Scepticism

Brian Weatherson

Winter 2012

Contents

Last revised: **April 10, 2012**

Chapter 1

Sceptical Arguments and Sceptical Scenarios

1.1 Epistemic Disasters

We're all familiar with sceptical arguments that make appeal to epistemic disaster scenarios. In the 17th century the canonical disaster scenario was the evil demon Descartes worries about in the first Meditation. In the 20th century it was the brain-in-a-vat scenario. Later it became the *Matrix* scenario, or perhaps now *Inception*. But we should also worry about scenarios where it isn't quite true that *nothing* is as it appears, but rather scenarios where a lot of things are not as they appear.

As Peter Klein (2011) notes, *The Truman Show* presents a slightly different worry. Klein notes that the protagonist of that scenario has the ability to work out that it is really all fake. But we might also note that until that happens, it isn't true that quite *everything* is fake. When Truman appears to pick up a book, or look at his hands, he really picks up a book, or looks at his hands. What he is mislead about, rather systematically, concerns the motivations of others. He's facing a kind of sceptical scenario concerning other minds, where the most natural inference from his perceptions to facts about other minds will almost always go wrong.[1] Along these lines, we might consider Russell's discussion of the world that popped into existence five minutes ago, full of dinosaur bones, apparent memories, etc (Russell, 1921/2008, Lecture IX). We can use this scenario to argue for scepticism about the past, just as we can use the Cartesian scenario to argue for scepticism about the external world. Hume doesn't avert quite as

---

[1] It might be thought that this isn't quite right, since Truman doesn't go wrong in thinking that the people around him have minds. If scepticism about other minds is scepticism that other minds *exist*, then this isn't a real sceptical scenario. But note that some of our canonical disaster scenarios are not really cases where the external world does not exist. It certainly does exist in the brain-in-a-vat scenario. And those scenarios are still properly considered as key to certain arguments for external worlds scepticism. The short point is that one way to be a sceptic about X is to say we don't know that X exists, but another way is to say that we're systematically wrong in our beliefs about X.

explicitly to scenarios in which the future is completely unlike the past in his arguments for inductive scepticism. He prefers to consider scenarios in which a particular inference fails, rather than an inference in which all of our inductive inferences fail at once. But there is at least a family resemblance between Hume's discussions of worlds in which the future fails to resemble the past, and these scenarios where a certain class of inference systematically fails.

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 that the external world exists and is like she thinks it is.
- 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.

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 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 four sceptical arguments all have their appeal.

Sensitivity Argument

1. Were the disaster scenario actual, then S would still believe $\neg SH$, but that belief would be false.
2. If a belief is to constitute knowledge, then it 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 $\neg SH$.

Underdetermination Argument

1. S has the same evidence in the disaster scenario as in the real world.
2. Knowledge supervenes on evidence, so if S has the same evidence in two scenarios then she has the same knowledge in those scenarios, even if the scenarios are in different possible worlds.

_____

C. S does not know that O.

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 $\neg SH$.

Methods Argument (Quantified)

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

_____

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

Each of these four arguments, especially the last, could do with some tightening up of the premises. But I think they're useful enough to start with.

Note that the arguments do not all have the same conclusion. Three of them 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 three 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 $\neg SH$.

But from this premise, and the intermediate conclusion that S does not know that $\neg SH$, nothing much follows. We can't use *modus tollens* unless we know that the consequence of our new conditional is false. And it doesn't follow from S's not knowing that $\neg SH$ that she *couldn't* know it.

But actually this looks like a fairly easy hurdle to get around. If the above arguments are sound, then 'modalised' versions of them, where each new premise says that the old premise must be true, and the new conclusion says that the old conclusion must be true, should be sound as well. The key point here is that the premises don't seem to rest on any kind of contingent fact about the world. (This is perhaps less clear when we consider the methods argument. We'll return to this point in a bit.) But the new, modalised versions of the arguments have as their conclusion that S *couldn't* know that $\neg SH$. And that, combined with our additional premise, implies that S doesn't know that O.

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. Indeed, I'm sympathetic to a reasonably strong sceptical view about consciousness in non-humans. I think we simply can't know whether certain entities, especially entities with the ability to react to changes in their environment in highly sophisticated ways, but with very different physiologies to ours, are conscious. Or, if they are conscious, we can't know very much about what it is like to be them. And if I tried to articulate my reasons for holding such a sceptical view, they wouldn't be that dissimilar to some of the arguments above. So you might think that these arguments will be helpful to a 'limited' sceptic, who simply wants to argue for scepticism about, say, silicon minds, or 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.

1.2   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. 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. One thing that we can take away from the discussion of scepticism in Moore (1959) is that we can even use the falsity of these forms of scepticism as premises in our reasoning. For instance, we can use the fact that trivialism (the thesis that all propositions are true) is not only false, but knowably false, as a premise in various arguments.[2] 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 I guess *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

---

[2] It's worth keeping trivialism in mind when we think about scepticism and circular arguments.

supervenience premise is false. Timothy Williamson (2000a) 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, independently of whether we think the conclusion is plausible.

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. These arguments say that since we know a lot of things, some particular theory must be true, since only if that theory is true can the sceptic be defeated. Another way to put this kind of argument is that the theory in question is the only thing that stands between us and scepticism. We can see this most clearly in the modern literature in epistemological arguments for externalist theories of perception. (See, for example, Campbell (2002) or Brewer (1999).) But we also see something similar, I believe, in Lewis's argument for contextualism, and indeed in any argument that promotes contextualism by appeal to how the contextualist can respond to the sceptic. 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 both these cases, and in many other transcendental arguments that have been offered over the years, the answer is that the theory doesn't help with any premise of any plausible sceptical argument.

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. I'm interested in an argument that two-dimensionalism (of the broad kind upheld by Frank Jackson and David Chalmers (2001)) is committed to every premise of what we'll call the Humean argument for scepticism. That's not an argument that scepticism is true obviously; it's an argument that this kind of two-dimensionalism is false. But I think the only way to see whether that's right is to drill down fairly deep into the details of arguments for scepticism. And a cheap dismissal of the subject matter of scepticism will prevent us doing that.

Chapter 2

Descartes

2.1 Descartes' Sceptical Arguments

When we introduce the evil demon in introductory epistemology classes, it is usually to motivate a kind of scepticism about the external world. But that isn't what Descartes (1641/1996) uses the demon for. Indeed, it is arguable that the demon doesn't play a large role in argumentative structure of the First Meditation. If the evil demon isn't a core part of the argument, what is it doing? I think Descartes has two aims in the First Meditation.

1. He wants to argue that most of our beliefs do not constitute knowledge, and his argument for this turns on the unreliability (or at least the non-hyper-reliability) of our ordinary belief forming practices.
2. He wants to show that it is psychologically possible to doubt almost all of our previous beliefs.

If you are prepared to countenance the obvious anachronisms involved, you can think of the second aim as a response to a certain kind of pragmatist. Think of the kind of philosopher who objects to foundationalism by an airy appeal to Neurath's boat. Such a philosopher may argue that since it is psychologically necessary to manage our doxastic states from within those states, foundationalism is inappropriate as a theory of human epistemology.

The foundationalist picture is that how justified our beliefs are depends on their relationship to their foundations. In principle, it might be impossible to get from where we are, doxastically, to where we should be. This is a problem for a lot of theories of justification, but it seems to be a particular problem for the foundationalist. If our beliefs are not coherent, it's easy to see how we can move towards more coherence. Indeed, bringing more coherence to our beliefs is one aim of everyday reasoning. But if there are systematic discrepancies between our beliefs and what we should believe given the foundations, it is going to be hard to

correct any one belief without changing the network of other beliefs that we use to get from foundations to conclusions.

Now one way the foundationalist could reply to the objector wielding Neurath's boat is to argue that the picture they give is a kind of ideal. Even if it is unobtainable, they might argue, it is good to know what the perfect state is so we can know what we imperfect beings can aim at. I'm rather dubious of the value of such projects. Perhaps it is intrinsically useful to know what the ideal state is. But we shouldn't pretend that this will be of much practical use. As a rule, unless a system is only one 'step' away from the ideal state, it is an open question whether making it more like the ideal will make it better or worse. (For many demonstrations of this, see Lipsey and Lancaster (1956-1957); Kennett and Smith (1996); Williamson (1998).) In a lot of cases, it will make it worse. So unless we want to give up on the idea that epistemology should be useful, we need to show how our theories can be implemented.

Descartes, at least on my reading, decides to confront the Neurath objection head on. He claims that it is possible to doubt (almost) all of our beliefs. It isn't *easy* to do so. In most circumstances it isn't even advisable. But if we secure for ourselves a safe, quiet (perhaps even dark) room, and carefully reflect on the kind of beliefs we have, it might just be possible. And this, I think, is where the evil demon eventually comes in. But let's build up to that.

Descartes starts off by endorsing the following two, closely related, principles:

- We should reject any belief which is not 'completely certain and indubitable'.
- Once the 'foundation' on the basis of which a belief is held is undermined, there is no reason to hold the belief.

We might wonder what it takes to demolish the foundations of our belief in the way Descartes suggests. It's plausible that if a belief that p is based on a belief that q, and q is false, then the belief that p is somewhat defective. It's arguable, for example, that nothing inferred from a falsehood is knowledge, and that we should only believe what we know. But it seems Descartes needs something stronger here. He needs it to be the case that if q is doubtful, and q plays a role in forming the belief that p, then p too is doubtful. This is a plausible epistemological standard in mathematics, but it doesn't obviously extend to empirical methodology.

In the third paragraph of the First Meditation, Descartes offers a very strong constraint on knowledge: **"It is prudent never to trust those who have deceived us even once."** It is tempting to attribute to Descartes the principle that any method that fails even once is not a method that can produce knowledge.

But we have to be a little careful about just how strong a sceptical principle we really say Descartes commits to here. Note that Descartes gives a quick, and compelling, argument that perception is not perfectly reliable. That's because some things look different from different vantage points, so not all of the appearances can be accurate. But he doesn't think that this shows we can't have perceptual knowledge. Indeed, at this stage he still thinks it looks mad to think that perceptions of our immediate environment are mistaken.

It's to get over this idea that doubting perception is madness that Descartes brings in dreams. Just how this is relevant to the argument depends on just how we interpret Descartes' sceptical reasoning. Here are two distinct kinds of argument that might make dreams central.

Reliability-Based Dreaming Argument

1. *S* uses the same methods to form perceptual-based beliefs about her immediate environment as she does when she forms beliefs about her apparent immediate environment in a dream.
2. When *S* uses these methods in dreams, the methods lead to false beliefs.
3. Only methods that never lead to false beliefs are capable of grounding knowledge.

C. *S*'s perceptual-based beliefs about her immediate environment do not amount to knowledge.

Underdetermination-Based Dreaming Argument

1. *S* has the same evidence for her perceptal-based beliefs about her immediate environment as does her counterpart who is dreaming that she is an environment just like *S*.
2. *S*'s counterpart does not know, and indeed cannot know, that she is in an environment just like *S* is in.
3. Knowledge supervenes on evidence, or at least, if two people have the same evidence, and the one knows that *p*, then the other can know that *p*.

C. *S*'s perceptual-based beliefs about her immediate environment do not amount to knowledge.

Let's look through the premises in turn.

The first premise in each argument puts forward a similarity between *S* and a dreamer. In the reliability-based argument, the similarity is that they use the same methods. I think Descartes thinks this is true, and I think he thinks that because he has a broadly phenomenological conception of methods. What matters is whether it seems to us that it is the same kind of reasoning. In the underdetermination-based argument, the similarity is that the evidence is the same. Provided we think that evidence is phenomenological, and we allow that counterparts might be in other possible worlds, this premise is reasonably secure.

You might think that the second premise in each argument is beyond dispute. But there is one small ground for doubt about the second premise in the reliability-based argument, namely that it isn't obvious that we form genuine *beliefs* in dreams. This is a tricky issue about just what a *belief* is, and I don't have a lot to add to recent work by Jonathan Ichikawa (2008, 2009) and Ernest Sosa (2007) on the matter. I will assume that we *do* form beliefs in dreams, but note that this is a substantial assumption, and certainly something one may question in **Reliability-Based Dreaming Argument**.

The third premise in each argument is a very general theoretical premise. In the reliability-based argument, it is a hyper-reliability premise. In the argument based on underdetermination, it is a supervenience premise. The primary difference between the two premises concerns which kinds of failures they take to undermine knowledge. The hyper-reliability premise takes failures of this method, in this or nearby worlds, possibly based on *different* evidence, to undermine knowledge. The supervenience premise takes failures in possibly quite remote worlds, but based the exact same evidence, to undermine knowledge. In a way, you can think of both of these as safety principles, with the difference being in which kind of failures you care about.

There are two reasons to think that Descartes thinks the reliability-based arguments are the arguments that are most in need of a response. The first is his own response to the threat of dream scepticism at the end of the Sixth Meditation. His line there is that there's another method we can use to form beliefs about our immediate environment. We can double check what we seem to see, and see whether our different senses, and each sense over time, is producing a coherent picture. In dreams, this is usually not the case. Different senses do not cohere, and picture each sense presents changes radically over time. I think this is an argument that if we are being careful in our investigations of the world, we are not using a method that actually fails when applied in dreams. (Though, to be sure, he could be arguing that given God's benevolence, a dreamer could never have exactly our evidence.)

The other is that it is hard to see how to generalise the underdetermination-based argument to the worries about arithmetic knowledge that Descartes goes on to consider. Descartes worries that he may be bad at basic arithmetic, even at things that seem quite obvious to him. This could be because God is a deceiver, or because he was created by some defective craftsman who makes creatures who become quite certain of simple falsehoods.

It is very hard, to say the least, to work this kind of worry into the framework of the underdetermination argument. Assume that $2 + 3 = 5$. (Hopefully a safe assumption!) And assume that Descartes believes it. Then there isn't any agent in any possible world who has the same evidence as Descartes who is misled by his evidence. That's because Descartes's evidence includes his strong sense that $2 + 3 = 5$. There could be someone who is very confident that $2 + 3 = 6$, but that would clearly be a different phenomenal state, and not at all the same evidence. If we are to turn the worries about deceiving Gods into an argument for scepticism, it seems that we need to rely on *something like* the reliability assumption. Here's how it could go.

Reliability-Based Deceiving Gods Argument

1. The method that Descartes uses to form simple arithmetic beliefs is that of trusting his God-given cognitive system on matters where it issues a confident judgment.
2. Sometimes that method leads to false beliefs.
3. Only methods that never lead to false beliefs are capable of grounding knowledge.

_____

C. Descartes' simple arithmetic beliefs do not amount to knowledge.

But there is a puzzle here about the domain of the quantifier in premise 2. We can't simply say that it ranges over all agents in all possible worlds. It's absurd to think that a method can't produce knowledge because it leads to a false belief in *some* possible world. Or, at least, it seems palpably absurd to me; though maybe someone could try to defend that. On the other hand, if we restrict the quantifier to actual cases, it becomes less clear that the premise is actually true. Once we can't simply imagine counterexamples, it is much harder to argue that a method leads to failure. So this doesn't look like the strongest sceptical argument ever. Can we come up with a better argument to attribute to Descartes here?

2.2   Scepticism and Iterated Knowledge

Stepping away from Descartes exegesis for a minute, there are a few arguments around here whose evaluation relies on being careful about the distinction between knowing something, and knowing that one knows it. Start with the following argument, which is plainly invalid.

Ignorance of Reliability Argument

1. The method that Descartes uses to form simple arithmetic beliefs is that of trusting his God-given cognitive system on matters where it issues a confident judgment.
2. Descartes does not know whether that his God-given cognitive system is reliable.

_____

C. Descartes' simple arithmetic beliefs do not amount to knowledge.

Since that's clearly invalid, it isn't much of an argument for arithmetic scepticism. And the gap between the premises and the conclusions is no small matter. A reliabilist about arithmetic knowledge may think that the premises are true and the conclusion false.

We could obviously make the argument valid by adding as a premise that one can only get knowledge through a method if one knows that the method is reliable. But it is hard to know exactly what would be the justification for such a premise. If a method works, arguably that's enough for knowledge, even if it isn't known to work.

The following argument is arguably valid (at least in a plausible epistemic logic), but it doesn't quite have the sceptical conclusion we're looking for.

Ignorance of Knowledge Construal Argument

1. The method that Descartes uses to form simple arithmetic beliefs is that of trusting his God-given cognitive system on matters where it issues a confident judgment.
2. Descartes does not know whether that his God-given cognitive system produces beliefs that amount to knowledge.

_____

C. Descartes does not know whether his simple arithmetic beliefs amount to knowledge.

It's an interesting question whether this is a sufficient interesting sceptical conclusion to worry us on its own.[1] It is a kind of scepticism, but it isn't arithmetic scepticism, it is scepticism about epistemic capacities. We could turn it into an argument for arithmetic scepticism by adding a premise:

- If Descartes knows something, he knows that he knows it.

Premises like these are often called **KK-principles**, because of their representation in epistemic logic. If we let Kp mean that the salient agent knows that p, then KKp will mean that this agent knows that she knows that p. Then the principle in question is just $Kp \rightarrow KKp$. The converse of this is trivially true, since knowledge is factive, so we could write this principle as $Kp \leftrightarrow KKp$.

In *Knowledge and Its Limits*, Timothy Williamson offers several arguments that tell against these kinds of principles. The following kind of argument is a small variant on the kind of argument he offers.

Imagine that we're trying to figure out what percentage of Michigan students use Brand X of toothpaste. We run a careful survey of students that says that 34% of students use Brand X. After checking the possible sources of error in the survey, we decide (correctly) that its accuracy is $\pm 5\%$. So we conclude that the proportion of students who use Brand X is in the range $[29\%, 39\%]$. Call that p. Let's say it turns out that exactly 34% of students do use Brand X. Then plausibly we know that p. We've formed the belief that p using a reliable method, and we believe p because p is true. But we don't know much more than p. There was a margin of error in our methods. It's consistent with what we know that just 29% of students use Brand X. Consider the world where that's true, i.e., the world where 29% of students use Brand X, we run the same test we actually ran, and we got the answer that 34% of students use Brand X. We're still careful, so we still believe p. And p is still true, although it's a close call. Do we know p in that world?

There are a few ways to argue that we don't. One way is to note that there are very close worlds to that one in which we still believe p, but p is false. There is a world where only 28% of students use Brand X, and we get the same result, and still believe p. There's a world where the facts are still the same, still 29% of students use Brand X, but our results come in a little bit differently, and we end up concluding that the proportion of students who use Brand X is in the range $[30\%, 40\%]$. That is, we end up drawing a false conclusion from our methods.

---

[1]If we don't know whether we have arithmetic knowledge, then we presumably don't know whether anyone else has arithmetic knowledge either. And that might put pressure on cases where we are relying on other people being arithmetically competent in order to coordinate. Such cases are the bread and butter of game theory.

So neither the belief we formed, nor the method we used to form it, look like knowledge in that world. And that world is one that, for all we know, is actual. So for all we know, we don't know p. That is, we don't know that we know p.

The underlying theoretical point is easy enough to see if you think about things spatially. Knowledge that p requires that things go well, at least with respect to p, in nearby worlds. Knowing that you know p means that things go well, at least with respect to Kp, in all nearby worlds. And you only know p in a world if things go well in worlds nearby to that. So knowing that you know p requires that things go well in all worlds nearby to worlds that are nearby.[2] Now not all worlds that are nearby to a nearby world are themselves nearby. Compare: Not all points that are within a mile of points that are within a mile of here are within a mile of here. Some of them are two miles away. So the conditions for knowledge can be satisfied without the conditions for knowing that one knows being satisfied.

So in principle, an argument that shows one cannot know that one knows p is not yet an argument for scepticism about p. It is a kind of sceptical conclusion, but it is scepticism about epistemic capacities, not about the underlying subject matter.

2.3   Is Descartes' Reply Circular?

There are a few different interpretations of what Descartes is trying to do in Meditations Three and Four. I'm going to mention two interpretations in order to set them aside. This isn't because I think those interpretations are wrong; I simply don't know enough to contribute to these exegetical debates. But the view I'm most interested in, and which may be Descartes' differs from these interpretations.

What we might call the *psychological* interpretation of Descartes says that he is trying to show us that a number of propositions are psychologically beyond doubt. (Such an interpretation is found in Rubin (1976) and Larmore (1984), and there is an important discussion of the broader role of psychological necessity in Descartes in Loeb (1990), though without endorsing this resolution of the circle.) We simply cannot doubt that if we are thinking, we exist, for instance. And he wants to argue on the basis of things that we cannot doubt that we have justification for a lot of beliefs. This doesn't mean that the things we cannot doubt are themselves justified. On this interpretation, the issue of justification may not arise for these propositions since they are psychologically necessary. (There is

---

[2]There's a fairly clear equivocation over 'nearby' here, which is why this isn't the formal argument against $Kp \rightarrow KKp$. But I don't think this undermines the intuitive picture I'm sketching.

a further exegetical issue here about what normative significance Descartes attached to psychological necessity, but I won't enter into that debate.)

