Poverty of Stimulus

The first worry is that children simply lack sufficient data to judge that people around them are reliable. Following Chomsky (1980) we might call this a "poverty of the stimulus" objection. Chomsky's concern was to argue that certain syntactic rules, in particular rules about the non-well-formedness of certain expressions, must be grounded in innately known facts about language, since children didn't get enough information to learn these facts. Children learn, by observation, which expressions are well-formed, but they don't learn which expressions are not well-formed.

We could try to argue in a similar way that children simply don't have enough data to decide who is reliable and who is not reliable. Children are told all these things about the wide world around them, but how could they possibly tell whether any of them are true without having independent access to the wider world?

On closer reflection, this doesn't look like a particularly strong argument. The first thing to note is that we have to specify which children we're talking about. It isn't obvious that very young children (say, under 3 months) do learn a lot by testimony. By the time a child is old enough that intuitively they are learning by testimony, they have heard a lot of words. According to one famous study, children hear between 600 and 2100 words per hour (Hart and Risley, 1995). Most of the uptake of that study has been about the variation between the high and low end of the scale. But across the scale, there are a lot of words there for children to use as inputs to a theory of reliability.

But how can they tell whether the words are true? Well, some of the words are about their immediate vicinity, and they can check those by perception. And some of the words, both across informants and across times, concern a common subject matter, so they can check those at least for coherence, and judge that someone who disagrees frequently with others is not in fact reliable.

So there seems to be sufficient information around for children to pick up reilability judgments. But can children use that information?

19.5 Children - Processing Power

That's the point where a lot of people wanted to reject the reductionist theory. Here's how Jennifer Lackey (2005) put the objection.

1. According to reductionism, a hearer, H, is testimonially justified in believing that p on the basis of S's report that p only if H has non-testimonially based positive reasons for accepting S's report that p.
2. However, infants and young children lack the cognitive capacity for acquiring and possessing non-testimonially based positive reasons.
3. Therefore, infants and young children are incapable of satisfying the reductionist's requirement.
4. Infants and young children do have testimonial justification for at least some of their beliefs.
5. Therefore, reductionism is false. (Lackey, 2005, 166, lightly edited for consistency with our terminology)

Lackey's own response to this is a kind of *tu quoque*; she says that if there is a problem here, it is also a problem for any reasonable anti-reductionist theory. And that may be right. But I think we can do more to reply directly. Recent evidence suggests that children are much better at tracking complex data than we might have thought.

Alison Gopnik, along with many colleagues over the years, has done a lot of work looking at how children infer to the existence of a causal relationship. Naturally enough, children rely on data about correlations in order to infer to the existence of a causal relationship. But it turns out that we can say much more about this. Gopnik developed a case where two possible causes, C1 and C2, were equally well correlated with an effect E in the data set that the child would be shown. But while C1 was an unreliable cause of E, C2 was not a cause. And this could only be detected if the child was tracking not just how often C was followed by E, but what else was happening in the cases where C was followed by E. In the cases Gopnik developed, there was a third cause C3, and C2 was never followed by E unless C3 was happening as well, while C1 was sometimes followed by E without C3. So the children are not just tracking correlations, as between C1/C2 and E, but conditional correlations, as between C1/C2 and E conditional on C3 obtaining or not. Importantly for our purpose, she found that children as young as two were doing this. Now the two-year-olds certainly would not be able to describe what they were doing in terms of conditional correlations that they were tracking. But the best explanation of what they were doing was that they tracked these conditional correlations. And that in turn suggests they do have the intellectual capacity to track the reliability of informants. (For more information on these experiments, see Gopnik et al. (2001) and Gopnik (2009, Ch. 3).)

Jenny Saffran and colleagues investigated a different problem facing young children: how do they separate the stream of speech they hear into words. As you may be able to tell by casual observation, spoken speech features nothing like the spaces between words we use in writing to break up words. If you are not a speaker of, say, Hungarian, it is hard to even tell where one word in spoken Hungarian ends and the next starts, even though this is easy to do with written Hungarian. How do children even start this process?

Saffran proposed that they use the following ingenious trick. Word boundaries feature statistically improbable sequences of phonemes. Look at that last sentence again: note in particular the part that you would pronounce 'turestat'. That isn't a particularly natural or common sequence. I don't think it is part of any English word. (Note that in English it's 'thermostat', not 'temperaturestat', which would have the sequence I'm describing.) So here's an idea for how to solve the problem: record a very long sequences of phonemes in speech, and hypothesise that there is a word boundary between any two that don't normally appear one after the other.

Saffran's team produced experimental evidence that that's now adult subjects were able to learn about word segmentation an artificial language (Saffran et al., 1996b). It's harder to prove that infants use just this method, but it would be plausible if it is how adults learn about word segmentation, and infants have the capacity to track this information. And in two follow-up studies (Saffran et al., 1996a; Aslin et al., 1998), they argued that the latter condition is met. In particular, 8 month old infants who have been exposed to a stream of syllables from a made-up language would listen longer to 'words' from the initial setup than they would to non-words. This 'listening longer' is usually taken to be evidence of surprise; they had expectations that were violated when they got unusual streams of syllables. And note here that unusual did not necessarily mean new; some of the non-words that were used in the test condition consisted of the last syllable of a familar word, followed by the first two syllables of a distinct familiar word.

The best interpretation of the data was that the infants, who were only 8 months old, were tracking the frequency not just of the syllables they heard, but of the relative frequency with which particular syllables appeared after other syllables. And this information was being fed into their decisions to attend more or less closely to novel stimuli.

It seems to me that this is a much more complex problem than the problem of tracking whether a particular speaker is speaking truthfully or not. If infants as young as 8 months old can track syllable frequency with such care and detail as is needed to compute word segmentations, they should be able to track reliability of informants. That is, they should be able to do this unless there are some specific reasons to think that they have a particular problem with judging reliability. We'll start next time with one such reason.

Chapter 20

False Beliefs and False Reports

20.1 False Beliefs

For a long time, roughly the late 1980s to the mid 2000s, the orthodox opinion in psychology and philosophy was that it took a long time for children to become aware of the possibility that people could have false beliefs.[1] Recently a flurry of evidence has turned up suggesting that orthodox opinion was totally wrong. And even more recently, some doubts have begun to appear about the 'flurry of evidence', suggesting there was more to the orthodox story than was apparent.

Whether young children can understand that other people have false beliefs is important in its own right. But it is also relevant to thinking about testimony. If children can't even comprehend the possibility of others having false beliefs, then they can't plausibly be thought to track the reliability of the beliefs of others. By hypothesis, they think the reliability of the beliefs of every person they meet is one. Someone so confused about the nature of mind does not have any kind of reliable beliefs that others are reliable, hence does not know that others are reliable, hence can't acquire testimonial knowledge in the way that reductionists posit.

When I talk about 'young children' here, I mean children up to about 42 months. That's actually quite far along developmentally. A 39 month old child can do some sophisticated cognitive tasks. We saw evidence in the last chapter about what they can implicitly do. But by that age they are capable of doing some rather impressive tasks explicitly to. Yet there are some tasks that they reliably fail at.

The first of these tasks has become so iconic that it sometimes just is referred to as the false belief task. It involves the child (again, picture this being a child who is 36–42 months old) watching a short play. On stage, there are two boxes, easily distinguished. I'll call them the blue box and the red box. A doll comes onto the stage, and puts something into one of the boxes, let's say the blue box. The doll then clearly leaves

[1] Note that I'm not going to cite every experiment performed here one by one; though that would be required in anything written for publication. And, indeed, you should do it in papers you are turning in! The relevant citations can all be found, however, in the bibliographies of recent surveys by Peter Carruthers (2013) and Cecilia Heyes (2014).

the stage. A different doll comes in, takes the object from the blue box, and puts it into the red box. In different versions of the play, different clues are given to make it clear that this is malicious; the second doll is trying to hide the object. Then the first doll returns. And the child is asked, where will that doll look for the object? And a very large percentage, usually a majority, will say that she'll look in the red box. The orthodox interpretation of this is that the child themselves knows that the object is in the red box, and can't imagine that the doll has the false belief that it is in the blue box.

This study has been the subject of massive variation and replication, and the result I've just reported is extremely resilient. There are, as we'll see, many challenges to the interpretation of the result. But the result that the child will say "Red Box" is well supported.

The disposition to say this turns off reasonably suddenly, some time late in the fourth year, and does not seem to correlate with any other developmental milestones. Indeed, it happens among children with Down's Syndrome, who are often (at least considered as a group) much later in reaching developmental milestones. The prominent group of children who are still getting this wrong around age 4.5 are autistic children, who indeed can keep getting it wrong much much later, and this fact has played a major role in theorising about autism. But we're not focussing on that here; what we care about are the 39 month olds who collectively get it wrong.

There is another experiment that seems to have a similar conclusion. This involves asking the child about their own beliefs at a different time, rather than about the beliefs of others. The experimenter shows the child a familiar package, say the package of a popular brand of candy. They ask the child what's in the package. The child says "Candy," as is reasonable. They then open the package up, and show the child that it has an unexpected filling, say it is full of pencils, not candy. They then ask the child "What did you think was in the package?". And the child will, typically, say pencils, not candy. The videos of this, where the children just contradict their own reports from seconds earlier, are really quite striking.

As I said, the standard conclusion drawn from both of these experiments is that children under 3.5 cannot understand the possibility that someone, either a different person or a different stage of themselves, has a false belief. But notably, both pieces of evidence for that come from the child's failure to complete what seems to us like a simple verbal task. In recent years, thanks to technological advances, we've been able to get evidence from children's non-verbal behavior. And that evidence suggests that children do understand what's going on.

There are two kinds of evidence that have recently been put forward. One involves surprise reactions, primarily looking time, and the other involves anticipatory looking.

It is a reasonably well established theory that children, like adults, will look longer at things that they find surprising. So here's a variant of the original false belief task.

We let the first doll go to one or other box, and pick up the object. And then we watch the infant to see how long they look at the play - i.e., at how surprised they are. I say 'infant' here because we don't need to wait until the subject is old enough to speak to run this experiment; we can do it with children as young as 15 months. And it turns out, on average, that they look longer at the situation where the doll goes to where the object was moved to. That is, we think, they find that more surprising.

There is another feature that children share with adults. Their eyes will look ahead of the action. In general, a person looking at an object will look somewhere between where the object is, and where they expect them to be. Like a camera operator, we're constantly framing our view so that the subject at the center of attention is moving into the center of the picture, not out of it. Now we don't even have to run the play through to completion. Just have the first doll come back into the room, make it clear that they are going to retrieve the object, and see which way the child starts looking. Again, from a very young age, they start to look towards the blue box; towards where the object was originally left.

So what to make of all this? The short version of the data is that children between 15 and 39 months (which is such an enormous range) have non-verbal behavior that is as if their theory of mind is quite reasonable, and verbal behavior as if their theory of mind is completely crazy, and doesn't include the possibility that people can have false beliefs. I have a bias towards theories that say people are reasonable, so I like the interpretation of the data that says young children do understand that false beliefs are possible, but for some reason do a bad job expressing this understanding in words. But it is hardly the only impression.

20.2 Children's Attitude To Testimony

So far we've been looking at what children could in principle do with the information they are presented about the reliability of those around them. It is worth noting that there are a number of studies already performed that address this directly, rather than having to take roundabout approaches. The studies are typically on small, idealised settings, and I think it is worth looking at the broader class of studies to see how we can generalise from the simple ones. But the simple ones are useful. In much of this section, I'll be reporting results that are discussed in Harris and Corriveau (2011).

One canonical study is reported in Koenig et al. (2004). Children are shown two informants, who provide conflicting information on five questions. On the first four, it is clear that the first is right, and the second is wrong. The child (who is always 3 or older in these experiments) is asked which of them they believe on the fifth. And they overwhelmingly choose the first. This is an extreme case, but it shows that children are at least minimally tracking reliability over time.

More importantly, this study has been replicated in ways that make it more plausible that it is detecting a real ability to track informants for reliability. If you make the first informant right three times out of four, and the second informant right on just the fourth, the child still trusts the first informant. If you show the fifth conflict a week later, and ask which of the informants is trusted, the child trusts the first informant. That shows that the child is storing this information about reliability, which is crucial for any real world discrimination between possible sources.

If you have the first informant be a stranger, and the second informant be someone the child has a prior relationship with, then three year olds will get confused (should I trust the familiar person, or the one who looks reliable), but four year olds will trust the reliable one. So children can be quite sensitive to new information about reliability, though note this skill kicks in much later than other skills we've discussed.

There are a number of other results that seem to support a broadly reductionist treatment of children's acquisition of testimony. That is, they support the idea that children are actively monitoring the world for cues about the reliability of informants, and selectively trusting those who fit the description.

Children do not accept just what they are told, even from previously reliable informants. The paradigm used above - show child a reliable and an unreliable informant, and watch them trust the reliable one - breaks down if the subsequent information is implausible. Jaswal et al. (2008) ran a version of this experiment where the test question concerned the proper formation of plural and past-tense forms of made up words. For example, the 'reliable' speaker would say that the plural of 'lun' is 'lean', while the 'unreliable' speaker would say it is 'luns'. Despite the fact that English does actually have a lot of irregular plural formations, and despite the reliability of the informant, the children believed that the right answer is 'luns'. That's what a reductionist should say; testimonial evidence is evidence, but it can be overridden by background knowledge. (Remember a key document in the history of reductionism is Hume's scepticism about testimony about miracles.)

Children don't just track the reliability of informants, they track something like their knowledgability. Einav and Robinson (2011) had the pair of informants be equally reliable, but the first informant would say the true answer unaided, while the second informant always needed help even with simple questions, but always said the right thing. The children were then asked which informant they wanted to ask a new question to, and they said the first. So they aren't just tracking reliability, but knowledge. This suggests a richer form of reductionism; we look for informants who know what they are talking about, not just who have true beliefs.

And even two year olds can do things reductionists expect. Birch et al. (2010) found that two year olds are sensitive to the levels of confidence with which people speak, and are more likely to trust more confident speakers. This is, you may recall, one of the strategies we suggested the reductionist could use in puzzles about learning from strangers; the form of the report could be a piece of evidence. And it is, it seems, evidence that we clue onto very quickly.

20.3 Hearing and Believing

There is one last important piece of evidence that certainly looks like it poses a problem for reductionism. Daniel Gilbert and colleagues argued, based on work done in the early 1990s, that people did not process information they heard before believing it. Rather, they simply believe it. Now that's not to say that people end up believing everything they hear; that would be absurd. The view, instead, is that people initially believe stuff they hear, then if it is absurd, go through the process of disbelieving it. But this is work, and won't get done if the person is under enough stress.[2]

If that's right, it is a bit of a problem for reductionism. It isn't an argument that people don't, in general, have the capacity to judge whether their informants are reliable. It's rather than this capacity is not, apparently, exercised in the usual process of coming to believe things. And that is a problem for at least the most attractive forms of reductionism: those that say we should and do go through some kind of inferential process before believing things.

Here's the basic experiment Gilbert and colleagues performed. Show the subject some sentences on a computer screen in succession. Some of them will be black, others red. The subject is told in advance that red is a form of negation; showing something in red means it is false. Then put the subject under some stress. For example, make the sentences go by very quickly, and make the subject do some other mental activity (e.g., count backwards by 7s) that takes up most of their processing power. At the end, ask the subject some basic general knowledge questions. They will, on average, answer as if they had been told the sentences in red as truths.

What should we make of this? Sperber et al. (2010) say that we should not read too much into it. In particular, they note, the experiment relies on the claims in red not being of any importance to the subject. Change that fact about the experiment, and it does not replicate. So perhaps we should have a different interpretation of the basic result.