What we might call the *natural light* interpretation of Descartes says that he is trying to show us that a number of propositions are epistemologically beyond doubt. These are propositions that we know by the 'natural light', and they are immune to the kinds of doubt that Descartes brings up in the First (and early in the Second) Meditation. But they don't, notably, include all propositions that are clearly and distinctly perceived. So it's obviously non-circular to use this special class of propositions to argue that all propositions clearly and distinctly perceived are true. Rickless (2005) endorses such an interpretation, and discusses several related interpretations. What many of these interpretations have in common is that they identify a class of propositions that are somehow immune to doubt, and argue that we can derive the reliability of clear and distinct perception from just propositions in this class.

Finally, what we might call the *ambitious* interpretation of Descartes says that Descartes aims to give an epistemological justification for all our beliefs, even the most foundational ones, and he aims to do this without positing a special class of beliefs that are immune to doubt. For an example of such a view, see Newman and Nelson (1999) and Newman (2010). They think there is an important *temporal* aspect to Descartes' argument. He needs to show that clear and distinct perception is reliable not so he can use it–we can use clear and distinct perception without knowing it is reliable, just as young children can use vision without knowing it is reliable–but so we can appeal at later points in time to things that we earlier clearly and directly perceived.

It is worth thinking at this point about the epistemology of mathematics. Imagine that I've just completed a short, but significant proof, such as for example Euclid's proof of Pythagorus's Theorem. At that stage I know Pythagorus's Theorem to be true. Do I need to know that mathematical proof is reliable (or, better yet, perfectly reliable) to know this? No; I can learn some mathematics without having to get any meta-mathematical results like that established first. In any case, even if I were to prove that proof is reliable, that would be useless if I needed to know proofs were reliable before using them. But now think about some later stage, when I am trying to establish some other result, and I need to use Pythagorus's Theorem. I have a memory of having proved it earlier. Is that enough? Here it is plausible that I need one of two things. Either I need to be able to recreate the proof, or I need to be sure that the existence of a proof is good enough to use Pythagorus's Theorem. Here then is where knowledge that proof is perfectly reliable may be useful. Let's say, what might be questioned, that to use the fact that I earlier proved that p to support the use of p in an argument for

q, I need to either (a) recall the proof, or (b) know that proof is reliable. Then if I want to prove anything whose proof is too large to apprehend at once, I will need to know that proof is reliable.

That's the kind of thing that Newman and Nelson think Descartes is up to in the argument that clear and distinct perception is perfectly reliable. He isn't trying to shore up the support for his earlier argument that God exists, and is good, and would not allow us to be deceived in ways that we could not overcome. Those things have been shown. But he is trying to give us reason to think we can use those facts in later arguments, without having to constantly recall their proof. This means he never uses the reliability of clear and distinct perception in the proof that clear and distinct perception is reliable, any more than a mathematician uses the reliability of mathematical proof in a proof that a particular proof system is sound, i.e., perfectly reliable.

Still, there is a circularity worry in the area. What remains true is that a key move in Descartes' anti-sceptical strategy, i.e., his strategy for showing that we know a lot about the external world. is to argue that everything clearly and distinctly perceived is true. (Call this the **C & D Rule**.) And it is also clear that Descartes' argument for this includes premises that he came to believe by clearly and distinctly perceiving them. In particular, he clearly and distinctly perceives something like the following propositions.

- He has the idea of God.
- Causes are greater than their effects.
- The only thing that could be greater than the idea of God is God.
- It is inconsistent with God's goodness that clear and distinct perceptions are false.

So there is at least a whiff of circularity about Descartes' moves here, as has been noticed from the very earliest commentaries on the Meditations. But is it badly circular? Let's go through the ways in which an argument can be circular.

Premise Circularity The conclusion is a premise of the argument.

 As it stands, Descartes' argument is not premise circular. He doesn't argue that all clear and distinct perceptions are true, and he clearly and distinctly perceives that he has the idea of God, so (via a few more steps) all clear and distinct perceptions are true. That would be a bad argument, because premise circularity does seem to be bad, but it isn't Descartes' argument. His premises are that he has the idea of God, that causes are greater than their effects, and so on.

It's tempting to think that it's too easy to avoid premise circularity by simply tweaking the premises. Perhaps it is bad to have premises that tacitly include the conclusion, say by implying it.

Implied Premise Circularity The conclusion is implied by the premises of the argument.

But this can't be a flaw in an argument. Indeed, when we're doing deductive reasoning, this is just another way of stating validity. In general, it is very easy when trying to come up with a problematic kind of circularity to end up defining valid arguments as circular. That's a trap we'll have to avoid.

As an illustration of how hard it is to avoid this trap, consider what happens if we relate our notion of circularity to the general notion of a defeater. There's something plausible to the idea of what's wrong with Descartes' argument is that the premises somehow require the conclusion. Even if the conclusion isn't one of the premises, if he thought the conclusion were false, he wouldn't believe the premises. If I thought that my eyesight was defective, I wouldn't use p as a premise solely because I saw that p. That doesn't mean that the non-defectiveness of my eyesight is a premise in every argument I use that involves perceived premises. So let's try turning that into a constraint.

Conclusion as Defeater Circularity I If one believed that the conclusion were not true, then one's justification for the premises, either collectively or individually, would be defeated.

But this is still too strong. Again, it seems to rule out all valid arguments. If p entails c, then believing $\neg c$, would defeat justification for p. So I don't think this can be quite right either.

Perhaps the mistake there was talking about the conclusion being *believed* to be false. Perhaps we should instead say that the premises are undermined by the conclusion actually being false.

Conclusion as Defeater Circularity II If the conclusion were false, then one's justification for the premises, either collectively or individually, would be defeated.

The first thing to note about this principle is that it requires a non-factive view about justification. If justification requires truth, then this circularity principle will once again say that all valid arguments are circular. Since Descartes is looking for a very high standard of justification, this could be relevant to its application to Descartes. It also isn't obvious that Descartes actually violates this principle.

From the falsity of the **C & D Rule** it only follows that some clear and distinct perceptions are false. It doesn't follow that the ones Descartes himself uses are actually false.

In a number of recent papers, Crispin Wright (2000, 2002, 2004) has noted that there are kinds of argument where the premises can't be used to support the conclusion, even though the premises entail the conclusion. One such example concerns a person who walks by a (soccer) field on which some people are kicking a football around. One of the people kicks the ball into the net. Consider the following argument.

Goal

1. A goal was just scored.

_____

C. A football game is being played.

Wright's thought is that we can't use the premise to strengthen our belief in the conclusion. It's possible, given our evidence, that it isn't really a game being played, and instead that the people are just kicking a ball around, or perhaps they are filming a movie and pretending to play a game for that purpose. If those possibilities are actual, then the premise and the conclusion are both false. Now it's true that we wouldn't normally reason through the fact that it's a football game to conclude that there was a goal scored, but still there seems to be a kind of circularity here. Let's note two related principles that this argument potentially violates.

Priority Circularity The person actually using the argument must justifiably believe the conclusion before they can justifiably believe the premises.

Learning Circularity The person actually using the argument could not have come to know the conclusion by coming to know the premise.

Note that both these principles don't relate to arguments as much as *uses* of arguments. Perhaps someone with very different evidence could have learned that a football game was being played by learning that a goal was scored. But given the background of this agent, they couldn't learn a game was being played this way. This gets the direction of learning wrong.

Do Descartes' arguments violate these circularity principles? I think it's arguable that they do not. Could one justifiably believe that the *cogito* is sound without first having a justified belief in the **C & D Rule**? Surely yes; there's a reasonable chance that you the reader are an example of someone who justifiably

believes the *cogito* is sound but has no belief at all, let alone a justified belief, in the **C & D Rule**. Could one justifiably believe that the **C & D Rule** is the best explanation of how we know the *cogito* to be sound without first justifiably believing the **C & D Rule**? Again, it seems that one could. Perhaps this conclusion could come about by a process of elimination, for example. So it doesn't look like Descartes is violating either of these two principles in his argument.

Finally, let's return to some notions to do with rule circularity. This looks like a bad kind of circularity.

Simple Rule Circularity The conclusion is that one of the rules used in the argument is a good rule.

But on reflection, it isn't clear that this is ever so bad. Compare the following piece of reasoning. Assume $(p \rightarrow q) \wedge p$. That clearly implies $(p \rightarrow q)$ and p. And those collectively imply q. So $(p \rightarrow q) \wedge p$ implies q. That is, $[(p \rightarrow q) \wedge p] \rightarrow q$ must be true. Now, depending on just what one thinks about schematic variables, that conclusion might just *be* a statement of the validity of modus ponens. But the reasoning doesn't look problematically circular to me. Indeed, it looks like the reasoning by which I learned that $[(p \rightarrow q) \wedge p] \rightarrow q$ in my undergraduate logic class. So I'm not convinced that Simple Rule Circularity is bad.

It's even clearer, I think, that this kind of circularity is not problematic.

Priority Rule Circularity The agent must have justification to believe the conclusion before she has justification to believe that the rules used in the argument are acceptable.

This isn't problematic because even if an argument has this feature, it doesn't imply that an agent can't use it to learn the conclusion. And that's because an agent doesn't have to know that a rule is good in order to reason in accordance with it in a way that extends their knowledge. To see this, note that any number of people do learn things by deductive reasoning without being able to properly state the rules that they are using, let alone know that those rules are correct.

This is not to commit to the strong view that all that matters for good reasoning is that the rules being used are in fact correct. Such a view is tempting, especially when thinking about cases like the beginning mathematician, but it probably isn't right. It might be that if we have positive reason to believe a rule is bad, then we can't use it, even if it is a good rule. Indeed, I suspect that if someone actually believes that a rule is bad, they can't acquire knowledge using it, even if it is a good rule. But such a clause wouldn't raise any distinctive problems for Descartes, and would be very different from a general bar on violating the anti-circularity principles mentioned here.

The Cartesian Circle is far from the only place where sceptical arguments and concerns about circular reasoning have run up against one another. Hume's arguments for inductive scepticism lean heavily on certain anti-circularity principles. We'll turn to those arguments next.

Chapter 3

Hume

3.1 Hume's Sceptical Argument

Hume (1748/2004), famously, argues for a kind of inductive scepticism. The broad structure of the argument is clear enough. All reasoning is either demonstrative or based on experience. We can't reason demonstratively for our inductive conclusions, since they could be false, and good demonstrative reasoning can't produce results that could be false. We can't reason on the basis of experience to inductive conclusions for a couple of reasons. First, such reasoning would be circular. Second, such 'reasoning' has odd features when it comes to how we deal with the accumulation of evidence. So our empirical beliefs are not grounded in reasons.

It is easy enough to find textual support for at least this much in Hume. If we try to go beyond that, we start to run into philosophical and exegetical issues. Here are some of the issues we must face.

On the Relationship Between Induction and Demonstrative Reasoning

1. What role do principles about conceivability and possibility play?
2. What role do principles about demonstration and possibility play?
3. Are Hume's premises here any more plausible than his conclusion?

On the Argument that Induction cannot be Based on Experience

1. Is the central challenge that such reasoning would be circular, or that it would fall into a regress?
2. If it is a kind of circle, what kind of circle is it?
3. Is what Hume says about the things we actually presuppose in reasoning actually true, and if not, does this matter?

4. How much importance should we place on the argument that reasoning cannot be sensitive to mere repetition of the data, in the way that induction appears to be?
5. If we are meant to place a lot of weight on this argument, is it any good?
6. Is the consideration that inductive reasoning must be simple enough for babies to use, and hence easy to describe, meant to be a key part of the argument?
7. And is it at all persuasive?

On the Conclusion of the Argument

1. Is the conclusion that inductive inference cannot be justified?
2. Is the conclusion that inductive inference is always unjustified?

Let's take all of these in turn.

3.2   Induction and Demonstration

Hume's argument that induction cannot be grounded in demonstration seems to occasionally be something like this.

1. In any inductive inference, we can conceive that the premises are true and the conclusion false.
2. Anything that we can conceive to be is possible.
3. Demonstrative reasoning can never prove the falsity of anything that is possible.

C. Inductive reasoning cannot be reduced to demonstrative reasoning

We can question every one of these premises. Indeed, the conclusion is arguably more plausible than any of the steps of the argument.

   The first premise is not obviously false, but it has been challenged by Robert Pargetter and John Bigelow (1997). They argue that Hume misstates the structure of an inductive inference. It's not true that our inductive arguments look like this:

1. I've seen lots of black swans, and no white swans.

C. All swans are black

Rather, it is more like the following:

1. I've seen lots of black swans, and no white swans.
2. That's all the relevant evidence I have concerning swan colour.

   _____

C. Given my evidence, it is probable that all swans are black.

And, they argue, it isn't clear that we can conceive that the premises are true and the conclusion false. This is a very contentious way to treat induction. And it is pretty concessive, since it agrees that we cannot in the first instance reason to conclusions like "All swans are black." (Though perhaps that is something we can reasonably conclude on the basis of it being probable that all swans are black.)

The second and third premises seem clearly false though. It is conceivable that water is atomic. But this isn't possible, since water is necessarily H_2O. And some things that are possible can be demonstratively shown to be false. Let Ap mean that p is actually true, and consider any sentence of the form $\neg(p \wedge \neg Ap)$. That could be true. Consider any p that is false, but could have been true. That is to say, it isn't actually true, but it could have been. Had it been true, it would have had a different truth value to what it actually had. In particular, it would have been true in such a world, though it is actually false. That is, $p \wedge \neg Ap$ would be true. So $\neg(p \wedge \neg Ap)$ is possible. But $p \leftrightarrow Ap$ is typically taken to be a theorem (or even an axiom) of the logic of actuality. So we can demonstratively prove some things that could have been false.

Perhaps though none of this matters. A lot of people will think that Hume's conclusion is plausible independently of the strength of his preferred argument for it. And some of the things that Hume says are clearly meant to support this kind of view. He notes how odd it would be to think we could justify a priori some of the things we have learned about mechanics demonstratively. And in general, it just seems odd to think that we could know a priori that we aren't brains in vats. I occasionally suspect this is just a pernicious kind of infallibilist intuition creeping in. (For a development of this thought, see Weatherson (2005).) But to most people this premise of Hume's looks plausible.

3.3 Induction Based on Experience

It's harder to bed down exactly what Hume's argument concerning the other horn of the dilemma is. At times he seems to think that grounding inductive reasoning in our experience will lead to a short, and vicious, circle. For instance, he writes

> We have said that all arguments concerning existence are founded
> on the relation of cause and effect; that our knowledge of that rela-
> tion is derived entirely from experience; and that all our experimen-
> tal conclusions proceed upon the supposition that the future will be
> conformable to the past. To endeavour, therefore, the proof of this
> last supposition by probable arguments, or arguments regarding exis-
> tence, must evidently be going in a circle, and taking that for granted,
> which is the very point in question (Hume, 1748/2004, 26-7).

At times he seems to think that it will lead to a problematic regress. Here is
how he puts a version of the argument in the *Treatise*.

> Your appeal to past experience decides nothing in the present case;
> and at the utmost can only prove, that that very object, which pro-
> duced any other, was at that very instant endowed with such a power;
> but can never prove, that the same power must continue in the same
> object or collection of sensible qualities; much less, that a like power
> is always conjoined with like sensible qualities. Should it be said,
> that we have experience, that the same power continues united with
> the same object, and that like objects are endowed with like pow-
> ers, I would renew my question, why from this experience we form
> any conclusion beyond those past instances, of which we have had
> experience. If you answer this question in, the same manner as the
> preceding, your answer gives still occasion to a new question of the
> same kind, even in infinitum; which clearly proves, that the forego-
> ing reasoning had no just foundation (Hume, 1739/1978, 91).

And at other times he puts a lot of weight on an allegedly odd feature of
our inductive practices, namely that more cases make the inductive argument
stronger, even if the cases are alike in all respects.

The circularity argument is easy to understand, at least in outline. Hume
thinks that when we ask ourselves what justifies our inductive practices, the an-
swer includes our presupposition that the future will resemble the past. But when
we ask what justifies that presupposition, we find that the answer is, at best, our
inductive practices. We have a tight circle.

There are, however, several responses we might make here. We might try
to argue that the canons of induction are really **rules** not extra premises, and
rule circular arguments are not problematic, for reasons discussed in the previ-
ous chapter. The problem with this approach is that it looks rather implausible
until we can state just what these rules are, and why, like the rules of deductive

inference, they are unjustified justifiers. The absence of much progress in saying what those rules are seems to tell against this response. In general, we should be a little sceptical of the free appeal to the difference between rule circularity and premise circularity; if anything could be a rule, and rules don't need to be justified, then nothing will need justification. One way to curb this proliferation of unjustified justifiers is to insist that rules are closely analogous to the rules that we use in deductive reasoning. And that may well require that the rules be precisely statable.

I also think there is something confused about treating inductive steps in reasoning as parallel to deductive steps. If we do make them parallel, we get an odd result. If inductive inferences are really the results of rules, then there can't be a rule R such that applying rule R leads to justified beliefs, but we are not justified in believing that R is universally truth-conducive. Consider, for instance, the following very plausible looking inference.

Infinite Rule

If we know there are infinitely many Fs, and that at most one of them is $\neg G$, and we know nothing about a except Fa, then we can infer Ga.

That seems like a pretty sound ampliative inference, if anything is. But now let's see how it works in practice. Let F be the property of being a positive integer, and note we can now reason as follows.

1. Let i and j be arbitrary positive integers.

2. By **Infinite Rule**, $i \neq j$. (In this application of the rule, F is being a positive integer, and G is being not equal to j.)

3. Since x was arbitrary, we can conclude $\forall x : x \neq j$. (This is a straightforward application of $\forall$-introduction.)

4. And from that we can conclude $j \neq j$, which is a contradiction.

So either **Infinite Rule** is a bad rule, or there's something wrong with how we applied it, or $\forall$-introduction or $\forall$-elimination, as usually stated in logic textbooks, are bad rules. Since the last three options seem unlikely to me, I conclude **Infinite Rule** is a bad rule. So I don't really think we can use rules in inductive reasoning; we simply have to reason.

3.4 The Varieties of Induction

We could try to argue, with some plausibility, that Hume has overstated our dependence on a simple principle of resemblance between the past and future, or between the observed and the unobserved. Elliot Sober (1988, Ch. 2) makes a version of this complaint. He says, in effect, that at least one version of Hume's argument turns on a confusion between a claim of the form $\exists\forall$ and a claim of the form $\forall\exists$. We can see the difference between these claims by looking at sentences like (1) and (2). (In these examples, Hx means that x is a human, and Byz means that y is the brain of z.)

(1) $\forall x Hx \to \exists y Byx$
(2) $\exists y \forall x Hx \to Byx$

Clearly (1) is true while (2) is false. In general, statements like these where the existential scopes over the universal are strictly stronger than the ones where the universal scopes over the existential. Note that in this case, for example, (2) entails (1), even in relatively weak logics.

The relevance to Hume's argument comes in the discussion of presuppositions of any inductive inference. Hume sometimes suggests that *every* inductive inference presupposes that the future will resemble the past. For instance, he writes

> For all inferences from experience suppose, as their foundation, that the future will resemble the past, and that similar powers will be conjoined with similar sensible qualities (Hume, 1748/2004, 28).

But this isn't true at all. For one thing, we sometimes infer from part of a group to the whole group, as in statistical sampling. For reasons related to the grue paradox, we know that the future can't resemble the past in *every* way. (We'll return to this point a lot in what follows.) If there is any presupposition, it is something more fine-grained, such as a presuppostion that the unobserved resembles the observed in the particular respect we're interested in.

Now one consequence of this is that there may be no single proposition that is presupposed in every inductive inference. Even if it is true that in every inference there is some empirical claim that is presupposed, it may be a different claim in different inferences. So there is no quick proof here that inductive inference is necessarily circular.

But, as I think Sober acknowledges, this only handles one version of Hume's complaint. In particular, this doesn't help with Hume's claim that inductive *reasoning* will inevitably lead to a regress, even if it is not circular. Now perhaps

we could supplement this with an argument that regresses are not as problematic as circles, but I don't see why that will be much easier than attacking Hume's argument directly.