Here is the one that they propose. Consider what happens when you are walking down a busy street. Consider, in particular, your attitude towards the people around you. In a typical case you will be, to use their helpful word, vigilent. You certainly won't be making conscious predictions about the behavior of every other person on the sidewalk, to think about which way they might move. If you're sufficiently distracted, you won't make any predictions about them at all. But you'll be disposed to start paying serious attention to any one of them if they start acting out of the ordinary (and you

[2] For more details on this work, see Gilbert (1991); Gilbert et al. (1990, 1993).

aren't otherwise distracted). That suggests that you are, at some level, monitoring the situation. You are keeping watch on whether there is anything that is worth spending conscious attention on. In normal cases, there isn't, so you just keep on walking. But the background processes keep chugging along. Walking is, hopefully, different to driving in this respect; a driver should be paying conscious attention to lots of things around them, while the walker can get by with unconscious attention.

Sperber et al. (2010) present a lot of evidence, which I won't try to summarise here, to the effect that that's how we think about informants. We are constantly monitoring the situation for signs of unreliability. If those signs appear, then we switch to conscious attention to reliability. If we're distracted, or not really invested in the process, we might miss some signs. (Even flashing red signs!) We might do the monitoring poorly; subconscious systems can be even more biased on grounds of race, sex, gender, class, etc than conscious ones. But we shouldn't, they argue, think that testimony becomes belief straight away. Rather, as the reductionists argue, it requires some background processes to tacitly approve it. If they are right, there is a good empirical argument for a kind of reductionism about testimony after all.

Chapter 21

Knowing about Knowing

In the last chapter of her book, Nagel (2014, Ch. 8) focusses on four related issues:

1. How is it that humans can tell what's going on inside the minds of others?
2. Are our judgments about knowledge, which we use all the time in philosophy, a sufficiently reliable enough foundation for any reasonable theory?
3. Are our judgments about knowledge prior to, or subsequent to, our judgments about belief?
4. Is knowledge itself analytically prior to, or posterior to, belief?

Here is one set of answers to those questions - roughly the answers that Nagel promotes in the chapter.

1. We make inferences from the behavior and appearance of agents to claims about what is going on inside their heads. These inferences are highly reliable. And that isn't just because we're smart and have lots of practice at this; we probably have a special part of our brain that is dedicated to the task of mindreading.
2. Yes - the fact that we're so good at mindreading in general, combined with the fact that knowledge is a mental state, gives us reason to think that we are reliable. Moreover, attempts to show that there is enough disagreement between judgments to undermine that reliability have failed.
3. Yes. The developmental evidence suggests that the ability to detect knowledge comes earlier than the ability to detect the broader class of beliefs. And probably this reflects something in the architecture of adult minds as well; detecting knowledge is a simpler task than detecting belief.
4. Yes. We should not think of knowledge as belief + X, as most theories of knowledge historically have. Rather, we should think of knowledge as the more fundamental category, and think of belief in terms of it. Perhaps belief is something like "attempted knowledge", in the sense of "attempt" where it is compatible with success.

These answers are obviously not compulsory - it wouldn't be philosophy if there was only one possible answer to a question - but nor is it compulsory to answer them all in the same way. In particular, the last two questions could, in principle, come apart. Consider the relation between the ideas of *being a bird* and *looking like a bird*. In terms of the metaphysics, which is what question 4 is asking about, *being a bird* is prior. The nature of birds explains what it is to look like a bird, and not vice versa. But in terms of cognition, it could be the other way around. It could be that in practice the way that we classify things as birds is that we first notice that they look like birds, and then infer that they really are birds. I'm not saying this is how the recognition of birds goes, but it isn't a completely implausible theory.

But there is some unity to Nagel's answers. And we can see that by picturing a set of answers that go the other way on at least the last three questions. Here's the rival view, one that was long the dominant view among philosophers. Knowledge = Belief plus X, for some X that is not, intuitively, a mental property. If that's right, the following answers suggest themselves.

1. The imagined philosopher can agree with Nagel on this question.
2. No, they are not. Humans are typically good at mind-reading, but they are not so good at either identifying what X is (look, we haven't done it so far), or identifying instances of it.
3. No, it does not. We come to believe that people know something by seeing that they believe it, seeing the belief (or the agent) is X, and putting these facts together.
4. No, it does not. In fact, belief is a part of the concept of knowledge.

The imagined philosopher might say that the reason we have such trouble with fake barn cases and the like is that we don't know precisely what X is, and we can't identify cases of it well, so once we get outside familiar territory, we just lose the ability to think reliably about knowledge. But that doesn't matter for practical purposes; what matters is identifying beliefs and desires, and perhaps identifying true beliefs, and we can do that while being lousy at identifying knowledge.

On the other hand, if belief is attempted knowledge, then the picture the imagined philosopher is presenting becomes less plausible. It would be surprising if we were really good at identifying attempts at knowledge, but poor at identifying successful attempts. And it would even be surprising if we identified attempts at knowledge prior to identifying successes. It isn't impossible; we could imagine a creature C and an act X such that the only way that C could tell that X was performed was that she could see someone trying to X, and could see that the try was successful. (Remember, it isn't crazy to think that we come to believe that there is a bird nearby by first noticing that something looks like a bird, and tacitly inferring that it is as it looks.) But if belief is attempted knowledge, the answers that Nagel gives do seem to fit neatly into place.

21.1 Behaviorism

The answer Nagel gives to the first question is by now orthodox opinion, and rightly so, but it's worth spending a little time on the history of how we got here. Consider the following argument, which is a slight caricature of the reasoning you see in classic mid-century behaviorist writers like Gilbert Ryle (1949).

1. If mental states like belief were states of people's souls, we could not know them unless we could look into people's souls.
2. We can know what people believe.
3. We can't look into people's souls.
4. So mental states like belief are not states of souls.

This is an example of a **transcendental argument**; we reason from the existence of some knowledge, and some theories about the requirements or constraints on knowledge, to the conclusion that the requirements and constraints are satisfied. Here the existence of knowledge is premise 2, and the constraint is given by premise 3. (Premise 1 is almost definitional, and at least isn't controversial.) And it isn't a terrible argument; if minds are immaterial souls, then perhaps it should be a lot harder than it seems to be to read them.

Some behaviorists took this much further, however. Here's a much more radical version of the argument.

1. We can know what people believe.
2. All our knowledge of other people comes from how they behave.
3. So, beliefs must be (possibly complex) facts about behavior.

And here the argument seems to go wrong. In general in life, we don't just know what systems do. We make guesses about why they do it, and then we test them. A typical test will be to look at competing explanations of how something works, not that the competing explanations will make different predictions about the circumstances in which the system will fail, and then put the thing in some of those circumstances. By seeing when the system does and does not fail, despite hard testing, we can judge different theories about what is going on underneath.

And when we do this for human minds, we take ourselves to be learning something about what really happens inside minds. We aren't just describing behavior. Our scientific practice here, as in many other places, is to posit underlying structure to account for the visible phenomena. When I say 'our' here, I don't just mean scientists; this happens in everyday life. When you think "Oh, she must think that such-and-such" in order to explain some otherwise puzzling behavior, you're doing the same thing.

As Nagel notes, there is reason to think that humans are distinctively good at this kind of exercise. So there is reason to think we get knowledge from doing it; it's the proper exercise of a cognitive skill. Nagel thinks that skills comes in relatively late; after the ability to pass verbal false belief tasks. I think it probably comes a little earlier, because the non-verbal tasks are telling us that younger children can mind-read too. But the big picture is the same.

21.2 Hard Cases and Experiments

As Nagel notes, in the early 2000s there was a flurry of interest in surveying people from as many walks of life as possible to see what their reaction was to various thought experiments that we use in epistemology. And at first, it seemed like these surveys revealed systematic divergences in how the terms were indeed used. Notably, it didn't seem that English speakers from East Asian backgrounds thought that cases involving inference from a false belief (like the Dharamottara and Gettier cases we discussed earlier) were really cases of knowledge.

If that finding held up, there would be a few things we could conclude. One natural conclusion, I think, would be that the term 'know' didn't pick out anything very important. So imagine one class of speakers uses 'know' to pick out Belief Plus X, and another uses it to pick out Belief Plus Y, and the beliefs in Dharamottara-style cases are Y but not X. If that were all true it would be terrible to argue that X is really philosophically, or even theoretically, more important than Y because it is what we pick out by 'know'.

And perhaps we could draw more radical conclusions than that. If different people mean different things by 'know', perhaps we should stop caring about knowledge in epistemology, and just care about belief, X and Y (and related properties). And if we think, as seems reasonable, that our linguistic dispositions are liable to influence our judgment about the relative importance of X and Y, then perhaps we should be broadly sceptical about our ability to easily resolve disputes about the importance of X and Y.

Those are some questions we would face if the original results had held up. They have not. It should have been surprising if they did. It would be very surprising indeed if every language had a term for Belief Plus Something Important, but the Something Important changed from language to language. And it would be even more surprising if languages differed greatly on what the Something Important was, but the term for Belief Plus Something Important was one of the most common verbs in their language. Yet we know from the cross-linguistic studies that terms for knowledge are found universally, and are very widely used.

In fact the data from the original experiments seemed to be generated in part by different assumptions that speakers made about how to fill in the background to the thought experiments offered, and in part by different levels of knowledge on the part of subjects about the subject matter of the experiment. For an example of the latter, some of the presentations of the thought experiments presupposed that the person reading the presentation would know that, say, a Buick was an American car. And this knowledge is obviously not equally widespread amongst different groups. Once that was corrected for, and once enough gaps in the story were filled in to take away points where it was easy to intepret the story in multiple ways, it turned out that most people, across various demographic categories, agreed that cases like Dharamottara's were not cases of knowledge.

The main experimental finding that does seem quite robust is that once you remind subjects of arguments for scepticism, their willingness to attribute knowledge even in ordinary cases goes down precipitously. Some people have taken that to be an argument that 'know' is a context-sensitive term, like 'empty'.

Think about what it would take to describe a fridge as empty. Often you would use it to mean there is no edible food in the fridge. Other times, you might use it to mean that there is nothing but the shelving there. (Is a fridge that has no food, but needs cleaning, empty?) Perhaps when you are thinking of moving it, you need to get all the shelves and drawers out, so it is only empty when everything movable is removed. Perhaps if you are using it for a chemistry experiment, you need to make a vacuum inside, so you need to empty out the oxygen. In general, when you'll use 'empty' to describe something can vary greatly with the purposes you put the word to. That's odd, because you might think 'empty' had a very clear meaning: contains nothing. But in practice, what counts as 'nothing' can vary.

Well, perhaps the same is true of 'knows'. What is it to know something - it's for there to be no possibility of error. What does no possibility mean? Well, it's like there being nothing in the fridge; it depends on you're interested in. This is the contextualist view that Nagel discusses in Chapter 7 of her book. I'm not going to go into that view in any detail, save to note that there is a very active research program in looking into how well the view squares with our ordinary practice of making knowledge ascriptions. Like Nagel, I'm personally sceptical that it ultimately provides a good explanation.

21.3 Hard Cases and Knowledge First

Let's conclude by thinking about how well the 'knowledge first' position Nagel sketches does at explaining the difficulty that subjects have with a couple of hard cases. In particular, let's spend a bit of time on sceptical doubt cases, and fake barn cases. The kind of sceptical doubt case I have in mind is like the one discussed in the previous section. For concreteness, imagine the following exchange, noting that B's responses are representative of a very large class of subjects in real experiments.

> A: Bob is at the zoo with his daughter. He sees an animal in an enclosure, and is sure it is a zebra, based on its shape, the stripes down its sides, and

> the sign saying "ZEBRA" in front of the enclosure. And indeed it is a zebra; there is nothing misleading about the situation. Bob says to his daughter, "That's a zebra". Does Bob know that there is a zebra in the enclosure?
> B: Yes.
> A: Now add one fact to the story. If the zookeepers had very carefully painted a donkey to resemble a zebra, and put it in the zebra enclosure, Bob would have been fooled. He's not so good at recognising zebras that he couldn't be tricked. He wasn't tricked, and the zookeepers aren't interested in those kind of tricks, but they had the capacity to trick him.Does Bob know that there is a zebra in the enclosure?
> B: No.

In this case, B changes her mind just by being reminded there is a kind of sceptical possibility that can't be ruled out. But there are always sceptical possibilities that can't be ruled out. Is that compatible with the 'knowledge first' program?

Here's the argument that it isn't. This is an easy case of belief. Bob believes that there is a zebra. It's hard to say whether he knows it, and in fact we can be easily manipulated into saying that he does or he does not. The simplest explanation of that is that knowledge is belief plus X, and B is being manipulated to change her mind on whether the X condition is satisfied. That's why B doesn't lose any confidence in whether Bob believes there is a zebra, while her standards for what it takes for something to be knowledge seem to be easily changed; her view about X is very unstable.

Here's the argument that it is. Does B really believe throughout that Bob believes there is a zebra? Perhaps she does not. Perhaps she attributes the belief that there is a zebra to Bob when she only cares about whether there is a zebra rather than some other non-disguised animal. But once sceptical doubts are raised, it isn't just that Bob fails to know there is a zebra, but that he isn't (on the most natural understanding of the story) even committed to it being a zebra rather than a cleverly disguised donkey. If A manipulates B's standards for knowledge by bringing up possibilities like this, she similarly manipulated B's standards for belief. And that's just what you'd expect if the knowledge first picture is right.

I'm not sure which side has the best of that exchange, so let's look at another case. A tells B the story of Henry driving through Fake Barn Country, and asks B, does Henry know there is a barn in that field? B is unsure, like many of us are. But B is sure that Henry believes it is a barn. This I think is a bit puzzling on the view that belief is attempted knowledge. We have to say that B is certain that Henry is attempting to know something, but we can't say whether the attempt is successful. And that's despite having all the facts that should matter in front of us. This is a little odd. If we don't know whether Henry knows there's a barn, that seems to be because we lack knowledge about some important property of knowledge itself. It isn't because the story hasn't been filled in enough, for example. If we lack that knowledge about knowledge, you'd expect that we should similiarly lack knowledge about attempted knowledge. But that isn't as clear; it's hard to get a fake barn case (or something similar) where it isn't clear whether Henry even believes there is a barn.

That's not conclusive, to put it mildly, but it does make me a bit worried that it's hard to see how knowledge could be conceptually prior to belief. This is, I think, going to be an active area of debate in epistemology in upcoming years.

Chapter 22

Testimonial Injustice

22.1 Ethics and Epistemology

Epistemology is a normative discipline. That is, we talk a lot about what people should and should not do, and we talk a lot about how it is better and worse for people to adopt certain epistemic practices. But we haven't talked a lot so far about how it interacts with the other big normative discipline: *ethics*. Yet there are obvious connections between the two. Here is one famous example linking the two fields, from William Clifford's *The Ethics of Belief* (Clifford, 1876).