There is a puzzle for rescuing a narrowly *Humean* argument here, in that Hume quite clearly intends his argument to rest in large part on empirical facts about how we reason inductively, but it's also quite clear that his empirical story is at best a massive over-simplification (Okasha, 2001). From the very top he says that every inductive inference we do involves cause and effect. Now either that has to be interpreted so narrowly as it can't actually ground an argument for Hume, or it isn't true. It isn't true in statistical sampling, for instance.

This doesn't mean that the *form* of Hume's argument can't be rescued. Hume is raising, in effect, a very hard question about the foundation, or foundations, of our inductive practices. Does that foundation consist on an unjustified presupposition that the future will resemble the past, or at least resemble it in certain respects? Does it consist of an inductive rule that we are entitled to use without having a justification for it? Does it consist of an in principle open-ended sequence of inferences, each of them justified by the conclusion of the next one? Even once we clear up the empirical facts, these hard questions are bound to remain.

3.5 Induction and Repetition

The worries about circularity, or ungroundedness, that Hume raises have been considerably discussed in the subsequent literature. But Hume himself puts just as much weight on a second argument that has not received nearly as much attention. Now that's probably because it isn't a very good argument. But I think it's actually kind of interesting to see why it fails, and since it was such an important argument to Hume we should address it. Here's one passage where Hume makes the argument I want to discuss.

> From causes which appear similar we expect similar effects. This is the sum of all our experimental conclusions. Now it seems evident that, if this conclusion were formed by reason, it would be as perfect at first, and upon one instance, as after ever so long a course of experience. But the case is far otherwise. Nothing so like as eggs; yet no one, on account of this appearing similarity, expects the same taste and relish in all of them. It is only after a long course of uniform experiments in any kind, that we attain a firm reliance and security with regard to a particular event. Now where is that process of reasoning which, from one instance, draws a conclusion, so different from that which it infers from a hundred instances that are nowise

different from that single one? This question I propose as much for
the sake of information, as with an intention of raising difficulties.
I cannot find, I cannot imagine any such reasoning. But I keep my
mind still open to instruction, if any one will vouchsafe to bestow it
on me (Hume, 1748/2004, 27).

The argument starts with the 'observation' that we feel more confident in
an inductive inference based on a large number of positive instances than one
based on just a few instances. Hume then argues that the extra instances can't
be any better evidence for our conclusion, since merely repeating evidence does
not make it any stronger. So he denies that any form of reasoning could have the
same patterns as our inductive inferences. Now there are some situations where
we do change in response to mere repetition. They are cases of what we'd now
call conditioning, or what Hume would call the acquisition of a custom or habit.
So induction does not look like a reasoning process, it looks like something less
cognitive.

I think the best response here is simply to challenge the premise. It isn't true
that when we collect more inductive evidence, what we have is sheer repetition.
Indeed, it isn't even clear that this is conceptually coherent. Let's say we're trying
to figure out whether all Fs are Gs. We find one F, call it x_1, and see that it is
G. Then we find another F, call it x_2, and see that it is also G. Could these be
exactly the same kind of evidence? It seems they could not. For one thing, one
of the data points is identical to x_1, and the other is not. One of them was first
observed by us first, and the other was first observed by us second. There are
bound to be many respects in which x_1 and x_2 differ. Now maybe they will turn
out to be insignificant, but saying that they are insignificant is a major claim, that
it isn't clear that the sceptic has reason to insist we accept.

In practice, the different Fs we see will usually have a lot of variability within
them. They won't vary with respect to whether they are G, at least not if the hy-
pothesis we are testing is actually true, but they will vary with respect to many
other properties. And we usually think that this is very important. In any text-
book treatment of enumerative induction, it will be insisted that a more diverse
sample is a stronger ground for inductive inference than a less diverse sample. It's
interesting to think about why this could be so. There are two plausible expla-
nations that I know of that could help here. Both of them, interesting, help us
respond to Hume's challenge.

In his *Treatise on Probability*, Keynes (1921) argues that the fundamental form
of inductive reasoning is reasoning by analogy. We should, think Keynes, reduce
enumerative induction to reasoning by analogy. The advantage, according to
Keynes, of a diverse sample is that it increases the probability that each of the

unobserved *F*s will resemble one of the observed *F*s. On Keynes's picture, mere repetition of the data would not increase the support that the evidence provides to the conclusion. If x_1 and x_2 are alike in all salient respects, then observing both of them will not increase the probability that an unobserved *F* is analogous to an observed *F*. So Hume's premise about repetition not being evidentially significant is fully accepted, but his broader conclusion about enumerative induction is not.

In a recent paper, Roger White (2005) has argued that we should ground enumerative induction in inference to the best explanation. If we only observe *F*s that are *G*s, that's a striking fact that needs to be explained. One candidate explanation is that all of the *F*s are in fact *G*. As the sample of observed *F*s gets larger and larger, competing explanations become less and less tenable. So again, increasing sample diversity increases the support that the observations provide to the conclusion. It isn't quite as clear on White's view as on Keynes's view hat we need variety in the sample to increase the support the observations provide to the conclusion. But it is consistent with White's view that they do not. So Hume just seems to be wrong about the nature of inductive inference.

It's also worth noting that, contrary to what Hume says, induction based on a single case is not invariably a bad inference. If I hear that one showing of a particular movie ends with the hero being killed, I'll infer that all the other showings will end the same way. If some physicists found that the mass of a particular electron is exactly m, I'd infer that all other electrons have mass m. Clearly there are a lot of cases where we don't make inferences from a single case, but in those cases, diversity of the observed sample becomes more important. So this argument of Hume's seems to be easy for supporters of induction to reply to.

3.6   Induction and Infants

There is a puzzling discussion at the end of §4 of the *Enquiry* about inductive reasoning in infants.

> It is certain that the most ignorant and stupid peasants-nay infants, nay even brute beasts-improve by experience, and learn the qualities of natural objects, by observing the effects which result from them. When a child has felt the sensation of pain from touching the flame of a candle, he will be careful not to put his hand near any candle; but will expect a similar effect from a cause which is similar in its sensible qualities and appearance. If you assert, therefore, that the understanding of the child is led into this conclusion by any process of argument or ratiocination, I may justly require you to produce

that argument; nor have you any pretence to refuse so equitable a demand. You cannot say that the argument is abstruse, and may possibly escape your enquiry; since you confess that it is obvious to the capacity of a mere infant. If you hesitate, therefore, a moment, or if, after reflection, you produce any intricate or profound argument, you, in a manner, give up the question, and confess that it is not reasoning which engages us to suppose the past resembling the future, and to expect similar effects from causes which are, to appearance, similar (Hume, 1748/2004, 29).

Hume seems to reason as follows.

1. Since no one has been able to come up with the intermediate steps that support inductive reasoning, whatever reasoning process we go through in induction cannot be simple.
2. Since young children can reason inductively, whatever reasoning process we go through in induction cannot be complex.
3. So, induction cannot be a reasoning process, since it cannot be either a simple process or a complex process.

We could question both premises here. In particular, a quick look at our best syntactic theories suggests that young children are in fact very good at following the complex rules that constitute the difference between well-formed and ill-formed strings. So perhaps there are some complex pieces of reasoning that we are going through, but which we should not expect to be any easier to state than the rules of syntax for our language.

But even if we grant the argument, it's not clear what the conclusion tells us. Let's make sure we keep clear the distinction between the person making an inductive inference, and the theorist evaluating that inference. Even if the person making the inference does not go through a particularly complicated bit of reasoning, it might be that what they do is something the theorist can applaud, and the theorist's reason for applauding might in turn be a complicated bit of reasoning. So Hume's point about complexity seems to pretty clearly miss the mark.

That is, it misses the mark if Hume's point is to argue something about the normative status of inductive practices. If Hume is simply trying to argue that as a matter of fact, our inductive practices are not a matter of reasoning through-and-through, then perhaps this argument looks stronger. But if that's Hume's point, it isn't clear that he's latched on to anything distinctive about *induction*.

We might well question, in this spirit, whether our *deductive* practices can really be based in reason all the way down. One upshot of Lewis Carroll's dialogue might be that they cannot be (Carroll, 1895). At some level in an argument, we simply have to move from a premise to a conclusion, without reasoning that the premise supports the conclusion. That's true in deductive reasoning, and it is true in inductive reasoning. It isn't obvious that Hume has found a point here that is different between induction and deduction. It's true that there are metaphysical differences between the two. In good deductive reasoning, the premise necessitates the conclusion. But it isn't clear how Hume, or anyone else, can turn that metaphysical fact into an epistemological, or even psychological conclusion.

3.7 Hume's Conclusion

If we assume that Hume's argument can be rescued from obvious flaws, then two closely related questions arise:

1. What is the conclusion of the argument?
2. If the conclusion is true, what should we do?

One reaction, basically Popper's reaction, is to say that the conclusion of the argument is that induction is irrational, and we should react to the argument by not using induction any more. I'm not sure there's either textual evidence to support that as a reading of Hume, nor any reason to think that Hume has given us a good argument to that conclusion.

A more plausible reading of the argument is that induction can't be justified by reason, not that it is unjustified all-things-considered. There are a couple of ways we might go from there, both of them I think compatible with various things that Hume says.

One option is to take the conclusion of the argument to be that induction is psychologically unavoidable, even though it cannot be in principle justified. On this kind of view, induction is arational, not irrational. This doesn't mean we should stop making inductive inferences. Since we cannot, on this view, do this, it doesn't make sense to say that we should stop. We should, perhaps, stop trying to mimic our natural reasoning processes in regions outside the realm they were designed/evolved for. So it is foolish to apply canons of inductive reasoning to 'metaphysical' questions. But in general we should just learn to live with the fact that we have certain inductive habits, and not treat this as something for normative evaluation.

Just how close this is to Hume's position I can't really say. But it doesn't seem to be a particularly plausible position. We know that there are all sorts of better and worse pieces of inductive, or at least of ampliative, reasoning. There

are cases where people are too quick to infer causation from correlation. There are cases where people confuse the inability to reject the null hypothesis with the truth of the null hypothesis. There are cases where people apply formal rules too precisely, and accept any thesis that's supported by a statistically significant experimental finding, even if we should expect that we'll have merely fluky 'statistical significant' findings literally every day of the week.

We know not just that all of these mistakes are made, but that they can be corrected. And they don't just have to be corrected by a process of rehabituation. We can correct them by reasoning. Now the Humean may say that what's really unavoidable, and beyond the realm of rational assessment, is not the 'high-level' inferences just described, but something more foundational. And perhaps that's right. But it's not clear what follows from this, either for the evaluation or revision of our practices.

Perhaps then a better option would be to weaken the intended conclusion even further. What Hume really concludes is that reason cannot justify our inductive practices. But Hume thinks "Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my little finger." But this doesn't mean it is OK, all things considered, to have this preference. If you have this preference, you're a psychopath, and you shouldn't be a psychopath. There is nothing in reason alone that can show why being a psychopath is wrong, but this just shows there is more to life, and more to normativity, than reason.

On this kind of reading, our Custom and Habit are taken to have normative significance. There are actually a few different ways that this could happen. One way is that we might adopt a kind of normative relativism. Induction is good because it involves conformity to *our* customs and habits, and that's a good thing to do if you're one of us. Another way is that we might observe that our customs and habits seem to be working, and that as long as our customs keep working, we are justified in following them.

It might be thought that this involves an objectionable kind of inductive presupposition. On this picture, our justification for inductive practices rests on the success of our practices. But without a theory of induction, how can we have reason to believe that the practice will continue to succeed? So the practice cannot be justified by its successes. This is to assume too 'internalist' a picture of justification. A practice may be justified by its actual success, even if we had no reason, or even no justification, to believe that it would succeed when we adopted the practice. Compare what Hume says about political authority:

> Princes often seem to acquire a right from their successors, as well
> as from their ancestors; and a king, who during his life-time might
> justly be deemed a usurper, will be regarded by posterity as a lawful

prince, becase he has had the good fortune to settle his family on the throne, and entirely change the antient form of government. Julius Caeser is regarded as the first Roman emperor; while Sylla and Marius, whose titles were really the same as his, are treated as tyrants and usurpers. Time and custom give authority to all forms of government, and all successions of princes; and tha power, which at first was founded only on injustice and violence, becomes in time legal and obligatory. Nor does the mind rest there; but returning back upon its footsteps, transfers to their predecessors and ancestors that right, which it naturally ascribes to the posterity, as being related together, and united in the imagination. The present king of France makes Hugh Capet a more lawful prince than Cromwell; as the established liberty of the Dutch is no inconsiderable apology for their obstinate resistance to Philip the second. (*Treatise*, Book III, Part II, Section X.)

In a similar vein, we might think that a community which adopts a set of customs, can be evaluated retrospectively for its successes and failures. And if it has succeeded, that gives a legitimacy to the practices it was engaged it.

The practical upshot of this way of looking at Hume's argument is a kind of anti-revolutionary view. (But wasn't Hume a kind of revolutionary? Only of a sort - he wanted to return to more well established practices.) If we thought that the best inductive practices could be established by pure reason, then we would be justified in throwing out all of our existing practices and replacing them with this ideal. But in reality no such discovery is possible, so no such epistemological revolution can be justified. Any reforms we make to our practices are ultimately justified by other parts of our practices. That suggests we should be always ready to reform our practices, particularly if the practices do not live up (our own) standards of coherence. But, just as in politics, we should be reformists not revolutionaries.

This is, in a way, a deeply *anti-sceptical* reading of Hume. He thinks that not only are inductive inferences justified, but they are justified in a way that is resistant to root-and-branch defeat. To be sure, the kind of justification we end up with is perhaps not what we wanted to start with. Perhaps we wanted justification in terms of pure reason alone, and we can't have that. We can't have it for our political systems, and we can't have it for our inductive practices. But the result is not a kind of political scepticism, which denies that the existing systems are legitimate, nor a kind of inductive scepticism, which denies that the existing inferential practices are legitimate.

Chapter 4

Formulations of Sceptical Arguments

4.1 Brueckner on Underdetermination

In his "The Structure of the Skeptical Argument", Anthony Brueckner (1994) argues that the underdetermination argument for scepticism is more powerful than the closure based argument. In particular, he argues that the closure based arguments tacitly assume an underdetermination principle, and that principle is strong enough to generate a sceptical conclusion on its own. Let's start with the simple closure argument that he begins with, and follow roughly his naming conventions in calling it **Argument A**. In these arguments, p is a proposition we ordinarily take ourselves to know, and SK is a sceptical counterpossibility to p.

> **Argument A**
>
> 1. If I know that p, then I know that $\neg SK$.
> 2. If do not know that $\neg SK$.
>
> ———————————
>
> C. I do not know that p.

Both premises stand in need of some justification. The first premise is a particular instance of a closure premise. We can support it intuitively by imagining someone going through the deductive reasoning from p to $\neg SK$. For instance, if p is that I have hands, and SK is that I'm a handless brain in a vat, then we can imagine a thinker reasoning, "I have hands, so I'm not handless, so I'm not a handless φ for any φ, so in particular I'm not a handless brain in a vat." The person who wants to reject premise 1 has to say which is the first step that the thinker does not know.

This turns out to be surprisingly hard. Let's follow Nozick's theory of why it is that premise 1 is false (Nozick, 1981). Nozick says that it is a necessary condition on knowledge that one's belief be sensitive. A belief in A is sensitive iff one wouldn't believe A if A were false. That's not *sufficient* for knowledge, but

let's assume we're looking at propositions for which all the other conditions on knowledge are met. Then we imagine the thinker going through the following four steps.

1. I have hands.
2. I'm not handless.
3. $\forall \varphi$, I'm not a handless φ.
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, assuming we can make sense of the second-order quantifier it contains, her third belief is sensitive as well. The nearest world in which it is false is one where there is a φ such that she is a handless φ. So perhaps it is a world in which she's a handless human, or a handless self-identical thing. But if she had one of those properties, she wouldn't believe she had hands. So that belief would be sensitive. Amazingly, it's just the step of $\forall$-elimination to get from the third to the fourth step where we lose sensitivity. So the sensitivity story about why premise 1 fails would say that our thinker could lose knowledge by performing a step of $\forall$-elimination. This seems improbable, to say the least.

So perhaps we should accept that premise 1 of **Argument A** is true, and look to premise 2. Why might we accept it? Brueckner's first suggestion, on behalf of the sceptic, appeals to sensitivity. If we hold that it's only plausible to know things that are sensitively believed, then premise 2 is true. Brueckner offers two objections to this move, neither of which seem particularly compelling to me.

The first response is that this kind of defence of premise 2 undermines premise 1. The worry I guess is that since the property of being sensitively believed isn't closed under (known) entailment, then if sensitivity is needed for knowledge, then knowledge won't be so closed either. But there is no reason to think that properties of necessary conditions on knowledge are always transferred to being properties of knowledge. That is, it is consistent to say that all knowledge is sensitive, and that knowledge is closed under known entailment. If we know only necessary truths, both conjuncts will be true.

There's a more general point here about closure that is worth making. Brueckner goes on later to say "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, and yet $p \wedge \neg p$ is non-contingent, while p is contingent. 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.

The other point Brueckner makes about sensitivity and premise 2 concerns intuitions. He wants to dispute the claim that premise 2 is intuitive because our belief that $\neg SK$ is insensitive. The argument here is a little tricky to follow. He says, roughly, that even if we were disposed to have fewer beliefs were we brains in vats, the sceptical intuition would still be strong. I'm not sure about that, because I'm not sure what it would be like to be the kind of person who forms beliefs if they are in a normal world, but not if they are in a sceptical scenario. Consider, in particular, the sceptical scenario Hume worries about, where hitherto reliable correlations break down. What would the mental life of someone who is disposed to form beliefs about the future iff they are not in such a world be like? It's hard to imagine they are anything like us, since our dispositions are not typically triggered by future events in anything like this way. So it's hard to know what to make of this thought experiment.

After dismissing sensitivity as an explanation for premise 2, Brueckner suggests we explain the premise in terms of justification. The agent doesn't know $\neg SK$ because she isn't justified in believing $\neg SK$. And this is to be derived from more general principles about justification. Here is Brueckner's 'underdetermination principle'.

(UP) $\forall S, \varphi, \psi$, if S's evidence for believing that φ does not favour φ over some incompatible hypothesis ψ, then S lacks justification for believing that φ.

Now to get from there to a sceptical argument, we need some reason for saying that one's evidence does not favour p over $\neg SK$. Brueckner eventually says (in the last paragraph of the paper) that the sceptic's argument for this is infallibilism about support, and that this is implausible. So Brueckner ultimately does not think that any sceptical argument based on **(UP)** will work. But first, let's see how he thinks the sceptic can try to use **(UP)**. After fussing a little with the right statement of the closure principle, he offers the sceptic this argument.

Argument B

1. If I have justification for believing that p, then I have justification for believing that $\neg SK$.
2. If my evidence for believing that $\neg SK$ does not favour $\neg SK$ over SK, then I lack justification for believing that $\neg SK$.
3. My evidence for believing that $\neg SK$ (my sensory evidence) does not favour $\neg SK$ over SK.
4. If I lack justification for believing that p, then I do not know that p.

_____

C. I do not know p

Note that Brueckner's own formulation of the argument includes a couple of extra steps that are derived from these premises, and takes what I've labelled as premise 4 to be clear enough to go without stating, but I don't think this changes the structure of the argument.

The key point that gets made next is that both of the first two premises are specific instances of more general principles. And these principles can be stated as conditionals with a common antecedent. One of them is a reformulated version of **(UP)**, the other is a closure principle, which is called **(C1J″)**. (The numbering comes from some variants that we've skipped over. I assume S is the 'one' in the principles!)

(UP′) $\forall S, \varphi, \psi$, if one has justification for believing that φ, and φ and ψ are incompatible, then one's evidence for believing that φ favours φ over ψ.

(C1J″) $\forall S, \varphi, \psi$, if one has justification for believing that φ, and φ and ψ are incompatible, then one has justification for believing that $\neg \psi$.

Brueckner goes on to make two claims about these principles which are somewhat in tension. First, he wants to argue that the two claims are equivalent, or at least are equivalent in a large range of cases. Second, he wants to argue that the sceptical argument can be stated without appeal to a closure principle, and with only appeal to the underdetermination principle. The tension should be fairly obvious. If the sceptic doesn't need a closure principle, but instead only needs something equivalent to a closure principle, that doesn't really move the ball far down the field. It's true that sometimes replacing a premise with a provably equivalent premise can make the argument intuitively stronger, because the new premise might be intuitive. (Intuiveness is *not* closed under logical entailment, or even under logical equivalence, as a quick glimpse at the list of statements equivalent to the axiom of choice will confirm. As the joke goes, "The Axiom

of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?") But it's not like the closure principle was the *unintuitive* part of the sceptical argument. Perhaps it is best to look at the two points in turn without worrying too much about the relationship between them.