> A shipowner was about to send to sea an emigrant-ship. He knew that she was old, and not overwell built at the first; that she had seen many seas and climes, and often had needed repairs. Doubts had been suggested to him that possibly she was not seaworthy. These doubts preyed upon his mind, and made him unhappy; he thought that perhaps he ought to have her thoroughly overhauled and refitted, even though this should put him to great expense. Before the ship sailed, however, he succeeded in overcoming these melancholy reflections. He said to himself that she had gone safely through so many voyages and weathered so many storms that it was idle to suppose she would not come safely home from this trip also. He would put his trust in Providence, which could hardly fail to protect all these unhappy families that were leaving their fatherland to seek for better times elsewhere. He would dismiss from his mind all ungenerous suspicions about the honesty of builders and contractors. In such ways he acquired a sincere and comfortable conviction that his vessel was thoroughly safe and seaworthy; he watched her departure with a light heart, and benevolent wishes for the success of the exiles in their strange new home that was to be; and he got his insurance-money when she went down in mid-ocean and told no tales.
>
> What shall we say of him? Surely this, that he was verily guilty of the death of those men. It is admitted that he did sincerely believe in the soundness

> of his ship; but the sincerity of his conviction can in no wise help him, because he had no right to believe on such evidence as was before him. He had acquired his belief not by honestly earning it in patient investigation, but by stifling his doubts. And although in the end he may have felt so sure about it that he could not think otherwise, yet inasmuch as he had knowingly and willingly worked himself into that frame of mind, he must be held responsible for it. (Clifford, 1876, 289–90)

Here's the principle that Clifford wants you to draw from this case. Believing that you are not harming others is no defence against the moral charge that you have actually harmed them, if the belief was itself not well formed in the first place. If that's right, there is a very tight connection between ethics and epistemology; unjustified beliefs can produce immoral actions, even if the belief is sincerely held.

We might note that Clifford's case is rather over the top. The ship-owner doesn't merely form an unjustified belief, he uses a simply awful belief forming strategy to get to that belief. He has to, in Clifford's telling, actively suppress doubts. He is a paradigm of what we'd now call motivated thinking.

Perhaps this matters. Imagine changing the case in a few ways. The ship-owner no longer engages in such motivated reasoning. Indeed, he takes his motivations to align perfectly with the welfare of the passengers. (Perhaps the regulatory costs of having a poor ship are so great that it's in his best interest to keeping the ship running.) If he had the faintest doubt that the ship was unsafe, he would immediately do the repairs, since he is sure (perhaps we'll say correctly) that the cost of repairs will be less than the legal costs of being found to run a poor ship. Yet somehow, the ship-owner misses the tell-tale signs that the ship is about to fail, and fail it does.

In other words, imagine that the ship-owner is unjustified in believing the ship to be sea-worthy, but the ship-owner has not committed any procedural wrongs in coming to this belief. He's simply got it wrong; catastrophically wrong as it turns out. Question: Is he still 'verily guilty' for the deaths of the migrants? In my mind, this isn't clear; malice and stupidity are separate failings, and the general lesson Clifford wants to draw might run them too closely together here.

I don't want to push any particular theory based on this example. What I want to draw from it are two lessons. One is that there are hard and interesting philosophical questions at the intersection of ethics and epistemology, and in particular about the ethical consequences of unjustified beliefs. The other is that when answering these questions, we might want to look not just at how the agent currently is (what they believe, what evidence they have for that belief, etc), but how they got there.

22.2 Basic Case

Thanks to Miranda Fricker (2007), there has been considerable attention paid in the recent literature to a special class of cases where it might seem that an error in belief has moral consequences. We'll work up to the case Fricker is most interested in by first looking at a much broader class of cases.

We often have the wrong opinion about another person. And, oftentimes, those wrong opinions concern matters where it would be very good for the person we're thinking about if we had the right opinion of them, and very bad for them if we had the wrong opinion. It is easy to think of cases from your own lives where this is so; think of how bad it is for you when professors, potential employers or potential romantic partners, for example, treat you differently than they would if they had a better and more accurate impression of you.

One important instance of this is when the impression that a person, who we'll call H, has of the *credibility* of another person, who we'll call S, is lower than is warranted given S's actual credibility. Following Fricker, we'll say in this case that S suffers from a credibility deficit. This does not mean that S really is not credible; it means she is not taken to be credible.

There are many ways we can carve up the cases of credibility deficits. But note for now it is defined purely factively; there is a deficit as long as H's opinion of S's credibility is lower than would be warranted given how reliable a speaker S actually is. As Fricker notes, there may be the occasional case where this is helpful to S. But the standard case will be that this is not good news for S.

22.3 Harms and Wrongs

There are three importantly different kinds of ways in which it could be bad for S if H takes her to be less credible than she actually is.

First, this might **directly harm** S. What I mean by this is that it might be an interest of S that she be believed by H. That is, things would be better for S if she is believed by H, merely in virtue of being believed. If H doesn't believe her, then arguably H harms her. This might seem like a strange interest to have, but recall our discussion of strangers asking for directions. There is something bad about simply not being believed; for whatever reason, we care about whether we are taken to be trustworthy sources about various facts.

Still, the more likely outcome is that S will be **indirectly harmed** by not being believed. A patient may not intrinsically care about how credible the medical professionals take her to be, while caring quite a lot about being healthy, and being harmed quite substantially if the medical professionals do not believe her, and hence do not properly treat her. In the example from Harper Lee that plays a central role in Fricker's presentation of her theory, Tom Robinson is very badly harmed in virtue of not being believed. He is unjustly imprisoned, and killed in prison.

There is a distinction that's worth being careful about here between *harming* someone and *failing to benefit* them. There are all sorts of ways each of us could random confer benefits on the people around us, most of which we don't do most of the time. That doesn't mean we're harming people; harm means making someone worse off relative to an important baseline. The baseline need not be the actual state; a doctor who refuses to treat a person may harm that person without making them worse off, if it is reasonable to expect that the person would get medical treatment. So the distinction between harming and merely failing to benefit is not always easy to draw, but it does seem morally significant.

From the other direction, there is a distinction between **wronging** someone and merely harming them. Here are two ways in which we can harm someone without wronging them. First, the harm might be neither intentional nor even careless. If I'm talking to someone in a relatively secluded area, and gesticulating wildly to help make points, and then out of nowhere someone walks up and walks right into one of my hand gestures, I may well harm them, but unless I was being careless, I didn't wrong them. It was just an accident. Second, the harm might be justified. Punching someone to stop them killing a third person is clearly a way of harming the person you punch, but it doesn't wrong them, since you had a compelling reason to do what you did.

We're going to be interested in cases where the speaker S is **wronged** by being assigned an insufficient credibility. It shouldn't be thought obvious that there are such cases; before thinking about the examples, it is possible that S is only ever harmed, and not wronged, by being assigned insufficient credibility. And in fact Fricker thinks that cases where S is wronged are in fact relatively circumscribed.

22.4 Three Varieties

So consider three possible cases of credibility deficit.

The first is a case where H's belief about S, while false, is perfectly reasonable. Fricker mentions an example of this. On the basis of knowing that S works in a medical school, H takes them to not be particularly credible on the fine points of contemporary philosophical debates. This isn't an unreasonable generalisation; most doctors don't have the relevant philosophical background, or the time to keep up with the relevant philosophical literature, to contribute to these debates. But it is far from an exceptionless generalisation; it is possible that S is one of the exceptions. The particular exception Fricker mentions is a medical ethicist who happens to be employed in a medical school, but you can probably imagine others. Still, it isn't unreasonable for H to think that S's philosophical background will be roughly like that of most other people in medical schools, and this could lead to a credibility deficit.

That, says Fricker, is not an injustice. I'm assuming here that someone suffers an injustice, in the sense Fricker is interested in, only if they are wronged. The difference between the two concepts is largely in whether we focus first on S or H. In the case where H makes an innocent mistake about S's credibility, H does not wrong anyone (even if he harms both S and himself), and S suffers no injustice.

Fricker says the same is true in cases where the mistake is ethically innocent, but it isn't epistemologically innocent. So imagine, she says continuing the above case, that H actually had, or very easily could have had, good evidence of S's philosophical ability. But H isn't engaged in any kind of motivated reasoning, or familiar kinds of stereotyping. He just blunders, forming a false belief that he shouldn't have formed, given his evidence.

Fricker says that this too is not an injustice to S, and I think this means not a wrong by H. She's trying here, with good reason, to keep ethics and epistemology apart for as long as she can. A silly error is just silly, and not an injustice or a wrong. Or, at least, that's the position Fricker takes. It is worth spending some time thinking about whether you agree. Could someone commit a wrong, cause another to suffer an injustice, merely by errors in thought that are not themselves ethically wrong?

The case that Fricker does want to focus our attention on is when the source of the credibility deficit is something that seems unethical. And that's when the basis of the credibility deficit is the prejudice that H has towards the group that S is part of.

22.5 Varieties of Prejudice

There are a number of ways in which certain credibility deficits based in prejudice stand out from other kinds of credibility deficit. These 'ways' aren't perfectly correlated, but they are strongly correlated; enough so that they arguably form a natural kind.

One way in which they stand out is that they are **systematic** rather than **incidental**. A systematic credibility deficit is one that affects the victim of it across a variety of fields, not just in a particular area. For a long time, for all I know through to the present day, there was a very widespread view around the world that English cooking, and English food, was basically awful. When an English person was around a group who had that widespread view, she might well suffer a credibility deficit when she tried to talk about cooking. This is an injustice; you shouldn't be dismissed because of lazy prejudicial thinking even on something like cooking. But it wasn't systematic; an English background would usually give one if anything a credibility boost. The kind of prejudice that, say, Tom Robinson faces is not like that; it isn't restricted to a carefully defined field.

Another way in which they stand out is that they are **identity-based**. If H doesn't believe S because of S's ethnic background, or national background, or sex, or gender, or sexual orientation, that feels rather different to not trusting S because of how they dress or their manner of speaking. This doesn't mean that it's good to indulge in stereotypes relating credibility to clothes or vocal style, but it does seem a different kind of case to the identity-based cases.

Finally, the identity-based prejudices will usually be **persistent** rather than **transient**. That is, they will stick around. If I suffer a credibility deficit because I'm wearing a purple shirt, and people around here are biased against speakers in purple shirts, then things may go badly for me today. But they won't keep going badly; unless the bias is against people who have ever worn purple shirts. Tomorrow I'll be back to my old, if anything over-trusted, normal state. Obviously the kind of wrong I suffer is radically different to someone who is distrusted day-in day-out, across a variety of fields.

Chapter 23

Stereotypes and Injustice

23.1 Stereotypes

The core case of testimonial injustice, for Fricker, is when a speaker S is given less credibility than they deserve by hearer H, and hearer H's reason for giving S less credibility than they deserve is morally problematic.

In theory, this could happen for any reason at all. In practice, it often happens because H is relying on stereotypes. And, in particular, H is relying on stereotypes in a morally problematic way.

Stereotypes themselves are not necessarily problematic. Indeed, it is hard to see how we could get by without them. And there are uses of them that seem perfectly reasonable. If a person is extremely hesitant about reporting what happened, and refuses to make eye contact while talking, and keeps changing details of their story, then it's not obviously unreasonable to think they are being dishonest. And that's because they fit a quite reasonable stereotype; that of the dishonest person who feels guilty about their dishonesty, and is acting in the way that guilty people feel. As always, we have to be careful. Those mannerisms might also be caused by shyness, or by the trauma of the facts the person is reporting. But it isn't clearly unreasonable, or even unreliable, to make judgments about credibility on the basis of these kind of inferences from body language.

There are other cases that seem even less problematic. We could, in principle, have reason to believe that group G is particularly trustworthy, at least on some subject. Maybe the group is a professional group, and there is an extremely strong professional norm against deceit and dishonesty. In that case, learning that S is a member of group G, and is speaking on professional matters, might provide a reason to assign S a very high credibility. And that could be reasonable even if S is an outlier; someone who doesn't follow the standard norms of her professional group. We'd still be relying on a stereotype, the stereotype of G's as honest, but it is hard to see how this is problematic.

Some authors react to cases like these by terminological stipulation: 'stereotype', they say, is only to be used for negative attitudes. Fricker rejects this, and I think for a good reason. Some common associations between groups and attributes may be hard to classify as being positively or negatively valenced. Indeed, they may be attributes that we admire in some settings and not in others. It would be unhelpful to say that the association between the group and the attribute was a stereotype if we were in one setting, but not in another.

Fricker suggests that a stereotype is an association between a social group and one or more attributes. Let's note two points about this account.

First, it requires that the stereotype be an *association*, and not necessarily a *belief*. Believing that group G has the attibute is one way of associating the group with the attribute. But it isn't the only way. By defining the terms this way, Fricker has avoided getting drawn into debates about whether implicit biases are really beliefs. Whatever associations underlie implicit bias, they need not be beliefs to be stereotypes.

Second, it doesn't say how strong the association has to be. Presumably it does not have to be a universal. I can have stereotyped views while still recognising exemptions. On the other hand, it better be stronger than an existential. I can reject a negative stereotype about a group, while acknowledging that the stereotype accurately depicts 0.01% of the group's membership.

Fricker doesn't put it this way, but one natural way to understand the content of a stereotyped attitude is as a **generic**. Consider the sentence *Cows eat grass*. That doesn't say that all cows eat grass; that isn't true because of cows in feed-lots. And it says more than that some cows eat grass; the parallel sentence *Monkeys eat grass* is false, even if some odd monkey somewhere is a grass-eater. Rather, it says something like that the normal state of a normal, general cow is to be a grass eater. The details of exactly how these generic sentences, with no pronounced determiner, work is a matter of some dispute. But they seem to be at the heart of stereotypes.

23.2 Prejudice

As its etymology suggests, Fricker takes a prejudice to be a pre-judgment. To be prejudiced against group G, you have to have an association between G and some negative attribute, and to have that association in advance of the evidence. It isn't a prejudice to think poorly of Nazis; that's just a sensible response to the evidence. It would have been a prejudice to think poorly of them on the first sight of Hitler's moustache, with no other evidence of their attributes other than the facial hairstylings of their leader.

So prejudice is defined in part by its relation to the past; it is an association that comes in advance of the appropriate evidence. It is also defined by its relation to the future, or at least to the holder's dispositions towards the future. For Fricker says that a stereotype is an association that is resistant to counter-evidence. So even if the holder of the prejudice gets evidence that they are mistaken, they will hold onto it.

In practice, you might think these two temporal features - lack of evidence in the past, resistance to counter-evidence in the future - will run together. And they probably will. But it is worth thinking through the cases where they come apart.

Imagine first a person (similar to one Fricker describes) who has formed a belief that women are not very good at theoretical reasoning on the basis of very weak evidence, but is about to give that belief up without too much evidential 'pushing'. Right now, while they still have the belief, are they prejudiced? Maybe a little, I think, even without the resistance to counter-evidence.

On the other hand, imagine a person who has quite a lot of evidence for an association between a group and a negative attribute. The link is so firmly established for them, that they are disposed to dismiss counter-evidence. (Not that they have any yet.) This isn't unusual; humans tend to require more evidence than our best models suggest is necessary to be bumped out of a belief state. Is this person prejudiced in light of their resistance to revising their (well-grounded) negative attitude? Again, maybe a little, though probably less so than the previous case.

I've simplified Fricker's account of prejudice a little, and it's time to restore the complexity. Fricker says that a prejudice doesn't just involve a resistance to counter-evidence, but a resistance that is due to 'affective investment'. It isn't clear whether that much of a difference to the case though. Imagine that I have lots of evidence associating group G with a particular negative attribute. And, like most humans, I'm a little more resistant than ideal when it comes to changing my beliefs. Well, the fact that I believe group G has this negative attitude will probably lead to a negative affect toward group G. Does that mean I'll satisfy the relevant clause? Actually, it's hard to say. Perhaps I have the affective investment in my belief, but that isn't why I'm resistant to counter-evidence here; I'm just generally stubborn that way.