What Brueckner first wants to prove is that when S's justification for φ is evidentially based (or perhaps evidentially justified), and φ and ψ are incompatible, the following biconditional is true.

- S's evidence for believing that φ favours φ over ψ iff S has justification for believing $\neg\psi$.

The right-to-left direction is reasonably clear. I'm not sure exactly how to make sense of the 'favouring' talk, but it's very plausible that if evidence justifies believing $\varphi \wedge \neg\psi$, it favours φ over ψ. But the left-to-right direction is harder to follow. Brueckner says, quite naturally, that the left-hand-side of the biconditional is consistent with S not having strong enough evidence to believe either φ or ψ, or for that matter $\neg\psi$. But he says that when S is also justified in believing φ, as is assumed here, that possibility vanishes. As Stewart Cohen (1998) points out, it isn't at all clear how this is supposed to help. It seems still possible that S could be justified in believing φ, but in staying neutral on ψ. If a theorist thought for whatever reason that sensitivity was necessary for justification, then it seems they could accept (**UP′**) and reject (**C1J″**).

Of course, if this biconditional fails, it means that it is more pressing to figure out whether an underdetermination based argument for scepticism can be made that does not appeal to a more general closure principle. Here is Brueckner's attempt to do just that.

Argument C

1. If my evidence for believing that p does not favour p over SK, then I lack justification for believing that p.
2. My evidence for believing that p (my sensory evidence) does not favour p over SK.
3. If I lack justification for believing that p, I do not know that p.

C. I do not know that p.

The crucial idea here is that if we can apply the underdetermination principle directly to p and $\neg SK$, rather than applying it to SK and $\neg SK$. That means we can argue directly that S lacks justification for p.

We'll look more closely at possible responses to this argument when we turn to the discussion by Stewart Cohen (1998) of the relationship between the arguments Brueckner discusses, but let's first quickly note how one might object to each of the premises.

The first premise looks hard to dispute. Brueckner seemed to be on shaky ice when he said that our evidence for propositions like SK was sensory. It is plausible that we don't have anything naturally described as 'evidence' for propositions like SK, and for other 'hinge' propositions. So our evidence does not, and need not, favour it over in order for us to be justified in believing it. But when it comes to propositions like p, which are meant to be ordinary everyday propositions, it looks plausible. If our evidence doesn't favour them over absurd claims like SK, we aren't justified in believing them.

The second premise is, as Brueckner states, hard to motivate. If the motivation is just that our evidence does not entail $\neg SK$, then it is implausible. But this probably understates the strength of the sceptic's position, for reasons that we'll frequently return to.

The third premise is hard to deny, much like the first. Even if one wants to deny, as Crispin Wright (2004) does, that we have 'justification' for believing Wittgensteinian hinges, it is hard to see how I could know an everyday proposition, which is what p is assumed to be, without having justification for it.

So it will really all turn on the second premise here. And the second premise doesn't look like a closure principle. So there is some *prima facie* plausibility to the claim that the sceptic doesn't need a closure principle for their argument to work.

4.2   Cohen on Closure

In "Two Kinds of Skeptical Argument", Stewart Cohen (1998) argued that Brueckner understated the force of closure-based sceptical arguments. He thinks that there is a closure-based sceptical argument that is better for the sceptic than the argument based on underdetermination that Brueckner develops. I think I'm basically on Cohen's side of this dispute, though perhaps not quite for the same reasons. Here is the argument that Cohen wants to develop, slightly modified for consistency with the terminology here.

Argument D

1. If my evidence justifies p, then my evidence justifies $\neg SK$.
2. My evidence does not justify $\neg SK$.
3. If my evidence does not justify p, then I do not know p.

—————————————

C. I do not know p.

Note that the closure principle being appealed to here is implausible if read in a certain, reasonably natural, way. Assume again that I have a justified belief in p, and consider the following instance of a closure principle.

- If my evidence justifies p, then my evidence justifies $p \vee \neg p$.

If p is a regular proposition about the world, then it is plausible that the antecedent is true and the consequent false here. I'm justified alright in believing $p \vee \neg p$, but not on the basis of my evidence. Rather, I'm justified in believing it on the basis of purely *a priori* considerations. I take it this is an uncharitable, and unintended, reading of the closure principle. What matters is whether someone in my evidential situation, broadly construed, is justified in believing $p \vee \neg p$, not whether the evidence, in any sense, is the source of this justification.

This distinction matters a lot for Cohen's preferred argument for premise 2 of **Argument D**. He appeals to the following principle, for some reason called **(Z)**.

(Z) $\forall S, \varphi$, if the truth of φ would explain S's evidence, then S's evidence does not justify $\neg\varphi$.

I think the plausibility of **(Z)** turns in part on equivocating between these two readings of 'justify'. If we read E would justify A as meaning that a belief in A is justified on the basis of E, then perhaps the principle is plausible, but of course on that reading it doesn't help support premise 2, at least as premise 2 is supposed to be interpreted. On the other hand, if we read E would justify A as meaning that a person with evidence E is justified in believing A, then the principle is obviously false. It fails whenever φ is inconsistent, as Cohen appears to note.

Now this seems to matter for the argument that Cohen offers. He says that we should at this point restrict **(Z)** to non-contradictions. But why should this be a plausible restriction? If we thought that some non-tautologies could be justified without an evidential base, the way that tautologies can be, then **(Z)** would fail for just the same reason.

More generally, we might worry that **(Z)** just is a form of infallibilism about justification. Let's say that S can justifiably believe H on the basis of E, when E does not entail H. Since E doesn't entail H, then $E \wedge \neg H$ is not a contradiction. Now $E \wedge \neg H$ seems to explain S's evidence; indeed it entails it! So S's evidence doesn't justify $\neg(E \wedge \neg H)$, at least if **(Z)** is true. So either S's evidence doesn't justify H, as we supposed that it did, or S is not in a position to make a simple move of $\vee$-introduction and a DeMorgan transformation to get from H to $\neg(E \wedge \neg H)$. Neither option seems attractive.

So I'm not convinced that the direct closure based argument that Cohen offers is a particularly strong sceptical argument. But let's step back from that and look at the relationship between the two arguments. As Cohen notes, **Argument C** and **Argument D** have some steps in common. Each of them derive the unjustifiedness of a belief in p. Let's isolate the steps in those arguments, and use Cohen's labels for the restricted arguments.

(UND)

1. If my evidence does not favour p over SK, then my evidence does not justify p.
2. My evidence does not favour p over SK.

   C. My evidence does not justify p.

(DC)

1. If my evidence justifies p, then my evidence justifies $\neg SK$.
2. My evidence does not justify $\neg SK$.

   C. My evidence does not justify p.

Now Cohen notes that the premises of **(UND)** together entail premise 1 of **(DC)**. And premise 2 of **(UND)** arguably entails premise 2 of **(DC)**. So, says Cohen, it is impossible to show that one of the premises of **(DC)** is false without showing that one of the premises of **(UND)** is false. But it is possible to show that one of the premises of **(UND)** is false without showing that one of the premises of **(DC)** is false. So, in a sense, **(DC)** is a stronger argument than **(UND)** since less can refute it.

I think there are a couple of issues here about what it takes to respond to, or as Cohen says, 'refute' an argument. In a footnote, Cohen says that refuting an argument 'entails showing that one of its premises is false.' (Cohen, 1998, 153n21) There's a crucial scope ambiguity here. Cohen could be saying one of the following two things.

- An argument is refuted if it is shown that there is at least one false premise in the argument.
- An argument is refuted if it has a premise which is shown to be false

If the first of these principles is true, then Cohen is right that any refutation of **(DC)** will be a refutation of **(UND)**. But that's not a very plausible account of what it is to refute an argument. Here's my refutation of the Sorites argument: I'm not bald, so any argument that entails I am bald has a false premise! On the other hand, if we take the second disambiguation, then it won't be true that any refutation of **(DC)** is a refutation of **(UND)**. An argument showing that premise 1 of **(DC)** does not show that either disjunct of **(UND)** is false, though it does show that at least one of them is false.

Jonathan Vogel (2004) makes this point with a nice analogy, which I'll slightly modify. Consider this argument for scepticism, call it **(NC)**.

(NC)

1. If my evidence justifies p, then my evidence entails that p is a theorem of first-order logic.
2. My evidence does not entail that p is a theorem of first-order logic.

_____

C. My evidence does not justify p.

The two premises of **(UND)** together entail the first premise of **(NC)** by showing that its antecedent is false. And the second premise of **(UND)** entails, at least with some overwhelmingly plausible assumptions, that the second premise of **(NC)** is true. So in a sense any refutation of **(NC)** will be a refutation of **(UND)**. But in practice this isn't really true. We can refute **(NC)** by noting that its first premise is absurd, and this doesn't help at all with responding to **(UND)**.

That's not to say that **(NC)** and **(DC)** are alike in all respects. In one crucial respect they are very different. The first premise of **(DC)** is very plausible, while the first premise of **(NC)** is absurd. But the two arguments are alike in the respect that Cohen highlighted, and that suffices to show that his argument does not show that **(DC)** is a stronger sceptical argument.

4.3 Byrne on Priority and on Evidence

We'll close this week by looking quickly at two scpetical arguments that Alex Byrne (2004) considers. He finds both of them to be bad for fairly straightforward reasons, though we might want to spend a little more time on each. We'll look at them in reverse order to Byrne, so we'll start with the priority arguments. Byrne goes through a few different arguments; here are two.

> **Priority Sceptical Argument - Circularity Variant**
>
> 1. In order for *S* to know any empirical claim, she must **antecedently** know that she is not in a sceptical scenario where her empirical methods of inquiry fail.
> 2. Whether *S* is in a sceptical scenario where her empirical methods of inquiry fail is an empirical question.
> 3. It is impossible for *S* to know that she is not in a sceptical scenario where her empirical methods of inquiry fail antecedently to knowing that she is not in a sceptical scenario where her empirical methods of inquiry fail.
>   _____
>
> C. *S* does not know any empirical claim.

The third premise seems like a platitude about what we mean by antecedent knowledge. The second premise looks reasonably plausible as well. If by 'empirical' we mean something that can't be decided *a priori*, then perhaps this question isn't 'empirical', since arguably it is knowable *a priori* that we are not in sceptical scenarios. But this looks like a poor definition of 'empirical'; on a better definition, premise 2 looks like it should work.

So all of the work is being done by premise 1. And it doesn't look very plausible. Note that it immediately entails that anyone who lacks the cognitive capacity to think about sceptical scenarios, such as a very young child, does not know any empirical claims. But this just looks clearly false, since intuitvely children can know a lot before they can think about sceptical claims.

Perhaps it was wrong to think about something as complex as a sceptical scenario in the argument, since what's really relevant is whether we know that the capacities we are actually using are reliable. This suggests a kind of regress argument, of a kind that Hume was particularly concerned with.

Priority Sceptical Argument - Regress Variant

1. In order for *S* to know any empirical claim, she must **antecedently** know that the methods she used in forming that belief were reliable.
2. It is an empirical question whether any method used in forming a belief about an empirical question is reliable.
3. The relation 'is antecedently known' is acyclic.
4. *S* has finitely many beliefs.

C. *S* does not know any empirical claim.

The validity of this argument may not be completely obvious, and for this reason I'm a little worried about whether this is the very best way to capture regress reasoning. But the idea is that for any empirical belief that *S* has, if this is to amount to knowledge, there must be a distinct belief she has about the method she used to form that belief, and this belief must itself amount to knowledge. So for any belief she has, there is a belief that stands in the 'antecedently known to' relation to it. Now it is a general fact about models that if there is a relation *R* defined on a model such that $\forall x \exists y : xRy$, then either *R* has a circle in it somewhere, i.e., a chain $x_1, ..., x_n$ such that $\forall i < n : x_i R x_{i+1}$, and $x_n R x_1$, or the model is infinite. But by premises 3 and 4, neither of these options are possible. So it isn't true that *S* knows anything empirical.

Premises 2 and 3 seem more or less definitional. But premises 1 and 4 are questionable. Premise 4 relies on having a way of counting beliefs that we don't obviously possess. And premise 1 seems false for the same kind of reasons that premise 1 of the 'circularity variant' seems false. A young child can, intuitively, gain knowledge before she has the notion of a reliable belief-forming method. So it isn't true that this knowledge of reliability must be antecedent to empirical knowledge.

We'll say much more about this issue when we get to James Pryor's discussion of dogmatism (Pryor, 2000), so for now we'll just lean on the 'young children have knowledge' point, which does seem to be Byrne's central objection to this argument.

Byrne says a little more about the underdetermination argument, but we'll pass over it reasonably quickly because it is the focus of next week's discussion. His main point is that it isn't at all clear that when it comes to external world scepticism, that our evidence actually does underdetermine our conclusions in a way that is helpful for the sceptic. His point is that external world scepticism typically relies on a picture of perception that is not widely shared by people who work on perception. In particular, it assumes a picture of perception on which

we have the same perceptual states as someone who is radically deceived. But that isn't obviously correct; if perception is at all 'wide', then we have different perceptions to a brain in a vat.

Now there are a few of distinct questions here that could do with answering. One is whether Byrne is right about perception. But that's a question for a whole different course. Another is what the connection is between perception and evidence. Perhaps even if we have different perceptions to someone in a sceptical scenario, we have the same evidence. Byrne discusses that question briefly, but we'll look at it in more detail when we get to Williamson. And finally, there's the question of whether this kind of move, noting that our evidence would be different if we were in a sceptical scenario, can respond to all sceptical arguments. Byrne does not claim it does. He explicitly notes that he isn't responding to other minds scepticism, for instance. But Williamson is more ambitious. He thinks any sceptical argument falls to this kind of move. So we'll turn next to Williamson's arguments about externalism about evidence, and how they relate to scepticism.

Chapter 5

Williamson on Evidence

5.1 Discrimination Arguments

We've so far been focussing on underdetermination arguments for scepticism. Timothy Williamson (2000b) starts with a subtly different kind of argument, an argument for scepticism from indiscriminability. Here's how we might state this argument.

Indiscriminability Argument

1. For S to know that she is not in a sceptical scenario, she must be able to discriminate the sceptical situation from the actual scenario.
2. S cannot discriminate her 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_1 and not o_2. But if we individuate situations finely, as I think we should, this will lead to some odd results. 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_1 be the actual situation, and o_2 be a situation where I've been in London for the past 6 weeks. Intuitively I can discriminate o_1 from o_2; if I had been in London I would know it! But I can't know that I'm in o_1, since o_1 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_1 and not o_2. 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. If we make 'can know' factive, then consider the case where o_1 is a scenario where I've been in London for the past 6 weeks, and o_2 a scenario where I've been in Barcelona 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 London and Barcelona 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 in the Tate Modern. Indeed, I was there not long ago and it wasn't too hard to tell I was there. That's (one of) the ways in which I can discriminate being in London from being in Barcelona. So those situations are discriminable.

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. (This is of course just the old 'speckled hen problem' (Chisholm, 1942), translated to a setting I have clearer intuitions about.) But imagine that my 'margin of error' in detecting building heights is 5 stories, and the building is 72 stories tall. Then I would know that it 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.

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 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 dead from being alive, but she can't discriminate being alive from being dead. (I'm not sure whether it better tracks the English usage to define things this way or with o_1 and o_2 reversed. The key point here is that the relation is asymmetric.)

But perhaps that's a feature of an oddly quirky example. What we want to say is that the agent's capacities for discrimination are epistemic *dispositions* of the agent. And we know that we shouldn't analyse disposition claims in terms of counterfactuals, since some dispositions are *finkish* (Lewis, 1997). That is, something can be disposed to φ in circumstances C, even if were C to obtain, it wouldn't φ since it would have lost that disposition, or perhaps the disposition would have been 'masked'. Maybe we can say the same thing about discrimination. S is disposed to be able to tell whether she's dead or alive, even though, were she to be dead, she wouldn't know she were dead, because she would have lost that capacity.

If there were a nice analysis of 'finkishness' to appeal to at this point, we could perhaps strengthen the sceptic's argument by noting that dispositions to discriminate scenarios are symmetric, even if counterfactuals about discrimination are not. But I think here the sceptic faces a very uncomfortable dilemma.

If she says that discriminatory capacities are finkish because they would be lost were we to be in scenarios where we aren't as good at knowing stuff, then she loses her best argument for saying that S can't tell the real world from the sceptical scenario. What's clearly true is that if S were in the sceptical scenario, she wouldn't know that she wasn't in the real world. But it's also clearly true that in the sceptical scenario, S has radically lowered epistemic capacities. So arguably this counterfactual is false not because she lacks the relevant capacity, but because the capacity is finkish, and were it to be called on, it would be lost. In other words, if we make these capacities *too* finkish, the best argument for premise 2 would fail.

But if we don't make capacities that finkish, it seems that premise 3 is bound to fail. If it is possible that not all changes in capacities make for changes in the dispositions that are relevant to what one can discriminate, then it seems that it could be possible that one could be able to discriminate a situation in which one had lost a capacity from the actual situation in which one has that capacity, but not vice versa. At the very least, the sceptic needs a good argument that such a pair of scenarios cannot arise. The anti-sceptic thinks that there is a clear candidate pair: the actual world and the sceptical scenario. And the sceptic needs some argument to show that the anti-sceptic is wrong.

5.2 Williamson on Evidence and Knowability

Williamson 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 should be familiar from his general work on attacking 'luminosity' assumptions (Williamson, 1996). Assume, for *reductio*, the general princple that agents know exactly which properties their evidence has. 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. That is presumably motivated by general principles about knowledge and indiscriminability, plus a general view that there's not much we can perfectly discriminate. Then consider a sequence of similar cases, starting with one where one's evidence is clearly F, and ending where it is clearly $\neg F$. 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 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.

5.3   Hawthorne on Coziness

There are a few ways in which we might try to help out the sceptic here. The reasoning Williamson attributes to the sceptic goes as follows:

1. Everyone knows exactly which properties their evidence has.
2. So if agents have different evidence in sceptical and normal scenarios, then the victim in the sceptical scenario knows that her evidence has properties that it would not have in the normal situation.
3. So agents have the same evidence in sceptical and normal scenarios.
4. So we have very little external world knowledge.

The last inferential step is dubious, though many non-sceptical philosophers (including Williamson) accept it. The main thing to note is that the reverse inference, from 4 back up to 1, is clearly absurd. Step 1 isn't the only reason to believe step 2, and step 2 isn't the only reason to believe step 3, and step 3 isn't the only reason to believe step 4, though perhaps it is difficult to motivate step 4 in a theory that denies step 3.

Looked at this way, it seems Williamson's offer of an argument to the sceptic is one that they should decline. The argument Williamson offers includes premises that are much stronger than needed to support the steps they motivate in the argument. Weakening the premises dramatically could preserve the plausibility of the premises without reducing their power to support sceptical conclusions.

John Hawthorne (2005) suggests, in effect, a variant on step 1 that is immune to Williamson's objections, but which is strong enough to support step 2. He suggests the following principle:

- If an agent's evidence is **determinately** F, then the agent knows that her evidence is F.

This principle is immune to the actual objection Williamson offers, but as Williamson in effect notes, it seems that it could be refuted with a very similar argument. Consider a case where the possible bodies of evidence I could have differ in small but discrete ways. One such case is the famous speckled hen. The following principle seems reasonably plausible:

- If I know that evidence I have is as of looking at a hen with at least n speckles on her facing side, then evidence I have is as of looking at a hen with at least $n + 1$ speckles on her facing side.

Now it isn't at all clear what the difference is between my evidence being as of looking at a hen with at least n speckles on her facing side, and being determinately as of looking at a hen with at least n speckles on her facing side. Assuming the speckles are not themselves vague objects, then a hen has n speckles iff she determinately has n speckles. So I think we could rerun Williamson's argument using a sequence of speckled hens to refute Hawthorne's suggestion.

But the kind of recipe Hawthorne suggests might be able to generate other kinds of objections. This principle does look immune to Williamsonian refutation, and not just because it is incredibly vague.

- If an agent's evidence is F, then she knows that it is roughly F.

Now we might reason as follows, acknowledging that it is more than a little informal. On an externalist picture of evidence, the evidence that agents have in the normal and sceptical scenarios isn't even close to each other. So if an agent in the sceptical scenario knows even roughly what her evidence is like, she'll be able to deduce that she's not in the normal scenario. And then the sceptical argument that Williamson offers will run. On an internalist picture of evidence, that reasoning won't fly, but on such a picture we'll be able to argue directly that the agents in the normal and sceptical scenarios have the same evidence, or at least close to the same evidence. So we don't need a stronger knowability principle than this one, or so it might seem.