23.3 Putting the Two Together

Putting together the accounts of stereotypes and prejudices, we get the following:

> a negative identity-prejudicial stereotype is a widely held disparaging association between a social group and one or more attributes, where this association embodies a generalization that displays some (typically, epistemically culpable) resistance to counter-evidence owing to an ethically bad affective investment. (Fricker, 2007, 35)

Note that Fricker is focussing on the **identity-based** prejudices here. A prejudice against people with a nervous manner of speaking, for example, would not be identity-based. Fricker is assuming, I take it, that the nervous do not form a **social** group. On the other hand, she is taking social group to be fairly broad-ranging; prejudice against conservatives, or against gays, or possibly even against the disabled, would count as a social group.

And note, as we said above, that Fricker here takes the underlying mental state to be an **association**, not necessarily a belief. Someone can be in the grip of a negative identity-prejudicial stereotype while explicitly believing that the content of it is not true. Indeed, there is evidence that such cases are common. People who expressly avow egalitarian beliefs don't do considerably better on implicit association tasks than the general population.

23.4 Mistakes and Wrongs

Consider the following three cases.

1. H1 has a lower opinion of S1's credibility than is correct. But this is because H1 has been presented with misleading evidence, and he is reacting as an ordinary reasonable person would react to that evidence. For example, H1 knows nothing about S1 except that she is a used car saleswoman, and she knows (in the world of the example - I don't know if this is true in reality) that used car salespeople are typically dishonest, so he infers that S1 is probably dishonest.
2. H2 has a lower opinion of S2's credibility than is correct. And this is unreasonable; either H2's evidence doesn't support this, or H2 could have and should have acquired evidence that changed his view. But this isn't because of a moral failing on H2's part, it's just an epistemic shortcoming. For example, H2 has come to believe that Wikipedia is unreliable relative to other similar sources, and no amount of studies showing its reliability can move him. This is too bad for H2, because it would often be useful for him to be able to rely on Wikipedia for everyday information, and isn't related to any general moral failing on H2's part.
3. H3 has a lower opinion of S3's credibility than is correct. And this is unreasonable; either H3's evidence doesn't support this, or H3 could have and should have acquired evidence that changed his view. And this is all because H3 has a visceral dislike of the social group that S3 is a member of, and so is determined to maintain a low opinion of them.

It's only the last case where Fricker thinks that we have epistemic injustice. It's clear why that is not the case in the first case; H1 hasn't done anything wrong, so it is hard to say he has wronged S1. Fricker gives a number of examples of versions of this case, and the verdict on all of them seems to be that it is simply a case of bad luck all around. But it isn't so clear why that is so in the second case. The intuition she's using seems primarily to be that a morally bad result requires a moral failing as a kind of input. If the reasoning wasn't itself something we can identify as morally bad, then it isn't something we should hold the agent culpable for. Whether Fricker is right about this is something that we'll want to discuss more!

Bibliography

Adler, Jonathan. 2015. "Epistemological Problems of Testimony." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2015 edition.

Alston, William. 1985. "Concepts of Epistemic Justification." *The Monist* 68:57–89.

Aslin, Richard N., Saffran, Jenny R., and Newport, Elissa N. 1998. "Computation of Conditional Probability Statistics by 8-Month-Old Infants." *Psychological Science* 9:321–324, doi:10.1111/1467-9280.00063.

Ayer, A. J. 1954. "Freedom and Necessity." In *Philosophical Essays*, 3–20. New York: St. Martin's.

—. 1956. *The Problem of Knowledge*. London: Macmillan.

Birch, Susan A. J., Akmal, Nazanin, and Frampton, Kristen L. 2010. "Two-year-olds are vigilant of others' non-verbal cues to credibility." *Developmental Science* 13:363–369, doi:10.1111/j.1467-7687.2009.00906.x.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Brueckner, Anthony. 1994. "The Structure of the Skeptical Argument." *Philosophy and Phenomenological Research* 54:827–835.

Burge, Tyler. 1993. "Content Preservation." *Philosophical Review* 102:457–488.

Carruthers, Peter. 2013. "Mindreading in Infancy." *Mind and Language* 28:141–172.

Chalmers, David J. 2005. "The Matrix as Metaphysics." In Christopher Grau (ed.), *Philosophers Explore the Matrix*, 132–176. Oxford: Oxford University Press.

Chisholm, Roderick. 1957. *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.

Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Blackwell.

Church, Ian M. 2013. "Manifest Failure Failure: The Gettier Problem Revived." *Philosophica* 41:171–177, doi:10.1007/s11406-013-9418-5.

Clark, Michael. 1963. "Knowledge and Grounds. A Comment on Mr. Gettier's Paper." *Analysis* 24:46–48, doi:10.1093/analys/24.2.46.

Clifford, William J. 1876. "The Ethics of Belief." *Contemporary Review* 29:289–309.

Coady, C. A. J. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.

Cohen, Stewart. 1984. "Justification and Truth." *Philosophical Studies* 46:279–295.

—. 2002. "Basic Knowledge and the Problem of Easy Knowledge." *Philosophy and Phenomenological Research* 65:309–329.

—. 2005. "Why Basic Knowledge is Easy Knowledge." *Philosophy and Phenomenological Research* 70:417–430, doi:10.1111/j.1933-1592.2005.tb00536.x.

Conee, Earl and Feldman, Richard. 2004. *Evidentialism: Essays in Epistemology*. Oxford: Oxford University Press.

Descartes, René. 1641/1996a. *Meditations on First Philosophy, tr. John Cottingham*. Cambridge: Cambridge University Press.

—. 1641/1996b. *Meditations on First Philosophy, tr. John Cottingham*. Cambridge: Cambridge University Press.

Edgington, Dorothy. 1995. "On Conditionals." *Mind* 104:235–327.

Einav, Shiri and Robinson, Elizabeth J. 2011. "When Being Right Is Not Enough: Four- Year-Olds Distinguish Knowledgeable Informants From Merely Accurate Informants." *Psychological Science* 22:1250–1253, doi:10.1177/0956797611416998.

Fodor, Jerry A. 1990. *A Theory of Content and other essays*. Cambridge, MA: MIT Press.

Fricker, Miranda. 2007. *Epistemic Injustice*. Oxford: Oxford University Press.

Gendler, Tamar Szabó and Hawthorne, John. 2005. "The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies." *Philosophical Studies* 124:331–352, doi:10.1007/s11098-005-7779-8.

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23:121–123, doi:10.2307/3326922.

Gilbert, Daniel T. 1991. "How Mental Systems Believe." *American Psychologist* 46:107–119.

Gilbert, Daniel T., Krull, Douglas S., and Malone, Patrick S. 1990. "Unbelieving the Unbelievable: Some problems in the rejection of false information." *Journal of Personality and Social Psychology* 59:601–613.

Gilbert, Daniel T., Tafarodi, Romin W., and Malone, Patrick S. 1993. "You Can't Not Believe Everything You Read." *Journal of Personality and Social Psychology* 65:221–233.

Goldman, Alvin I. 1976. "Discrimination and Perceptual Knowledge." *The Journal of Philosophy* 73:771–791.

Gopnik, Alison. 2009. *The Philosophical Baby: What Children's Minds Tell Us About Truth, Love, and the Meaning of Life*. New York: Farrar, Straus and Giroux.

Gopnik, Alison, Sobel, David M., Schulz, Laura E., and Glymour, Clark. 2001. "Causal Learning Mechanisms in Very Young Children: Two-, Three-, and Four-Year-Olds Infer Causal Relations From Patterns of Variation and Covariation." *Developmental Psychology* 37:620–629, doi:10.1037//0012-1649.37.5.620.

Greco, John. 2009. "Knowledge and Success from Ability." *Philosophical Studies* 142:17–26, doi:10.1007/s11098-008-9307-0.

Harris, Paul L. and Corriveau, Kathleen H. 2011. "Young Children's Selective Trust in Informants." *Philosophical Transactions of the Royal Society B* 366:1179–1187, doi:10.1098/rstb.2010.0321.

Hart, Betty and Risley, Todd R. 1995. *Meaningful Differences in the everyday Experience of Young American Children*. Baltimore, MD: Paul H Brookes.

Hawthorne, John. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

—. 2005. "Knowledge and Evidence." *Philosophy and Phenomenological Research* 70:452–458, doi:10.1111/j.1933-1592.2005.tb00540.x.

Hawthorne, John and Stanley, Jason. 2008. "Knowledge and Action." *Journal of Philosophy* 105:571–90.

Hazlett, Allan. 2010. "The Myth of Factive Verbs." *Philosophy and Phenomenological Research* 80:497–522, doi:10.1111/j.1933-1592.2010.00338.x.

Heyes, Cecilia. 2014. "False belief in Infancy: A Fresh Look." *Developmental Science* 17:647–659, doi:10.1111/desc.12148.

Holton, Richard. 1997. "Some Telling Examples:Reply to Tsohatzidis." *Journal of Pragmatics* 28:625–8.

Hume, David. 1739/1978. *A Treatise on Human Nature*. Oxford: Clarendon Press, second edition.

Ichikawa, Jonathan Jenkins and Steup, Matthias. 2013. "The Analysis of Knowledge." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2013 edition.

Jackson, Frank. 1987. *Conditionals*. Blackwell: Oxford.

Jaswal, Vikram K., McKercher, David A., and VanderBorght, Mieke. 2008. "Limitations on Reliability: Regularity Rules in the English Plural and Past Tense." *Child Development* 79:750–760.

Keller, Simon. 2004. "Friendship and Belief." *Philosophical Papers* 33:329–351.

Klein, Peter. 2013. "Skepticism." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2013 edition.

Koenig, Mellisa A., Clément, Fabrice, and Harris, Paul L. 2004. "Trust in Testimony: Children's Use of True and False Statements." *Psychological Science* 15:694–698, doi:10.1111/j.0956-7976.2004.00742.x.

Kripke, Saul. 2011. *Philosophical Troubles*. Oxford: Oxford University Press.

Lackey, Jennifer. 2005. "Testimony and the Infant/Child Objection." *Philosophical Studies* 126:163–190, doi:10.1007/s11098-004-7798-x.

—. 2006. "Knowing from Testimony." *Philosophy Compass* 1:432–448, doi:10.1111/j.1747-9991.2006.00035.x.

—. 2008. *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

Lewis, David. 1973a. "Causation." *Journal of Philosophy* 70:556–567. Reprinted in *Philosophical Papers*, Volume II, pp. 159-172.

—. 1973b. *Counterfactuals*. Oxford: Blackwell Publishers.

—. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13:455–476. Reprinted in *Philosophical Papers*, Volume II, pp. 32-52.

—. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74:549–567, doi:10.1080/00048409612347521. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 418-446.

Luzzi, Federico. 2010. "Counter-Closure." *Australasian Journal of Philosophy* 88:673–683, doi:10.1080/00048400903341770.

Malmgren, Anna-Sara. 2006. "Is There A Priori Knowledge by Testimony?" *Philosophical Review* 115:199–241, doi:10.1215/00318108-2005-015.

Moore, G. E. 1959. "A Defence of Common Sense." In *Philosophical Papers*, 32–59. London: Allen and Unwin.

Nagel, Jennifer. 2014. *Knowledge: A Very Short Introduction*. Oxford: Oxford University Press.

Nozick, Robert. 1981. *Philosophical Explorations*. Cambridge, MA: Harvard University Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Pritchard, Duncan. 2009. "Safety-Based Epistemology: Whither Now?" *Journal of Philosophical Research* 34:33–45.

Pryor, James. 2000. "The Sceptic and the Dogmatist." *Noûs* 34:517–549, doi:10.1111/0029-4624.00277.

—. 2001. "Highlights of Recent Epistemology." *British Journal for the Philosophy of Science* 52:95–124.

—. 2013. "Problems for Credulism." In Chris Tucker (ed.), *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*, 89–131. Oxford: Oxford University Press.

Radford, Colin. 1966. "Knowledge–By Examples." *Analysis* 27:1–11.

Rodriguez-Pereyra, Gonzalo. 2006. "Truthmakers." *Philosophy Compass* 1:186–200, doi:10.1111/j.1747-9991.2006.00018.x.

Russell, Bertrand. 1921/2008. *The Analysis of Mind*. Allen and Unwin. Retrieved from Project Gutenberg - `http://www.gutenberg.org/ebooks/2529`.

—. 1948. *Human Knowledge: Its Scope and Limits*. London: Allen and Unwin.

Ryle, Gilbert. 1949. *The Concept of Mind*. New York: Barnes and Noble.

Saffran, Jenny R., Aslin, Richard N., and Newport, Elissa L. 1996a. "Statistical Learning by 8-Month-Old Infants." *Science* 274:1926–1928.

Saffran, Jenny R., Newport, Elissa L., and Aslin, Richard N. 1996b. "Word Segmentation: The Role of Distributional Cues." *Journal of Memory and Language* 35:606–621.

Sainsbury, Mark. 1995. "Vagueness, Ignorance and Margin for Error." *British Journal for the Philosophy of Science* 46:589–601.

Sosa, Ernest. 1991. *Knowledge in Perspective*. Cambridge: Cambridge University Press.

—. 1999. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13:137–49.

—. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.

—. 2009. *Reflective Knowledge: Apt Belief and Reflective Knowledge*, volume II. Oxford: Oxford University Press.

—. 2010. *Knowing Full Well*. Princeton: Princeton University Press.

Sperber, Dan, Clément, Fabrice, Heintz, Christophe, Mascaro, Olivier, Mercier, Hugo, Orrigi, Gloria, and Wilson, Deirdre. 2010. "Epistemic Vigilence." *Mind and Language* 25:359–393, doi:10.1111/j.1468-0017.2010.01394.x.

Stoltz, Jonathan. 2007. "Gettier and Factivity in Indo-Tibetan Epistemology." *Philosophical Quarterly* 27:394–415, doi:10.1111/j.1467-9213.2007.493.x.

Stroud, Sarah. 2006. "Epistemic Partiality in Friendship." *Ethics* 116:498–524.

Turri, John. 2011. "Manifest Failure: The Gettier Problem Solved." *Philosophers' Imprint* 11:1–11.

Vogel, Jonathan. 1990. "Are There Counter-Examples to the Closure Principle?" In M. Roth and G. Ross (eds.), *Doubting: Contemporary Perspectives on Skepticism*, 13–27. Dordrecht: Kluwer.

Warfield, Ted A. 2005. "Knowledge from Falsehood." *Philosophical Perspectives* 19:405–416, doi:10.1111/j.1520-8583.2005.00067.x.

Weatherson, Brian. 2013. "Margins and Errors." *Inquiry* 56:63–76, doi:10.1080/0020174X.2013.775015.

Wedgwood, Ralph. 2013. "A Priori Bootstrapping." In Albert Casullo and Joshua C. Thurow (eds.), *The A Priori in Philosophy*, 225–246. Oxford: Oxford University Press.

Weisberg, Jonathan. 2012. "The Bootstrapping Problem." *Philosophy Compass* 7:597–610, doi:10.1111/j.1747-9991.2012.00504.x.

White, Roger. 2006. "Problems for Dogmatism." *Philosophical Studies* 131:525–557.

Williamson, Timothy. 1994. *Vagueness*. Routledge.

—. 2000. *Knowledge and its Limits*. Oxford University Press.

—. 2005. "Replies to Commentators." *Philosophy and Phenomenological Research* 70:468–491.

—. 2009. "Reply to Critics." In Patrick Greenough and Duncan Pritchard (eds.), *Williamson on Knowledge*, 279–384. Oxford: Oxford University Press.

—. 2013. "Gettier Cases in Epistemic Logic." *Inquiry* 56:1–14, doi:10.1080/0020174X.2013.775010.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. London: Macmillan.

Zagzebski, Linda. 1994. "The Inescapability of Gettier Problems." *Philosophical Quarterly* 44:65–73.

—. 1996. *Virtues of the Mind*. Cambridge: Cambridge University Press.