I suspect that if one tries to turn that into a rigourous argument, one will run into deep problems about just what exactly is meant by 'roughly'. Let's say

we try and make 'roughly' into an exact measurement of some kind. Then it won't be plausible that epistemically impoverished agents know, even roughly, what their evidence is. And the agent in the sceptical scenario is certainly epistemically impoverished. On the other hand, let's make roughly into a kind of circumstances-dependent concept. Then it won't be obvious that, from the perspective of the agent in the sceptical scenario, her situation and that of the person in the normal scenario are not even roughly equivalent. Relative to her, impoverished, position, lots of things are roughly the same.

This is hardly a knock-down argument that no such cleaned up principle can be made to do the work the sceptic needs to support step 2. But I don't think we've quite found the right principle to replace step 1 with just yet.

5.4 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. Hawthorne notes that there is a more viable opponent, 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 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.

Williamson's own view is that one's evidence is what one knows. We'll look at this in more detail next week, but let's note a few cases where this provides different answers to Hawthorne's preferred perceptual model of evidence. 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 presumably this belief is 'Gettiered', and does not constitute knowledge.

In each case, it is intuitive that the 'A' and 'B' worlds have agents with the same evidence. That is something that a perceptual model of evidence would suggest, but a more radically externalist view of evidence, such as Williamson's, would reject.

Now you might wonder why this is relevant. After all, on the perceptual view of evidence, agents in this world have different evidence to brains in vats. So the perceptual model of evidence is no help to the sceptic, or so you might think. This attitude is right so long as we are considering the most radical kinds of sceptic, such as the external world sceptic. It doesn't help when we are considering the inductive sceptic. A perceptual model of evidence is consistent with adopting an underdetermination argument for inductive scepticism, since on the perceptual model, whether our inductive practices keep working in the future does not affect what we are currently perceiving.

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.

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.

There is a crucial move here which seems 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.

In effect, Williamson seems to be assuming a kind of internalism here. He is assuming there is an internal connection between the pieces of evidence one has, and the strength of one's epistemic position. I don't see why anyone, let alone an externalist like Williamson, should accept that. Intuitively the same kind of evidence can be very good evidence for someone, and very poor evidence for someone else, at least as far as knowledge acquisition is concerned, simply because the world does not co-operate. The fact that a bank has unexpectedly changed its hours effects how useful our evidence is, not what evidence we have. Or so it seems to me! We'll be returning to this question in next week's notes.

5.5   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. There isn't, at least anywhere I can see in Williamson, any reason to believe that the order of explanation, and of philosophical support, is from 1 to 2, either in Descartes or in any of his followers. And I don't see why the Cartesian should accept such an interpretation of their views. 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 thing, 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, 1641/1996, 60)

Descartes is primarily making a claim about physiology here. But we can see, especially in the note that this is the best possible physiology, some normative import to his comments. 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. It isn't true, for instance, that the person who wants to deny that there is knowledge in Gettier cases owes us a theory of knowledge from which we can derive that Gettier cases are not cases of knowledge. We might suspect that the fact that inference from a false premise is involved is part of why the cases are not cases of knowledge. But the conviction that the cases are not cases of knowledge can, and should, persist through the raising of doubts over whether one can ever obtain knowledge by inference from a false premise.

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 6

Williamson on Evidence and Knowledge

6.1 The Concept of Evidence

As became clear last week, it isn't entirely clear just what a theory of evidence is meant to be a theory of. In part that could be because, as Ralph Wedgwood has stressed in some recent work (especially in a post on Certain Doubts) there is no easy translation of the English word 'evidence' into other languages. And of course it isn't a given that the right translation of the epistemologists' term 'evidence' into ordinary English is homophonic. So it would be good to say a little about what kind of thing we're looking for when we're looking for a theory of evidence.

One thing that might make our work easier is that we're ultimately looking for a theory of what it is for two people to have the same evidence. One way to work out what it is for two people to have the same evidence is to work out what evidence is, and then see whether the two people have the same of it. To slightly misquote Jerry Fodor (1998), if you want to find out whether two people have the same pumpkin, you want to find out what pumpkins are, and then work out whether two people have the same one. But this isn't the only approach we could take. We might try to identify what it is to have the same evidence directly, without going via a theory of evidence. For instance, we can say what it is for two rays to have the same direction quite easily; it is just that they are parallel. Saying this doesn't require that we identify what a direction is, or even that we believe in such things as directions. We'll keep both options open.

A natural starting point is the idea that evidence is what justifies. So two people have the same evidence iff they are justified in believing the same thing to the same degree.

The problem is that a lot of externalists might reject the idea that only evidence justifies. For instance, if two people believe that p because a thermometer says that p, but one of them is in a world where thermometers are reliable, and the other is not, then the reliabilist will say that one has a justified belief and the

other does not. But it does not seem that it is necessary for our imagined reliabilist to add that they have different evidence. The reliabilist in question may say that the reliability of thermometers is part of what justifies, but not part of the agent's evidence. At the very least, it does not seem to violate anything conceptually clear about evidence to say that the reliability of the thermometer is not part of the agent's evidence, but is part of what justifies the agent's belief.

One possible move here, one that was suggested when I discussed this material in a seminar at Rutgers, was to qualify the kind of justification that's relevant. For instance, we might say that what we care about is something like luck-independent justification. It is a little tricky though to say just what the relevant kind of luck is. Or we might replace talk of justification with talk about *responsibility*. The idea would be that two agents have the same evidence iff they are equally responsible in believing the same propositions. Perhaps, it could be argued, that our two agents are equally responsible in trusting their thermometers, even if one is justified and the other isn't. I don't think that we should rule out though the idea of an externalism about responsibility which said that different agents, even with the same evidence, could be differentially responsible if their reactions relate differently to the world.

So perhaps we should try a different option. One natural idea about evidence is that it is a kind of epistemic starting point, or an input to the cognitive process. We see this in Alvin Goldman's idea that evidence is *non-inferentially* justified belief. But we also see it in the role evidence plays in traditional epistemological theories. (Tom Donaldson made this suggestion, and offered the motivation for it.) The foundationalists thought that the foundations were the evidence, and were the starting point. The Bayesians think that the evidence, the stuff that one conditionalises on, is an exogenous input to the cognitive system. And the coherentists, who deprecate the notion of evidence, deny that there are really *starting points*.

But thinking about the coherentists suggests a restriction on the idea of evidence as inputs. Some coherentists think that involuntary perceptual beliefs are a kind of external constraint on the system. A belief isn't justified in virtue of its relation to those beliefs in particular. But as a matter of psychological necessity, those beliefs are indeed beliefs of ours, and the justification for our overall belief state comes from its coherence, which therefore includes its coherence with those beliefs. So in a sense this coherentist takes perceptual beliefs to be inputs, although they don't take them to be inputs.

One solution here is to say that evidence must not merely be an input, it must also play a justificatory role. I suggest, somewhat tentatively, that these two ideas are the core of our concept of evidence. Well-functioning cognitive agents use

evidence as inputs to their cognitive system, or at least are generally disposed to do so, and when they do so, the evidence is available to justify their subsequent beliefs. This is not to say that they use *all* their evidence; some of it might be merely potential input. And it is not to say that whatever they do with their evidence is justified; everyone makes mistakes. But if they have any evidence, that evidence is the kind of thing that is available to serve as an epistemic starting point, and to justify later conclusions.

This way of thinking about evidence is particularly congenial to me, since I think evidence is basic knowledge. And basic knowledge is both an input, and the kind of thing that can justify later inferences. There is potentially a problem here that we've defined evidence so narrowly that it makes it hard for any view other than the basic knowledge view can be correct. But I think we should think of this as a challenge for rival theories. If we say, as Williamson does, that all knowledge is among our evidence, what do we mean by that?

One answer, suggested to me by Jonathan Ichikawa, is that anything we know can be an input to some argument or other. This suggests, to me at least, a view that is like E=K in letter but not quite in spirit. We might say that the notion of evidence should really be relativised to inquiries. So something is evidence relative to an inquiry if it has the status of an input relative to that inquiry. If I'm investigating whether it would be a good idea to go skiing in Florida next summer, my knowledge that it won't snow in Florida next summer could be an input to that inquiry. So relative to that inquiry, it is part of my evidence. This suggests the real truth behind E=K is that any knowledge is evidence relative to some inquiry or other, but not that all our knowledge is in our evidence. I don't think this is quite correct, but it shows how we can in principle make E=K consistent with the idea that evidence is input that justifies.

6.2 Is Evidence Propositional?

Since only propositions are known, in the relevant sense, the only way that only knowledge can be evidence is if only propositions can be evidence. So Williamson starts his argument for E=K with an argument that only propositions can be evidence (Williamson, 2000a, Ch. 9). I think this discussion is rather too compressed. There are actually a few distinct arguments, but we'll focus here just on one of them, the argument from inference to the best explanation.

The argument at first glance seems to be that inference to the best explanation (IBE) is a good form of reasoning, and what is explained in IBE is our evidence, and the only things that are explained are propositions. So only propositions are evidence. But that argument is too quick. Why should we think that only propositions are the things to be explained?

At first Williamson seems to give a syntactic argument. Sentences of the form "Explain φ!" are ungrammatical unless φ picks out a proposition. So the things to be explained are propositions. Now it isn't clear that this is a particularly strong inference. All that the syntactic point could show is that in the philosophical theory embedded in English, the kind of thing Vic Dudman used to call 'swamp-philosophy', only propositions get explained.[1] But they do lots of things oddly in the swamp. To use Dudman's own example, it would be terrible to argue from the fact that English has a past and present tense, but no future tense, to the conclusion that there is no future, and only the present and past are real. Similarly, it would be bad to argue from a fact about the syntax of certain imperatives about explanation to the conclusion that really only propositions can be explained.

That is, it would be bad even if the syntactic premise were true. But surely it isn't true. As Williamson notes, we can say things like "Explain the bloody knife!". Now Williamson could say, and does, that this is really shorthand for something like "Explain the existence/placement/bloodiness of the bloody knife!". But we need an argument that requests to explain a *thing* are not in good order as they stand, and really are shorthand for requests to explain some proposition. And that argument better not suppose that it is propositions that get explained, because that was something that was meant to be supported by a syntactic argument, not something that fed into a syntactic argument.

There's a worry around here that we're not considering enough possibilities for what kind of thing evidence could be. When we look at sentences like "Explain the bloody knife!", it is natural to think that evidence could consist of *things* like the bloody knife. Perhaps that isn't right, and the evidence is something more structured. But even if that's so, it wouldn't show that evidence is *propositional*. It might be that the evidence consists of *events*. Indeed, the link between explanation and evidence that Williamson brings up suggests just this option. Explanation is naturally thought of as a causal notion, and causation is naturally thought of as a relation between events. So perhaps our evidence consists of events that we're involved with in the right way.

We might wonder what the difference is between an event view of evidence and a proposition view of evidence. After all, for every event there is the proposition that that event happens. So even if someone had a great argument that all evidence was events, we could just say that our evidence$_p$ consisted of all and only the propositions that say those very events happened, and then say with some

---

[1] "No doubt the metaphysics enshrined in English is a metaphysics of the swamp, awaiting illumination and even revision by great minds. None the less it remains a legitimate object of investigation in its own right. We can ask, for instance, how *do* past and future differ in primeval metaphysics?" (Dudman, 1991, 271)

plausibility that whether we identified the epistemologically significant evidence with the events, i.e., the 'evidence', or the propositions, i.e., the 'evidence$_p$', was a purely terminological question with no significance.

I suspect that's a bit quick, for two reasons. First, Williamson thinks that not only is evidence propositional, but that in principle any kind of proposition could be part of one's evidence. A theory that restricted our evidence to propositions of the form 'That event happened' would be a very different theory. Second, the propositional view of evidence is naturally paired with a theory of evidential justification which entails that if p is part of one's (propositional) evidence, then p is justified by one's evidence. An event based theory of evidence might not have that property. If we follow Quine (1960, 235) and identify evidence with events of sensory irritation (which are inputs, and which justify), we wouldn't want to go on to say that every true proposition about our own sensory irritation is justified. Indeed, we can give a luminosity-based argument against that.

So I think it isn't a mere terminological question whether we say evidence consists of propositions or events. In part it feels like a convenience to use propositions as evidence, and assume that the structure of the 'justifies' relation is such that whatever other complications it has, p still justifies p. But since it is convenient, I'll stick to that assumption for now.

6.3   Chain Reaction Argument

Having established to his own satisfaction that evidence is propositional, Williamson goes on to consider what relation an agent must stand in to a proposition for it to be part of her evidence. Williamson argues against the theory that p is part of S's evidence iff S has a justified true belief that p. Call this theory, somewhat unimaginatively, the JTB theory of evidence.

One argument against the JTB theory of evidence is what we might call the 'chain reaction' argument. Consider a case where we investigate a number of Fs, say 100, and find each of them to be Gs. Since we have investigated, we'll suppose, a moderately diverse range of Fs we might well be justified in believing that the nearest unexamined F to us is G. Assume this is true; indeed, assume that all Fs are Gs. But also assume that we aren't yet justified in believing that all Fs are Gs. This seems like it should be consistent. After all, if some of the Fs are in very distant environments, very unlike the environments in which we have investigated Fs, then we have no reason to project our current data into such environments.

But note that by hypothesis we are justified in believing that the nearest unexamined F is G. And that belief is true. So if we adopt the belief, it will be part of

our evidence. Now by similar reasoning we'll have sufficient evidence to believe that the second nearest unexamined *F* is *G*. And that's true, so it goes into our evidence. And now we have enough evidence to conclude that the third nearest unexamined *F* is *G*, and so that goes into our evidence, and so on. Eventually, we can conclude that all *F*s are *G*s. Or, if you're worried that such a conclusion would involve some problematic kind of transfinite induction, we can conclude that *F*s in very different environments are also *G*, so long as there is a sequence of environments each very similar to the one before it, with each containing an *F*, linking our environment with the distant enviornment.

This seems like a *reductio* of the JTB theory of evidence. But it also, as Alvin Goldman (2009) points out, seems like a *reductio* of E=K. After all, ordinary inductive reasoning doesn't merely let us form the justified true belief that the nearest unexamined *F* will be *G*. It also lets us *know* this. Or at least it will do so long as we are not inductive sceptics. And once we add that to our evidence, we'll have sufficient evidence to know that the next nearest *F* will be *G*, and so on through to the *reductio*.

The non-inferentialist about knowledge has an easy way to block the chain reaction. If we insist that only non-inferential justified true belief is evidence, then we can't keep expanding the evidence set by doing these little inductions. So the chain reaction can never get started. I think that's probably the right way to stop the chain reaction getting going, but I want to consider two other related solutions.

Jim Joyce (2004) suggests a gradational solution to the puzzle. He says that we shouldn't think that whether someone has some evidence or not is an all or nothing matter. Rather, the possession relation for evidence can come in degrees. He suggests a version of gradationism that pays some homage to E=K, in that any knowledge we have is always better evidence than any non-knowledge. But all that matters for blocking the chain reaction is that as we get further and further away, the claims about the *n*th nearest *F* being *G* get to be sufficiently bad evidence that they don't support further inductive inferences. One puzzle with this view is that we need to explain why it is that the quality of the evidence deteriorates as we go further down the chain of inferences. And we need to do so while saying that it is still true that the points towards the end of the chain are, in some sense and to some extent, part of our evidence.

Here's one way to do that while keeping the letter of E=K alive. Those of us who, like Williamson, reject a sensitivity constraint on knowledge think that some of our knowledge is sensitive, and some isn't. It could be argued that this makes a difference to what kind of uses we can put evidence to. Here's a simple, surely too simple, way we could do this:

- Sensitive knowledge is evidence that can be used in either deductive or inductive reasoning.
- Insensitive knowledge is evidence that can only be used in deductive reasoning.

Now that's probably too strong. For one thing, sensitivity is a matter of degree, and we might think that the more sensitive some knowledge is, the more inductive reasoning it can support, rather than the simple binary picture we have here. But it is one way in which E=K could avoid the chain reaction argument. I suspect simply restricting evidence to basic knowledge is a better move, but it's hardly a mandatory move.

6.4 Evidence and Intuitions

Another argument against the JTB theory of evidence is that it gives us results that are counterintuitive. This is true, but as far as I can tell, this works equally well as an argument against E=K.

Start again with a case where an agent draws an inductive inference, say inferring that the nearest unobserved F is G from the evidence that a large and diverse range of Fs have been observed to be Gs. Then intuitively, a world in which the nearest F is $\neg G$ is consistent with their evidence. They are justified in believing that such a world does not obtain, but its obtaining is not, at least intuitively, entailed by their evidence.

This is, I think, a powerful argument that the JTB theory of evidence issues in unintuitive results. We might question how reliable intuition should be thought to be around here. After all, the role of evidence is somewhat theoretical, and perhaps intuition is not particularly reliable on something so theoretical. But as long as we just focus on the relation between the JTB theory and intuition, it is clear that there is a clash here.

The problem for the E=K theory is that the very same clash can arise between intuition and E=K. Consider the case where the agent comes to know, by performing a simple inductive inference, that the nearest unobserved F is G. Now consider the proposition that this F is $\neg G$. Intuitively, it is still consistent with their evidence. This is, I'd say, the simplest way to capture the idea that the inference the agent made is fallible, even though it produces knowledge. But if the proposition that the nearest F is $\neg G$ is consistent with the evidence, then it can't be that the evidence is everything that is known.

Now for the same reasons as I gave two paragraphs back, I don't know how strong a consideration this is. It seems plausible to me that our intuitions here are a little askew. But if to save E=K, we have to ditch some intuitions about evidence

in particular cases, then we can't turn around and use intuitions to oppose E=K's rivals. Yet that, it seems, is just what Williamson does.

6.5   Williamson's Processing Argument

Williamson also has an argument against the view that evidence is what one could know. I think this proposal criticises E=K from the wrong direction, since I think E=K counts too much as evidence, not too little. But it's an interesting argument, if a little hard to make rigorous. Here is Williamson's description of the case.

> [S]uppose that I am in a position to know any one of the propositions $p_1, ..., p_n$ without being in a position to know all of them; there is a limit to how many things I can attend to at once. Suppose that in fact I know p_1, and do not know $p_2, ..., p_n$. According to E=K, my evidence includes only p_1; according to the critic [who identifies evidence with knowability], it includes $p_1, ..., p_n$. Let q be a proposition which is highly probable given $p_1, ..., p_n$ together, but highly improbable given any proper subset of them; the rest of my evidence is irrelevant to q. According to E=K, q is highly improbable on my evidence. According to the critic, q is highly probable on my evidence. E=K gives the more plausible verdict, because the high probability of q depends on an evidence set to which as a whole I have no access. (Williamson, 2000a, 203)

There is, I think, something very odd about Williamson complaining that a view makes justification depend on things to which the agent has no access. I'd have thought part of the point of the anti-luminosity arguments was that this was an inevitable feature of our condition.

But more generally, it is hard to make sense of how the example is supposed to work. I assume that the imagined critic identifies what one is in a position to know with what one could know with a little more thought. If we identify what one is in a position to know with what one would know if one, say, looked, we'd get crazy results. Then it could be part of my evidence that my favourite socks are in the laundry hamper, since if I were to look in the hamper, I'd see them and know they were there, even if I haven't seen those socks in days. But that is crazy; looking other places for the socks could be rational given my evidential situation.

Given that sense of 'in a position to know', it seems that the other evidence the narrator has must make each of $p_2, ..., p_n$ very probable. That's because given that evidence, the narrator is in a position to know each of $p_2, ..., p_n$. And the

way the story is written, the conjunction $p_1 \wedge ... \wedge p_n$ is supposed to be highly relevant probabilistically to q. But now it isn't at all clear how the other evidence could be irrelevant to q, as Williamson says, since that evidence makes probable something that makes q probable.

And it isn't clear how q could be very probable given $p_1 \wedge ... \wedge p_n$, but not very probable given some subset of that, say $p_1 \wedge ... \wedge p_{n-1}$. Call that conjunction C to facilitate expression, and let the other evidence I have be E. The case seems to require the following to be true. First, $\Pr(q|C \wedge E)$ is low. Second, $\Pr(q|C \wedge E \wedge p_n)$ is high. Third, $\Pr(p_n|C \wedge E)$ is high. The first two points are stipulated by Williamson, and the third follows from the fact that the narrator is in a position to know p_n. But the second and third points together entail that $\Pr(q \wedge p_n|C \wedge E)$ is high-ish, since $\Pr(q \wedge p_n|C \wedge E) = \Pr(q|C \wedge E \wedge p_n) \times \Pr(p_n|C \wedge E)$. And obviously $\Pr(q|C \wedge E)$ is at least as high as $\Pr(q \wedge p_n|C \wedge E)$. So I don't think the example can really be probabilistically coherent.

But even if we fix up those problems, it feels to me as if the intuitions being adduced rely on running together propositional and doxastic justification. Let's say that E suffices for knowledge of $p_1 \wedge ... \wedge p_n$, and that clearly makes q probable, but it is very hard to see why q follows from E without going via $p_1 \wedge ... \wedge p_n$, and it is very hard to infer each conjunct of that conjunction from E given processing constraints. Then is q justified by E? One natural response is that the answer is 'yes' if we care about propositional justification, and 'no' if we care about doxastic justification. It's true that E itself is sufficient justification for q, at least in the sense that some agent could in principle believe q on the basis of E. But since the narrator has not gone through the necessary steps to infer q from E, any belief the narrator has in q would not actually be a justified belief, since it would not be related to the evidential basis E in the right way.

In general, distinguishing carefully between propositional and doxastic justification seems to help make the differences between Williamson's view, on which all knowledge is evidence, and a view on which only basic knowledge is evidence clearer. I think that evidence is basic knowledge. This isn't to say that there are times when p is part of an agent's non-basic knowledge, and she infers q from it, and q comes to thereby be part of her knowledge. It's true that part of why her belief in q is doxastically justified is that it was inferred from p. But I want to insist that the propositional justification for q will, in general, just be the propositional justification for p; this justification will 'transfer' to q. From the point of view of the agent, or even from the point of view of evaluating the agent as a cogniser, what is relevant is the role that the non-basic knowledge p plays. But when we focus on propositional justification, what matters is the evidence which supports p, and which is transferred to support for q via the agent's inferences.

6.6   Factivity of Evidence

Williamson also argues that evidence should all be true. There are a few distinct arguments for this conclusion that are worth teasing apart.

The first argument is that if evidence is not guaranteed to be true, then it isn't clear what the point of evidence is. I read both Joyce (2004) and Goldman (2009) as offering a similar kind of response to this argument. If evidence is probably true, or reliably true, then it could be useful to mark our beliefs to our evidence even if the evidence isn't always truth. Something can be a good guide to the truth without being perfectly reliable, and perhaps that's the status of evidence.

Another argument concerns the point that false evidence rules out, in a sense conclusively, some things that are true. If p is false and part of our evidence, we'll be able to deduce the falsity of $\neg p$ by simple deduction from the evidence. Williamson thinks it is acceptable for evidence to make some truths improbable, but it is unacceptable for evidence to conclusively rule out some truths. I'm not sure how strong a consideration this really is. If we agree that evidence can be inductively misleading, then it isn't obvious why it couldn't be deductively misleading.

Perhaps a stronger way to put Williamson's point here is that if evidence can be simply false, then it isn't easy to characterise what conclusive evidential support is. If even entailment by the evidence is consistent with the falsity of the supported proposition, then it seems that evidential support will always be less than conclusive. Maybe we could argue that this is something we could live with; it is, we might think, just a reflection of the fact that we have only fallible means of approaching the world. But it does seem unintuitive to think that we couldn't, even in principle, have such a thing as conclusive evidential support.

Finally, if we allow evidence to be false, we should worry about the possibility of evidence being inconsistent. Most fallibilist theories of evidence have conditions that rule out both p and $\neg p$ being part of one's evidence. But they don't, at least typically, rule out the possibility that a larger set of inconsistent propositions could all be part of one's evidence. And if they don't, it might turn out that there are propositions like 0=1 which are entailed by the evidence, but not in any way supported by the evidence. I think this is a very good reason to insist that evidence be consistent, and the best way to do that is to insist that evidence is true.

Chapter 7

Easy Knowldge

7.1 Cohen on Easy Knowledge

We've spent a lot of time so far concentrating on one of the traditional arguments for scepticism, the argument from sceptical possibilities. (Perhaps it would be better to call this a family of arguments; we'll return to this.) It's time to consider the other traditional argument, the problem of the criterion.

Very roughly, the argument goes as follows. Any knowledge we get must be via some method or other. But we can't use a method to gain knowledge unless we know that it is a knowledge-producing method. We might gloss this as saying we need to know that the method is reliable, but what matters is that it is knowledge-producing. And knowing that a method is knowledge-producing is a piece of knowledge. So to know anything, there is something we need to know before we can know anything. That's impossible, so we know nothing.

The problem of the criterion is potentially a very strong argument. After all, the conclusion of the last paragraph was not that we know nothing about the unobservable, or about the external world, or even about contingent matters. It is that we know nothing at all. That even extends to philosophical knowledge. So the problem of the criterion is naturally an argument for Pyrrhonian scepticism, the view that we cannot know anything, even the truth of philosophical claims like Pyrrhonian scepticism.

For much the same reason, the view looks so strong as to be self-defeating. You might think that by the lights of the Pyrrhonian sceptic, we can't even assert Pyrrhonian scepticism, since we can't know it to be true. That's too quick I think, since the plausibility of the rule *Only assert what you know* in academic contexts is pretty low. But still, a view that says we can't know that we exist, we can't know that we are thinking, we can't know that $0 \neq 1$, and so on is just absurd.

And worse still, it is an argument for an absurd conclusion with really only one key premise. Sometimes arguments can have absurd conclusions, but at least

they present us with a challenge to identify where things have gone wrong. Not here! The mistake is obviously in the premise that we can't use a method to gain knowledge unless we know that it is a knowledge-producing method, since that's the only premise there is! (I'm assuming here that we are reading 'method' so weakly that it is uncontroversial that any knowledge is gained by some method or other.)

And so most epistemologists do indeed reject that premise. Reliabilists say that any reliable method can produce knowledge, whatever the user of that method knows about the method's reliability. More internalist philosophers might say that we can use induction in advance of knowing that induction is reliable. Even Descartes, who is more sympathetic to these kind of priority principles than most epistemologists, thinks we can use clear and distinct perception before we know that it is reliable, though he does think we need to go back and check that it is really reliable.

But Cohen thinks that there are problems with any such move. In particular, he thinks that rejecting the principle KR will lead to unacceptably counterintuitive consequences.

KR A potential knowledge source K can yield knowledge for S, only if S knows that K is reliable.

There is something very odd about the way KR is stated. As is clear from the opening paragraph of Cohen (2002), what we're interested in here are *priority* arguments. And KR doesn't clearly entail any interesting premises about priority. This is relevant to Cohen's eventual conclusion. He thinks we can't say that either the knowlede we gain through a source, nor our knowledge that the source is reliable, are prior to the other. So he rejects what we might call PKR, for Priority KR.

PKR A potential knowledge source K can yield knowledge for S, only if S has prior knowledge that K is reliable.

There is conceptual space for a theorist who wishes to hold on to basic knowledge and yet also reject KR. One might hold that the basic knowledge is such that any time one uses it, one is in a position to know that it is reliably used, but before it is used one is not in such a position. I'm not sure whether any such position is plausible, but it is open.

But perhaps it will be clearer to have a position on the table which avoids the problem of the criterion by rejecting not just PKR, but also KR. In general, forms

of foundationalist externalism will do this. Let's consider a form of foundationalist externalism that says a method can produce basic knowledge if the following two conditions are met:

- The method is in fact reliable; and
- The agent has no reason to doubt that the method is reliable.

This is basically an externalist version of Jim Pryor's dogmatism (Pryor, 2000). The second clause is, I think, supposed to do a lot of work for Pryor, but is somewhat frequently overlooked by some of his critics. (I think a few of the arguments in White (2006), for instance, don't go through because of the second clause.) Similarly, I think the second clause is somewhat useful in responding to the problem of easy knowledge.

Cohen focusses on one particular kind of foundationalism, a kind of evidentialism that says we can know that something is φ on the basis of it appearing to be φ, even if we don't know at the time that appearances are reliable. This kind of theory could be either internalist, such as in dogmatism, or externalist, as in reliabilism. He makes three observations about this kind of theory.

First, the theory allows for a fairly simple response to doubts grounded in sceptical possibilities. If something appears to be a red table, and so we come to know that it is is a table, we can simply deduce that we are not in a tableless room but deceived by an evil demon to think there is a table. This looks too quick, but as Cohen concedes, any response to scepticism will have some odd feature.

Second, the theory allows for a fairly simple response to more everyday doubts. This is the core of Cohen's objection to basic knowledge views. For instance, he notes that the kind of foundationalism that he considers would allow an agent to easily infer that they are not looking at a white table illuminated by red lights simply on the basis of the appearance of a red table. And this he thinks is absurd. This is, it seems, the upshot of the imagined conversations with his (then) 7 year-old son.

Third, the theory seems to allow for a fairly simple generation of grounds for an inductive argument. After all, if whenever something appears φ, we can know that it is φ, for any φ, then we can easily test the accuracy of our appearances just by looking. And the test will be passed every time, with flying colours! So we will have grounds for an inductive argument that appearances are an accurate guide to reality. This is the conclusion of the argument on page 317.

I want to go over the question of whether the intuitions about these very cases are correct in the next section. But let's start with a question about the significance of these cases. After bringing up intuitions about these few cases, Cohen makes some rather sweeping generalisations about the impossibility of a

plausible theory of basic knowledge. And I doubt that generalisation is really supported by the intuitions.

Let's start by thinking about how the addition of a defeasibility clause, as I added to my preferred version of foundationalist externalism, affects things. Cohen carefully distinguishes between inferences from everyday propositions to the falsity of outlandish sceptical claims, and inferences from everyday propositions (like *That's a red table*) to the falsity of other everyday-ish propositions (like *That's not a white table illuminated by red lights*). But the everyday-ness of the the latter conclusions is a reason for doubt about the original appearance. If the agent is aware that there are sometimes white things that look red, that these are not just philosophers' playthings but really turn up in the real world, then she has reason to doubt that the red table is really as it appears.

In other words, once the basic knowledge theorist adds a defeasibility clause, I don't think Cohen can avoid considering the kind of sceptical scenarios that he sets aside on page 313. It might be that the only things we can know by basic means are relatively simple anti-sceptical propositions, since we have reason to doubt everything else. Put another way, it's arguable that the unintuitiveness of Cohen's example is due to the fact that we have reason to doubt that the lighting is normal in a lot of examples. So my preferred foundationalist externalist will think it is not a case of basic knowledge. And anything they do think is basic knowledge won't be subject to these doubts.

To make this point more dramatically, consider the theorist (such as perhaps Descartes) who thinks that introspection is a form of basic knowledge. Is it so unintuitive that we can see, by introspection, that introspection is reliable? We can, I think, introspect that p and that we are introspecting that p, and so deduce that introspection worked on that occasion. At the very least, this isn't obviously wrong. Put another way, we mostly consider that our pain appearances are reliable indicators of actually being in pain. They may or may not be reliable indicators of bodily damage, but they are reliable indicators of being in pain. We have no non-introspective evidence about this reliability. So we must, at some level, assume that introspection is good evidence that introspection is reliable.

7.2 What's Wrong with Easy Knowledge

It's hard to put one's finger on just what is supposed to be wrong with easy knowledge. I'm going to spend a bit of time going through various puzzling features of easy knowledge, and argue that each of them is compatible with easy knowledge being, in fact, knowledge. I'm going to concentrate on the easy knowledge argument against foundationalist externalism, but some of what I have to say should generalise to other theories that allow for basic knowledge.

7.2.1 Sensitivity

> *Objection:* If you use perception to test perception, then you'll come to believe perception is accurate whether it is or not. So if it weren't accurate, you would still believe it is. So your belief that it is accurate will be insensitive, in the sense of Nozick (1981). And insensitive beliefs cannot constitute knowledge.

The obvious reply to this is that the last sentence is false. As has been argued at great length, e.g. in Williamson (Williamson, 2000a, ch. 7), sensitivity is not a constraint on knowledge. We can even see this by considering other cases of testing.

Assume Smith is trying to figure out whether Acme machines are accurate at testing concrete density. She has ten Acme machines in her lab, and proceeds to test each of them in turn by the standard methods. That is, she gets various samples of concrete of known density, and gets the machine being tested to report on its density. For each of the first nine machines, she finds that it is surprisingly accurate, getting the correct answer under a very wide variety of testing conditions. She concludes that Acme is very good at making machines to measure concrete density, and that hence the tenth machine is accurate as well.

We'll return briefly to the question of whether this is a good way to *test* the tenth machine below. It seems that Smith has good inductive grounds for knowing that the tenth machine is accurate. Yet the nearest world in which it is not accurate is one in which there were some slipups made in its manufacture, and so it is not accurate even though Acme is generally a good manufacturer. In that world, she'll still believe the tenth machine is accurate. So her belief in its accuracy is insensitive, although she knows it is accurate. So whatever is wrong with testing a machine against its own outputs, if the problem is just that the resulting beliefs are insensitive, then that problem does not preclude knowledge of the machine's accuracy.

7.2.2 One-Sidedness

> *Objection:* If you use perception to test perception, then you can only come to one conclusion; namely that perception is accurate. Indeed, the test can't even give you any reason to believe that perception is inaccurate. But any test that can only come to one conclusion, and cannot give you a reason to believe the negation of that conclusion, cannot produce knowledge.

Again, the problem here is that the last step of the reasoning is mistaken. There are plenty of tests that can only produce knowledge in one direction only. Here are four such examples.

Brown is an intuitionist, so she does not believe that instances of excluded middle are always true. She does, however, know that they can never be false. She is unsure whether Fa is decidable, so she does not believe $Fa \lor \neg Fa$. She observes a closely, and observes it is F. So she infers $Fa \lor \neg Fa$. Her test could not have given her a reason to believe $\neg(Fa \lor \neg Fa)$, but it does ground knowledge that $Fa \lor \neg Fa$.

Jones is trying to figure out which sentences are theorems of a particular modal logic she is investigating. She knows that the logic is not decidable, but she also knows that a particular proof-evaluator does not validate invalid proofs. She sets the evaluator to test whether random strings of characters are proofs. After running overnight, the proof-evaluator says that there is a proof of S in the logic. Jones comes to know that S is a theorem of the logic, even though the failure to deliver S would not have given her any reason to believe it is not a theorem.

Grant has a large box of Turing machines. She knows that each of the machines in the box has a name, and that its name is an English word. She also knows that when any machine halts, it says its name, and that it says nothing otherwise. She does not know, however, which machines are in the box, or how many machines are in the box. She listens for a while, and hears the words 'Scarlatina', 'Aforetime' and 'Overinhibit' come out of the box. She comes to believe, indeed know, that Scarlatina, Aforetime and Overinhibit are Turing machines that halt. Had those machines not halted, she would not have been in the right kind of causal contact with those machines to have singular thoughts about them, so she could not have believed that they are not halting machines. So listening for what words come out of the box is one-sided in the sense described above; for many propositions, it can deliver knowledge that p, but could not deliver knowledge that $\neg p$.

Adams is a Red Sox fan in Australia in the pre-internet era. Her only access to game scores are from one-line score reports in the daily newspaper. She doesn't know how often the Red Sox play. She notices that some days there are 2 games reported, some days there is 1 game reported, and on many days there are no games reported. She also knows that the paper's editor is also a Red Sox fan, and only prints the score when the Red Sox win. When she opens the newspaper and sees a report of a Red Sox win (i.e. a line score like "Red Sox 7, Royals 3") she comes to believe that the Red Sox won that game. But when she doesn't see a score, she has little reason to believe that the Red Sox lost any particular game.

After all, she has little reason to believe that any particular game even exists, or was played, let alone that it was lost. So the newspaper gives her reasons to believe that the Red Sox win games, but never reason to believe that the Red Sox didn't win a particular game.

So we have four counterexamples to the principle that you can only know p if you use a test that could give you evidence that $\neg p$. The reader might notice that many of our examples involve cases from logic, or cases involving singular propositions. Both of those kinds of cases are difficult to model using orthodox Bayesian machinery. That's not a coincidence. There's a well known Bayesian argument in favour of the principle I'm objecting to, namely that getting evidence for p presupposes the possibility of getting evidence for $\neg p$. The argument turns on the fact that this is a valid argument, for any values of E, H, x you like.

1. $Pr(H) < x$
2. $Pr(E) > 0$
3. $Pr(H|E) \geq x$

C. $Pr(H|\neg E) < Pr(H)$

Intuitively, we might read this as saying that if E raises the probability of H above any threshold x, then $\neg E$ would be evidence against H.

I haven't discussed that objection here, because I think it's irrelevant. When dealing with foundational matters, like logical inference, Bayesian modelling is inappropriate. We can see that by noting that in any field where Bayesian modelling is appropriate, the objection currently being considered works! What's not so clear, in fact what I think is most likely false, is that we can model the above four examples in a Bayesian framework. Bayesianism just isn't that good at modelling logical uncertainty, or changes in which singular propositions are accessible to the agent. But that's what matters to these examples.

7.2.3 Generality

> *Objection:* Assume we can use perception to come to know on a particular occasion that perception is reliable. Since we can do this in arbitrary situations where perception is working, anyone whose perception is working can come to know, by induction on a number of successful cases, that their perception is generally reliable. And this is absurd.

I'm not sure that this really is absurd, but the cases already discussed should make it clear that it isn't a consequence of foundationalist externalism. It is easily possible to routinely get knowledge that a particular F is G, never get knowledge that any F is not G, and no way be in a position to infer, or even regard as probable, that all Fs are Gs.

For instance, if we let F be *is a Turing machine in the box Grant is holding*, and G be *halts*, then for any particular F Grant comes to know about, it is G. But it would be absurd for her to infer that all Fs are Gs. Similarly, for any Red Sox game that Adams comes to know about, the Red Sox win. But it would be absurd for her to come to believe on that basis that they win every game.

There's a general point here, namely that whenever we can only come to know about the Fs that are also Gs, then we are never in a position to infer inductively that all, or even most, Fs are Gs. Since even the foundationalist externalist doesn't think we can come to know by perception that perception is *not* working on an occasion, this means we can never know, by simple induction on perceptual knowledge, that perception is generally reliable.

7.2.4 Circularity

> *Objection:* It is impossible in principle to come to know that a particular method delivers true outputs on an occasion by using that very method. Since the foundationalist externalist says you can do this, her view must be mistaken.

We're going to spend a lot of time on this in a later section, but for now, let's note a quick problem for the view. Consider the way we might teach an intelligent undergraduate that (7.1) is a theorem.

$$((p \rightarrow q) \wedge p) \rightarrow q \tag{7.1}$$

We might first get the student to assume the antecedent, perhaps reminding them of other uses of assumption in everyday reasoning. Then we'll note that both $p \rightarrow q$ and p follow from this assumption. If we're ambitious we'll say that they follow by $\wedge$-elimination, though that might be too ambitious. And then we'll draw their attention to the fact that these two claims together imply q, again noting that this rule is called $\rightarrow$-elimination if we're aiming high pedagogically. Finally, we'll note that since assuming the antecedent of (7.1) has let us derive its consequent, then it seems that if the antecedent of (7.1) holds, so does the consequent. But that's just what (7.1) says, so (7.1) is true. For a final flourish, we might note the generalisation of that reasoning to the $\rightarrow$-introduction rule. Or we might call it quits at proving (7.1); undergraduates can only handle so much logic at one time.

That's a pretty good way, I think, of teaching someone that (7.1) is true, and indeed why it is true. That is, it is a way the student can learn (7.1), i.e., come to know (7.1). Indeed, I think it's more or less the way that I came to learn (7.1). But note something special about (7.1). It's obviously closely related to one of the rules we used to prove it, namely →-elimination. Indeed, when Achilles first tries to do without that rule in Lewis Carroll's famous dialogue, he appeals to an instance of (7.1). Given how closely related the two are, you might think using →-elimination to come to learn (7.1) would be viciously circular, a paradigm of what some people call rule-circularity. But in fact it isn't; the argument we just gave is a perfectly good way of learning (7.1). I don't know whether we should say it is circular but not viciously so, or simply non-circular, but however we classify the argument it isn't a bad argument. So I conclude that this rule can be used to come to learn that the rule is truth-preserving.

It might be objected at this point that the undergraduate must have already known (7.1) if they consent to →-elimination. I think this turns on an unrealistic picture of what undergraduates are capable of. It looks to me that (7.1) is in fact a very complicated proposition. It has a conditional embedded in the antecedent of a conditional! Such sentences are not familiar in everyday life, and most introductory logic students have problems parsing such sentences, let alone knowing they are true. So I don't think this kind of Socratic objection works.

It might also be objected that perception is different to logical reasoning. Such an objection might even be made to work. But it isn't a way of defending the objection we opened this subsection with; it relies on a principle that's simply false.

7.2.5   A Priority

> *Objection:* Assume it is possible to come to know that perception is reliable by using perception. Then before we even perceive anything, we can see in advance that this method will work. So we can see in advance that perception is reliable. That means we don't *come* to know that perception is reliable using perception, we could have known it all along. In other words, it is a priori knowable that perception is reliable. (This objection is related to an argument in White (2006), though note his argument is directed against a slightly different target.)

This objection misstates the foundationalist externalist's position. If perception is working, then we get evidence for this every time we perceive something, and reflect on what we perceive. But if perception is not working well, we don't get any such evidence. The point is not merely that if perception is unreliable, then

we can't possibly know that perception is unreliable since knowledge is factive. Rather, the point is that if perception is unreliable, then using perception doesn't give us any evidence at all about anything at all. So it doesn't give us evidence that perception is reliable. Since we don't know antecedently whether perception is reliable, we don't know if we'll get any evidence about its reliability prior to using perception, so we can't do the kind of a priori reasoning imagined by the objector.

This response relies heavily on an externalist treatment of evidence. An internalist foundationalist is perhaps vulnerable to this kind of objection. As I've argued elsewhere (Weatherson, 2005), internalists have strong reasons to think we can know a priori that foundational methods are reliable. Some may think that this is a reductio of interalism. (I don't.) But the argument crucially relies on internalism, not just on foundationalism.

7.2.6 Testing

> *Objection:* It's bad to test a belief forming method using that very method. The only way to learn that a method is working is to properly test it. So we can't learn that perception is reliable using perception.

This objection is, I think, the most interesting of the lot. I have a lot to say about it, and I've included a lot of it in the next section of these notes. This point takes us a fair way from debates about scepticism, so could be skipped by people who want to concentrate on the main plotline of these notes.

But I think this objection is interesting because in some ways I think the first premise, i.e. the first sentence in it, is true. Testing perception using perception is bad. What's surprising is that the second premise is false. The short version of my reply is that in testing, we aim for more than knowledge. In particular, we aim for sensitive knowledge. A test can be bad because it doesn't deliver sensitive knowledge. And that implies that a bad test can deliver knowledge, at least assuming that not all knowledge is sensitive knowledge. Defending these claims is the point of the next section.

7.2.7 Multiple Properties

> *Objection:* Let's say we grant that none of the six properties you mentioned so far is individually compatible with knowledge. That doesn't show that every combination of them is compatible with knowledge. In general, $\diamond p$ and $\diamond q$ don't entail $\diamond(p \wedge q)$. So you haven't shown easy knowledge is possible.

I don't quite know what to think about this objection. It strikes me as completely wrong-headed. The 'no easy knowledge' intuition seems, to me at least, to rest on an overlapping set of plausible but confused judgments about the relationship between knowledge, evidence and justifiability. We've shown that any possible reason one could have to support the intuition that easy knowledge is not knowledge is false, or not strong enough to support that conclusion. Could it be that the reasons work collectively when they don't work singularly? It's logically possible, but I don't see any reason at all to suspect it is true. But maybe I am just missing something philosophically significant.

7.3 Testing

In response to the 'testing' argument for the intuition that easy knowledge is no knowledge at all, I suggested that we should distinguish between a test being in general *good* and a test being the kind of thing which can ground knowledge. I think that's true because I think that tests really aim at *sensitive* belief. A test can fail in this aim, but still produce knowledge, because sensitivity isn't necessary for knowledge. That's probably a fairly intuitive idea, but it can be backed up by reflection on some examples involving testing which suggest more strongly the importance of sensitivity to good testing. The following case is slightly idealised, but I hope not terribly unrealistic.

Inspection

In a certain state, the inspection of scales used by food vendors has two components. Every two years, the scales are inspected by an official and a certificate of accuracy issued. On top of that, there are random inspections, where each day an inspector must inspect a vendor whose biennial inspection is not yet due. Today one inspector, call her Ins, has to inspect a store run by a shopkeeper called Sho. It turns out Sho's store was inspected just last week, and passed with flying colours. Since Sho has a good reputation as an honest shopkeeper, Ins knows that his scales will be working correctly.

Ins turns up and before she does her inspection watches several people ordering caviar, which in Sho's shop goes for $1000 per kilogram. The first customer's purchase gets weighed, and it comes to 242g, so she hands over $242. The second customer's purchase gets weighed, and it comes to 317g, so she hands over $317. And this goes on for a while. Then Ins announces that she's there for the inspection. Sho is happy to let her inspect his scales, but one of the customers, call him Cus, wonders why it is necessary. "Look," he says, "you saw that the

machine said my purchase weighed 78g, and we know it did weigh 78g since we know it's a good machine." At this point the customer points to the certificate authorising the machine that was issued just last week. "And that's been going on for a while. Now all you're going to do is put some weights on the scale and see that it gets the correct reading. But we've done that several times. So your work here is done."

There is something deeply wrong with Cus's conclusion, but it is surprisingly hard to see just where the argument fails. Let's lay out his argument a little more carefully.

1. The machine said my caviar weighed 78g, and we know this, since we could all see the display.
2. My caviar did weigh 78g, and we know this, since we all know the machine is working correctly.
3. So we know that the machine weighed my caviar correctly. (From 1, 2)
4. By similar reasoning we can show that the machine has weighed everyone's caviar correctly. (Generalising 3)
5. All we do in testing a machine is see that it weighs various weights correctly.
6. So just by watching the machine all morning we get just as much knowledge as we get from a test. (From 4, 5)
7. So there's no point in running Ins's tests. (From 6)

Cus's summary of how testing scales works is obviously a bit crude but we can imagine that the spot test Ins plans to do isn't actually any more demanding than what the scale has been put through while she's been standing there. So we'll let premise 5 pass.(If you'd prefer more realism in the testing methodology, at the cost of less realism in the purchasing pattern of customers, imagine that the purchases exactly follow the pattern of weights that a calibrator following the guidelines of the officially approved methods of calibration.) If 3 is true, it does seem 4 follows, since Cus can simply repeat his reasoning to get the relevant conclusions. And if 4 and 5 are true, then it does seem 6 follows. To finish up our survey of the uncontroversial steps in Cus's argument, it seems there isn't any serious dispute about step 1.

(It may be epistemologically significant that human perception does not, in general, work the same way as balances in this respect. We may come to know p through perception before we know that p was the output of a perceptual module. Indeed, an infant may come to know p through perception even though she lacks the concept of a perceptual module. This may be of significance to the

applicability of the problem of easy knowledge to human perception, but it isn't immediately relevant here.)

So the contentious steps are:

- Step 2 - we may deny that everyone gets knowledge of the caviar's weight from the machine.
- Step 3 - we may deny that the relevant closure principle that Cus is assuming here.
- Step 7 - we may deny that the aim of the test is (merely) to know that the machine is working.

One way to deny step 2 is to just be an inductive sceptic, and say that no one can know that the machine is working merely given that it worked, or at least appeared to work, last week. But that doesn't seem very promising. It seems that the customers do know, given that the testing regime is a good one, and that the machine was properly tested, that the machine is working. And the inspector has all of the evidence available to the customers, and is in an even better position to know that the testing regime is good, so as step 2 says, she gets knowledge of the caviar's weight from the machine.

In recent years there has been a flood of work by philosophers denying that what we know is closed under either single-premise closure, e.g., Dretske (2005), or multi-premise closure, e.g. Christensen (2005). But it is hard to see how kind of anti-closure view could help here. We aren't inferring some kind of heavyweight proposition like that there is an external world. And Dretske's kind of view is motivated by avoidance of that kind of inference. And Christensen's view is that knowledge of a conjunction might fail when the amount of risk involved in each conjunct is barely enough to sustain knowledge. But we can imagine that our knowledge of both 1 and 2 is far from the borderline.

A more plausible position is that the inference from 1 and 2 to 3 fails to *transmit* justification in the sense that Crispin Wright (2002) has described. But that just means that Ins, or Cus, can't get an initial warrant, or extra warrant, for believing the machine is working by going through this reasoning. And Cus doesn't claim that you can. His argument turns entirely on the thought that we already know that the machine is reliable. Given that background, the inference to 3 seems pretty uncontroversial.

That leaves step 7 as the only weak link. I want to conclude that Cus's inference here fails; even if Ins knows that the machine is working, it is still good for her to test it. But I imagine many people will think that if we've got this far, i.e., if we've agreed with Cus's argument up to step 6, then we must also agree with step 7. I'm going to offer two arguments against that, and claim that step 7

might fail, indeed does fail in the story I've told, even if what Cus says is true up through step 6.

First, even if Ins won't get extra knowledge through running the tests on this occasion, it is still true that this kind of randomised testing program is an epistemic good. We have more knowledge through having randomised checks of machines than we would get from just having biennial tests. So there is still a benefit to conducting the tests even in cases where the outcome is not in serious doubt. The benefit is simply that the program, which is a good program, is not compromised.

We can compare this reason Ins has for running the tests to reasons we have for persisting in practices that will, in general, maximise welfare. Imagine a driver, called Dri, is stopped at a red light in a quiet part of town in the middle of the night. Dri can see that there is no other traffic around, and that there are no police or cameras who will fine her for running the red light. But she should, I think, stay stopped at the light. The practice of always stopping at red lights is a better practice than any alternative practice that Dri could implement. I assume she, like most drivers, could not successfully implement the practice *Stay stopped at red lights unless you know no harm will come from running the light*. In reality, a driver who tries to occasionally slip through red lights will get careless, and one day run a serious risk of injury to themselves or others. The best practice is simply to stay stopped. So on this particular occasion Dri has a good reason to stay stopped at the red light: that's the only way to carry out a practice which it is good for her to continue.

Now Ins's interest is not primarily in welfare, it is in epistemic goods. (She cares about those epistemic goods because they are related to welfare, but her primary interest is in epistemic goods.) But we can make the same kind of point. There are epistemic practices which are optimal for us to follow given what we can plausibly do. And this kind of testing regime may be the best way to maximise our epistemic access to facts about scale reliability, even if on this occasion it doesn't lead to more *knowledge*. Indeed, it seems to me that this is quite a good testing regime, and it is a good thing, an epistemically good thing, for Ins to do her part in maintaining the practice of randomised testing that is part of the regime.

The second reason is more important. The aims of the test are, I claim, not exhausted by the aim of getting knowledge that the machine is working. We also want a sensitive belief that the machine is working. Indeed, we may want a sensitive belief that the machine has not stopped working since its last inspection. That would be an epistemic good. Our epistemic standing improves if our belief

that the machine has not stopped working since its last inspection becomes sensitive to the facts. Before Ins runs the test, we know that the machine will work. If we didn't know that, we shouldn't be engaged in high-stakes transactions (like the caviar sales) that rely on the accuracy of the machine. But our belief that the machine will work is not sensitive to one not completely crazy possibility, namely that the machine has recently stopped working. After the test, we are sensitive to that possibility.

This idea, that tests aim for sensitivity, is hardly a radical one. It is a very natural idea that good tests produce results that are correlated with the attribute being tested. And 'correlation' here is a counterfactual notion. For variables X and Y to correlate just means that if X had been different, then Y would have been different, and the ways Y would have been different had X been different are arranged in a systematic way. When we look at the actual tests endorsed in manuals on how to calibrate balances, producing this kind of correlation looks to be a central aim. If a machine weren't working, and it were run through these tests, the tests would issue a different outcome than if the machine were working. But 'testing' the machine by using its own readings cannot produce results that are correlated with the accuracy of the machine. If the machine is perfectly accurate, the test will say it is perfectly accurate. If the machine is somewhat accurate, the test will say it is perfectly accurate. And if the machine is quite inaccurate, the test will say that it is perfectly accurate. The test Ins plans to run, as opposed to the 'test' that Cus suggests, is sensitive to the machine's accuracy. Since it's good to have sensitive beliefs, it is good for Ins to run her tests.

So I conclude that step 7 in Cus's argument fails. There are reasons, both in terms of the practice Ins is part of, and in terms of what epistemic goods she'll gain on this occasion by running the test, for Ins to test the machine. That's true even if she knows that the machine is working. The epistemic goods we get from running tests are not restricted to knowledge. That's why it is a bad idea to infer from the badness of testing our eyes, say, using our eyes that we cannot get knowledge that way. The aims of tests don't perfectly match up with the requirements of getting knowledge.

Chapter 8

Disagreement and Circularity

8.1 Bootstrapping and Equal Weight

In recent years, a number of philosophers have become interested in the 'Equal Weight View' (EWV) of peer disagreement. This view holds that if you and someone you properly antecedently took to be an epistemic peer disagree about some matter, you should, *ceteris paribus* take it to be equally likely that each of you are right. So if you started fully believing p, and your peer fully believed $\neg p$, you should end up thinking it is 50% likely that p.

There are weaker and stronger versions of this view depending on just what we build into the *ceteris paribus* clause. If one of the thing that stays equal is that it is proper to regard each of you as peers with respect to p, and peerhood simply means that you are equally likely to have gotten it right, then simple coherence requires that you move your credence in p to 0.5. So the weak versions of this view are correct, but too weak to be of much interest.

The interest in recent years though has been in much stronger versions of the theory. Adam Elga (2007) and David Christensen (2010) have both argued for very tight conditions on how we understand that clause. Here's Elga's way of putting it.

> Upon finding out that an advisor disagrees, your probability that you are right should equal your prior conditional probability that you would be right. Prior to what? Prior to your thinking through the disputed issue, and finding out what the advisor thinks of it. Conditional on what? On whatever you have learned about the circumstances of the disagreement. (Elga, 2007, 490)

That is, as long as you don't learn anything distinctive about the 'circumstances of the disagreement', you shouldn't change your judgment of peerhood. And it's clear that Elga does not regard the fact that your 'peer' made a judgment

which you have strong evidence is false is part of the circumstances. Rather, the circumstances are just things like whether your 'peer' is drunk, distracted, or otherwise predisposed to err. It isn't clear whether the fact that your advisor is, say, failing to carry a number in an addition, or affirming the consequent, or the like counts as a circumstance. From what Elga says about various cases, it seems like it isn't, but this isn't clear.

The reason for this strong constraint on what can go into *ex post* judgments about peerhood is that the alternative, Elga argues, would be a kind of circularity. If we were allowed to use our judgments to show that our judgments were more reliable, we'd have a problematic kind of 'bootstrapping'.

> You and a friend are to judge the same contest, a race between Horse A and Horse B. Initially, you think that your friend is as good as you at judging such races. In other words, you think that in case of disagreement about the race, the two of you are equally likely to be mistaken. The race is run, and the two of you form independent judgments. As it happens, you become confident that Horse A won, and your friend becomes equally confident that Horse B won.
>
> When you learn of your friend's opposing judgment, you should think that the two of you are equally likely to be correct. For suppose not – suppose it were reasonable for you to be, say, 70% confident that you are correct. Then you would have gotten some evidence that you are a better judge than your friend, since you would have gotten some evidence that you judged this race correctly, while she misjudged it. But that is absurd. It is absurd that in this situation you get any evidence that you are a better judge.
>
> To make this absurdity more apparent, suppose that you and your friend independently judge the same long series of races. You are then allowed to compare your friend's judgments to your own. (You are given no outside information about the race outcomes.) Suppose for *reductio* that in each case of disagreement, you should be 70% confident that you are correct. It follows that over the course of many disagreements, you should end up extremely confident that you have a better track record than your friend. As a result, you should end up extremely confident that you are a better judge. But that is absurd. Without some antecedent reason to think that you are a better judge, the disagreements between you and your friend are no evidence that she has made most of the mistakes. (Elga, 2007, 486-7)

Thomas Kelly (2010) has argued that not every theory other than Equal Weight is vulnerable to this objection. But whether or not that's true, it is worth noting that the final sentence is a very strong priority claim. It says that what someone judges can never be evidence for the reliability of their judgment. In practice, that's not how we work. If I find out that someone I had antecedently thought was reliable on most matters turns out to be a climate change sceptic, I'll simply judge that they aren't particularly reliable on climate change judgments. And it's a very odd priority claim; what could be better evidence that S has become a better judge than T than seeing a series of disagreements where S is right and T is wrong? Yet if Elga is right, that's exactly the kind of evidence that is barred to S.

The restriction on what can go into our judgments of peerhood, and their connection to anti-circularity principles, is even clearer in David Christensen's (2010).

> **Independence** In evaluating the epistemic credentials of another's expressed belief about P, in order to determine how (or whether) to modify my own belief about P, I should do so in a way that doesn't rely on the reasoning behind my initial belief about P.
>
> The motivation behind the principle is obvious: it's intended to prevent blatantly question-begging dismissals of the evidence provided by the disagreement of others. It attempts to capture what would be wrong with a P-believer saying, e.g., "Well, so-and-so disagrees with me about P. But since P is true, she's wrong about P. So however reliable she may generally be, I needn't take her disagreement about P as any reason at all to question my belief." (Christensen, 2010, 1-2)

To my eyes, this argument seems to involve a category mistake. Moves in a dialectic can be question-begging or not. But here Christensen seems to want to put restrictions on rational judgments on the grounds that the alternative would be question-begging. That seems like the wrong way to get the desired end. If we want to stop "blatantly question- begging dismissals" we can just remind people not to be rude.

I think the problem Christensen is trying to get at is not to do with question-begging, but to do with circularity. The problem he's getting at is that if we violate **Independence**, we can use our reasoning to conclude that our reasoning is reliable, and that's circular. Or, to be more accurate, it has a whiff of circularity about it. Trying to turn this into an argument for **Independence** though will be difficult.

Part of the difficulty is that it isn't easy to say exactly what the circularity involved is. Consider the following little example, where S and T are trying to figure out who dun it, and the only suspects are the butler and the gardener. S reasons as follows.

1. The butler had means, motive and opportunity to commit the crime
2. Independent witnesses, who had no motive to lie, say that the gardener was in Australia at the time the crime was committed in Scotland.
3. So the butler did it. (From 1 and 2)
4. I observe, by introspection, that I believe the butler did it.
5. I also have compelling testimonial evidence, from T, that T thinks the gardener did it on the basis of 1 and 2.
6. Since T and I disagree (from 4 and 5), and I'm correct (from 3), this is a case where T hasn't done as well as me at reasoning about a case, so even if we're both equally reliable in general, we're not equally reliable here.

It isn't clear to me just which step is meant to be circular. If S had reasoned as follows, I could see how we might take her reasoning to be circular.

1. The butler had means, motive and opportunity to commit the crime
2. Independent witnesses, who had no motive to lie, say that the gardener was in Australia at the time the crime was committed in Scotland.
3. When I reflect on 1 and 2, it seems to me that the butler did it.
4. Since I'm generally reliable, the butler did it.
5. When T reflects on 1 and 2, it seems to her that the gardener did it.
6. Since T and I disagree (from 4 and 5), and I'm correct (from 3), this is a case where T hasn't done as well as me at reasoning about a case, so even if we're both equally reliable in general, we're not equally reliable here.

If S reasons this way, the only reason for thinking she is right and T is wrong is her own judgment, which is exactly what is at issue in 6. But that isn't at all how people usually reason. Nor is it a sensible 'rational reconstruction' of their reasoning. Rather, the first inference listed here is much more like the way normal human beings do, and should, reason. And in this case the dispute between S and T is broken by a fact recorded at line 3, namely that the butler did it. And that fact is in turn supported by some evidence that's not at all about S's reliability, namely the evidence given at lines 1 and 2. So if there is some circular reasoning going on here, the circularity is fairly subtle, and it won't be easy to say just what it is.

Still, there is some vague feeling of circularity that goes along with the reasoning that both Elga and Christensen provide. And in principle we shouldn't say that some reasoning is acceptable just because we can't precisely articulate the sin it commits. (Compare: We shouldn't say that Gettier cases are cases of knowledge just because it is hard to say exactly what makes them not knowledge.) But I think we have independent reason to think that this vague whiff of circularity does not in fact track problematic features of arguments.

8.2 Eight Easy Arguments

I think the anti-circularity points that Elga and Christensen rely on can easily over-generate. It would be nice to have a completely compelling circular but acceptable argument. But those are a little tricky to find. Instead I'm going to try for proof by quantity. Here are eight arguments that seem vaguely circular in just the way that Elga and Christensen think the anti-EWV theorist is committed to vaguely circular reasoning. I'm inclined to think that all of them are acceptable. I'm more convinced that at least one of them is (a) acceptable, and (b) 'circular' in just the way that is supposed to be problematic for the anti-EWV theorist.

8.2.1 *Cogito*

Let's start with the most famous argument in Western Philosophy: Descartes' *Cogito*. There is something vaguely circular about this. Let's imagine someone who entertained serious doubts about their own existence. Could we use the *Cogito* to talk them out of these doubts? I'm not sure we could. Why, they could ask, should they accept the premise? We could try to persuade them of their awareness of their own thoughts. But it would be in order for them to reply that nothing can be detected but the existence of thoughts; the owner of those thoughts cannot be clearly detected. Given doubts about their own existence, that seems a reasonable reply.

I don't think even Descartes would disagree with that assessment. The point of the *Cogito* is not to repel doubts about one's own existence. Rather, it is to show that we do not have grounds for such doubts. It is an undercutting defeater of doubts, not a rebutting defeater, in more contemporary terminology. Put another way, the point of the argument is to reveal that we had warrant for belief in our own existence all along, not to provide a new warrant for belief in our existence. And that's a good thing; its circularity prevents it from providing such a new warrant.

8.2.2 Clear and Distinct Arguments

Descartes seems to give two arguments for the reliability of clear and distinct perception. As we saw earlier, it is a little tricky to get clear on what he thinks the relationship between them is, and it is possible that he doesn't really endorse the first. But let's look at both of them, since they are both potentially circular.

First, we'll look at the argument right at the start of *Meditation* 3 for the reliability of clear and distinct ideas. Descartes presents, without clearly endorsing, the following argument.

1. I can know that I exist.
2. My method of coming to know that I exist is the faculty of clear and distinct perception.

———————————

C. The faculty of clear and distinct perception can produce knowledge.

The obvious problem with this argument is not that it is circular. As far as I can see, it isn't problematically circular in any way. The problem is that premise 2 isn't supported, or really that easy to support. Since I don't think I believe in such thing as a faculty of clear and distinct perception, I don't even think it is true. But I certainly don't think Descartes is justified, at this stage of the *Meditations* in using it. So I think this argument fails, but not for reasons of circularity. Perhaps Descartes thought so too, since he moved on to this argument.

1. God exists.
2. If God exists, then maximally careful reasoning is sufficiently reliable to produce knowledge.

———————————

C. Maximally careful reasoning is sufficiently reliable to produce knowledge.

Of course, premise 1 is not really a premise for Descartes, but something he concludes from a longer argument. And the premises of that argument are things he gets via his faculty of clear and distinct perception. But I think it's interesting to think about what we would say about the argument were we to spot Descartes premise 1. And Descartes has views about what is maximally careful, in particular it is clear and distinct perception. But we can also bracket that bit of Cartesian theory in evaluating this argument.

I think, and as we saw back in chapter 2 the contemporary scholarly consensus agrees, is that the argument would then be perfectly acceptable. And that's true even though Descartes has come to believe premise 2 using the kind of maximally careful reasoning that he is trying to show is reliable in the conclusion.

After all, Descartes doesn't use the reliability of maximally careful reasoning as a premise in his reasoning. He doesn't reason from the fact that he was maximally careful in coming to believe that God's goodness is incompatible with the unreliability of his maximally careful reasoning. Rather, he just is careful when reasoning, and concludes that these things are incompatible. And he's right; a perfectly good God would not allow us to be mistaken in this way. (Whether a perfectly good God would allow us to be mistaken in any way is a harder question, but not one that's at issue here.) So I think that if Descartes can have premise 1, then this argument is good. And it is good despite being an argument to the conclusion that one of the methods used in discovering a premise is reliable.

8.2.3 Achilles and the Tortoise Revised [After Carroll (1895)]

Achilles has been reading a logic textbook. The book explains the modus ponens rule, and says that it is valid. Achilles has strong independent reason to think that the book is highly reliable. So he comes to believe that modus ponens is valid. He tells Tortoise this, who asks him if he's used this argument.

Tortoise's First Reconstruction

1. The book says modus ponens is valid.

———————————

C. Modus ponens is valid.

Tortoise doesn't think this is a very good argument, and neither does Achilles when it is presented that bluntly. After all, Achilles relied on the reliability of the book in his reasoning. So Tortoise offers him this reconstuction.

Tortoise's Second Reconstruction

1. The book says modus ponens is valid.
2. If the book says modus ponens is valid, then modus ponens is valid.

———————————

C. Modus ponens is valid.

Achilles is happy to endorse this. And he should be; it's an excellent argument. But it is blatantly rule circular. I conclude from this that rule circularity isn't as big a problem as is sometimes assumed.

8.2.4 More Achilles and the Tortoise

As he reads further through the book, Achilles comes to learn that $((p \rightarrow q) \wedge p) \rightarrow q$ is a valid schema. He uses modus ponens, which he just learned, in learning this. For reasons we went over above, it looks like this is both circular and acceptable reasoning. But we've said enough about this case earlier.

8.2.5   G. E. Moore

We're probably all familiar with this argument, often traced to Moore (1959):

Moore

1. Here is one hand, and here is another.

   _____

C. The external world exists.

There is a large argument about whether this argument is *merely* dialectically in-efficacious, or whether it is really a bad argument. The best argument for the prosecution is that it is not a kind of argument that could, even in principle, be used to overcome doubt about the conclusion. The best argument, I think, for the defence is that the argument could be used to come to belief in the conclusion by someone who neither believed nor doubted the conclusion. I'm on the defence side of this one, but I suspect this is the least persuasive of the eight cases I'm going through here.

8.2.6   Lewis on Dialetheism

David Lewis (1982) proposes the following argument against dialetheism, the theory that there are true contradictions.

Lewis

1. Dialetheism implies that some contradiction is true.

   _____

C. Dialetheism is not true.

This is rule circular, in at least some sense. The argument is valid, indeed it is a nice instance of a perfectly general valid form of argument: *reductio*. But that form of argument is what's at issue in any debate about dialetheism. So it uses a rule that anyone who doubted the conclusion would doubt. Still, it seems like a perfectly good argument to me. In particular, it seems like the best argument we actually have against dialetheism. And since we aren't actually dialetheists, that's a reason for thinking it is a good argument.

What could a non-circular argument against dialetheism look like? We might try to undermine the best argument from dialetheism, an argument from the semantic paradoxes, by providing a consistent resolution of the paradoxes. I think there are some consistent resolutions of the paradoxes. But they tend to be 'ugly' in various ways, either insisting on ad hoc restrictions to the T-schema,

or on quirky restrictions on proposition formation, or on compositional meaning rules. If the alternative is inconsistency, I think the prices of these resolutions is worth paying. But that's hardly a non-circular argument against dialetheism. It's an argument that starts with the assumption that dialetheism would be a very costly solution to the paradoxes, so we should accept some other resolution, even if it is not too happy.

It's possible that we can do better than this. We could argue that even on neutral grounds, there is no good dialetheist solution to the Curry paradox. A solution to the Curry paradox seems to require either the kind of ad hoc moves a consistent resolution of the Liar paradox makes, or some surprising tweaks to the structural rules of logic. We'd then try to argue that even on neutral grounds, the costs of the dialetheist solution are too high.

I don't know whether this kind of argument could work. But I do know dialetheism is false. So I know there are grounds for rejecting dialetheism independent of these non-circular grounds. And I'm pretty sure that most of you think that dialetheism is false, and not because it lacks an elegant resolution of the Curry paradox. So I think you're committed to the acceptability of circular objections to dialetheism.

8.2.7 Trivialist

Graham Priest (2000) in recent work has been worried about the trivialist, the person who says that every proposition is true. I think he's right to worry about the trivialist actually. The stronger an 'anti-circularity' principle we can justify, the better the dialetheist looks. But the trivialist threatens particularly strong anti-circularity principles.

It's tempting to argue as follows against the trivialist.

1. It's not the case that $0 \neq 1$

   _____

   C. Trivialism is false

But that doesn't quite do it, since the trivialist accepts the conclusion. The conclusion, after all, is a proposition, and propositions are, he says, all true. To even formulate the dispute with the trivialist, we need to take seriously the distinction between denying a proposition and asserting its negation. Since Frege showed that we can't treat all negations as denials, some philosophers have been hostile to the notion of denial. But there's no simple Fregean argument that *all* denials are assertions of negations. When we deny that $0 = 1$, we aren't merely asserting that $0 \neq 1$. If we were, the trivialist would agree with us, but the trivialist can't agree with the denial.

Still, it seems clear that any argument for denying trivialism will either be from steps the trivialist rejects on principled grounds, i.e., denials, or will be an argument the trivialist thinks is plainly bad, since it is an argument from some assertions to a denial. So it seems any argument against trivialism will be circular in some sense. But it also seems we should reject trivialism, and clearly so.

8.2.8   The Reason Sceptic

Our last example has been discussed a fair bit by Timothy Williamson (2007) in work on philosophical disputes. The reason sceptic denies that there are any reasons. Hence he also denies that there are any reasons for accepting that there are reasons.

The trivialist is an absurd creature, but I can vaguely see how the reason sceptic might be motivated. If you think Mackie-style arguments from queerness motivate to some extent moral scepticism (Mackie, 1977), I think you should find similar arguments for reason scepticism similarly motivating. After all, reasons as they are usually understood by philosophers are a little 'queer' in Mackie's sense. Of course, the reason sceptic can't say that Mackie-style arguments provide a *reason* to be a reason sceptic! But she can say that if there are such things as reasons, then Mackie-style arguments provide a *reason* to be a reason sceptic, and she might try to parlay that into a *reductio* of the idea that there are reasons. I don't mean for a second to suggest that this is a plausible position – there is literally nothing that can be said in its favour! But I think it is not nearly as far fetched as trivialism, or for that matter dialetheism.

Now how should we argue against the reason sceptic. Should we offer reasons that he is wrong? That would be circular, since whether there are such things as reasons is part of what's at issue. But it would be the right thing to do! When one is arguing, one should offer reasons. And one should do so even if playing the game of giving and asking for reasons presupposes that one's interlocutor is mistaken. So I think any good response, and hence the best response, to the reason sceptic, will be circular in some way. But I also think those are good responses, since reason scepticism is false.

8.2.9   Summing Up

I'm rather hostile to sweeping arguments from circularity. (As you might have guessed from the last 20 pages or so!) Very strong anti-circularity principles would, if consistently adopted, lead us to take trivialism, reason scepticism and the like as positions that we could not reject. But weak anti-circularity principles would not be sufficient to ground arguments like Cohen's argument against basic knowledge, or Elga's argument for the equal weight view of justification. I

haven't argued here that there can't be viable 'intermediate' anti-circularity principles that are weak enough to allow in at least some of the eight arguments listed here, but strong enough to rule out the kind of reasoning that Cohen and Elga want to block. But I don't see any good reason to think that such principles would be remotely plausible.

Chapter 9

Reasoning and Dialectics

9.1 Pryor's Classification of Epistemological Views

James Pryor (2004), distinguishes three different approaches we might take towards different methods in epistemology. The first two labels are his; the third is mine, but it seems like a natural extension of his terminology. In every case, we assume agent S used method M to get a belief in proposition p. And we'll say the proposition *M works* is the conjunction of every proposition of the form *(M represents that q)* $\supset q$ for every salient q. Then we have the following three views. (I'm modifying Pryor's views a bit to make these attitudes towards methods, rather than towards propositions, but I think this makes everything a touch clearer.)

Conservatism S gets a justified belief in p only if she antecedently was justified in believing that *M works*.

Liberalism S can in some circumstances get a justified belief in p without having an antecedently justified belief in *M works*, but in some other circumstances she can properly use M and not get a justified belief in p, because her prior evidence defeats the support that M provides for p.

Radicalism As long as S uses M correctly, and M genuinely says that p, and M actually works, then no matter what evidence S has against *M works*, she gets a justified belief in p.

Different methods will naturally be treated in different ways.

- Scientific advances naturally produce a lot of methods that we should treat **conservatively**. If I come up with a fancy model for the economy, I could over time use that model to form justified predictions about economic activity. But we should treat the method of using this model conservatively.

I only get justified beliefs by using the model if I have an antecedent reason to believe that the model works.

- Plausibly we should be **radicals** about our most fundamental methods, such as introspection. A child doesn't antecedently need to know that introspection is reliable to come to have introspective knowledge that she's in pain. As long as introspection works, it isn't clear this is defeasible. If as the child grows up, she hears from some fancy philosophers that there is no such thing as pain, she might get some reasons to doubt that introspection works. But when she introspectively (and perhaps involuntarily) forms the belief that she's in pain, she knows she is in pain.

- It is a little trickier to say which methods we should be **liberals** about. Pryor suggests that we should be liberals about perception. (And he seems to suggest that we shouldn't be radicals about any method, though I'm not sure if that's really his view.) Many epistemologists are liberals about testimony. They deny that we need antecedent reason to believe that testimony works before getting testimonial knowledge, but they do think that of course we shouldn't just believe everything we hear, so testimonial justification is defeasible. Plausibly we should also take a liberal attitude towards memory. We don't need to know memory works to get knowledge from memory, but reasons to doubt that memory works might block the justification which memory normally provides us.

Conservatism and radicalism are fairly well defined views. The main thing we need to say about distinguishing different types of conservativism is that some conservatives have supplementary views that greatly alter the effect of their conservatism. For instance, the Cartesian sceptic is a conservative about perception who denies that we can believe perception works without having perceptual beliefs. But some other philosophers are conservatives about perception who also believe that it is a priori that perception works. Those positions will be radically anti-sceptical. The point is that what effect conservatism has depends on what turns up elsewhere in the epistemology.[1]

In contrast to conservativism and radicalism, liberalism covers a wide variety of fairly disparate theories. The liberal essentially makes a negative claim, antecedent justification for believing that *M works* is not needed for getting a justified belief that p, and an existential claim, there is some way of blocking the

---

[1] You'd think the same would be true in politics. Being a conservative in Mississippi should be very different to being a conservative in Denmark, because you'd be trying to conserve very different things. But the term 'conservative' in modern English is more typically used for those who favour a very particular set of social and political arrangements, roughly those in place in much of the English speaking world in the mid to late 19th Century.

support *M* provides to *p*. We get differing views depending on how we might make that existential claim more precise.

A conservative-leaning liberal thinks that there are a lot of ways to block the support that *M* provides to *p*. One natural way to be a conservative-leaning liberal is to say that whenever *S* has any reason to doubt that *M works*, the use of *M* does not justify belief in *p*. Pryor's own view on perception is this kind of conservative-leaning liberalism is true about perception. If any kind of liberalism about testimony is correct, then presumably it is a very conservative-leaning liberalism, since it is easy to block the support that testimony that *p* provides to *p*.

A radical-leaning liberal thinks that there are very few ways to block the support that *M* provides to *p*, even if in principle there are some. One natural way to be a radical-leaning liberal is that the support is blocked only if *S* believes, or is justified in believing, that *M works* is false. An even more radical view says that the support is blocked only if *S* knows that *M works* is false. A fairly radical form of liberalism seems plausible for memory; we are entitled to trust memories unless it seems we have good reason to doubt them. Once we see that very radical forms of liberalism are possible, it might be possible to deny that pure radicalism is ever true, though I suspect radicalism is still right about introspection.

9.2 Liberalism and Circularity

Pryor also notes an interesting way in which arguments can seem to be circular. He doesn't give this a fancy name, but we'll call it **Defeater Circularity**.

Defeater Circularity An argument displays **Defeater Circularity** iff evidence against the conclusion would (to at least some degree) undermine the justification the agent has for the premises. (This is Pryor's Type 4 dependence; see (Pryor, 2004, 359))

It is important that Pryor uses 'undermine' here rather than something more general, like 'defeat'. Any valid one premise argument will be such that evidence against the conclusion will *rebut*, at least to some degree, the justification for the premises. But it won't be necessary that this evidence *undermines* that justification. If I reason *X is in Ann Arbor, so X is in Michigan*, then evidence against my conclusion will rebut whatever evidence I had that X is in Michigan. But that might not undermine my support. If I thought X was in Ann Arbor because a friend said that they just saw X, the counter-evidence need not impugn my friend's reliability in general. It might just mean my friend got this one wrong.

It is not crazy to think that arguments which exemplify **Defeater Circularity** are defective in some way. In particular, we might think that what's wrong with

arguments like Moore's proof of the external world is that they suffer from **Defeater Circularity**. Pryor, however, wants to argue that some arguments which exemplify **Defeater Circularity** are not bad arguments in some crucial respects. In particular, he wants to argue that some such arguments are such that reasoning through the premises to the conclusion can generate knowledge of the conclusion.

Pryor gives two arguments for this conclusion. First, he offers direct examples of arguments that he says exemplify **Defeater Circularity**, but which could, it seems, be used to form justified beliefs in their conclusions. As he notes, however, the intuitive force of these examples is not strong. His second argument is that **Defeater Circularity** arguments suffer from some other vice, such as a dialectical vice, and we confuse this for their not being sources of justification. I'll come back to that point in the next section. For now I want to talk about how liberalism relates to **Defeater Circularity**.

If liberalism about M is true, and S can sometimes observe that she is using M, then she should be able to make the following argument.

The M Argument

1. p
2. M says that p

_____

C. So M got this one right.

By hypothesis, this could be a way that S comes to know that M is working. Since liberalism about M is true, she doesn't need to know that antecedently to using p to get the first premise. But the conclusion is obviously entailed by the premises. So it looks like it could be learned by learning the premises an doing a little reasoning. (A kind of liberalism that says that whenever S recognises which method she is using, that method is blocked from providing support, would not licence this reasoning. But that's a kind of liberalism that doesn't seem particularly plausible.)

But the M argument does exemplify **Defeater Circularity**, at least if the kind of liberalism that is true if the correct theory about M is not *too* radical. If S got evidence against the conclusion, that would block the support that M would provide for p. And that would mean the first premise was *undermined*. So we have the conditions needed for **Defeater Circularity**. So if some not-too-radical form of liberalism is true, then some arguments that exemplify **Defeater Circularity** can generate knowledge, and are in that sense not viciously circular.

This seems to be relevant to the debate over disagreement. Views other than the equal weight view of disagreement are alleged (by Christensen and by Elga)

to licence unreasonably circular arguments. But it's hard to see how we could get anything more than an argument which exemplifies **Defeater Circularity** out of equal weight violations.

   Let's take a simple case of disagreement. *S* and *T* both watch a court trial, and hear all the evidence presented. The evidence is a strong case that the defendant is guilty, though not so strong that no rational person could think otherwise. (That's not to say that thinking otherwise is rational; believing the defendant is not guilty in this case might be the kind of irrational act that rational people sometimes do.) After hearing all the evidence, *S* judges that the defendant is guilty, and *T* judges that the defendant is not guilty. *S* had no antecedent reason to believe that she and *T* were anything other than epistemic peers, but she isn't moved by *T*'s disagreement to change her mind about the case. (This is the equal weight violation.) Now she'll be able to make the following little argument to herself. (Assume the first premise is established by the evidence.)

   1. The defendant had means, motive and opportunity, and no one else had even two out of those three.
   2. So the defendant is guilty. (From 1)
   3. *S* judged that the defendant is guilty, and *T* judged that he wasn't guilty.
   4. So *S* is right and *T* is wrong about whether the defendant is guilty. (From 2, 3)

Now it isn't at all clear to me what is meant to be circular about that argument. But maybe if we include some extra premises about how the evidence appears to *S* and *T* we'll get a hint of circularity.

   1. The defendant had means, motive and opportunity, and no one else had even two out of those three.
   2. That evidence struck *S* as strong evidence the defendant is guilty, but the other evidence suggested to *T* that the defendant was not guilty.
   3. So the defendant is guilty. (From 2)
   4. *S* judged that the defendant is guilty, and *T* judged that he wasn't guilty.
   5. So *S* is right and *T* is wrong about whether the defendant is guilty. (From 3, 4)

Now we can see some kind of circularity. The support that 2 provides to 3 is seriously undermined if 5 is false. After all, if *T* and not *S* is who to trust on this question (as would be the case if 5 were false), then 2 is not evidence for 3. So if *S* reasons this way, and if using arguments that exemplify **Defeater Circularity** is bad, then *S* reasons in a badly circular way.

But there are two obvious problems with this attribution of viciously circular reasoning to *S*. First, it isn't clear why *S* has to go through this detour into psychological premises. (Timothy Williamson (2007, Ch. 7) has much more on this point in his discussion of the perils of psychologising the evidence.) Second it isn't clear that using arguments which exemplify **Defeater Circularity** is a bad thing. Indeed, if liberalism about some methods is true, then it is permissible some of the time. So I'm still very unsure how we are supposed to get vicious circularity out of violations of equal weight.

9.3 Is Defeater Circularity bad?

Here are four arguments, or perhaps approaches, that we could use to show that exemplifying **Defeater Circularity** is not bad. Remember that by 'bad' here, we mean that a person couldn't use the premises as the basis for their belief in the conclusion, or come to know the conclusion by reasoning through the premises.

Direct

1. This particular argument, e.g., **Moore**, exemplifies **Defeater Circularity** and is not bad.

C. Some argument exemplifies **Defeater Circularity** and is not bad.

Direct Method

1. Liberalism is true about some particular method of forming beliefs, e.g., perception or testimony or memory.
2. As argued above, if liberalism is true about some method of forming beliefs or other, then some arguments exemplify **Defeater Circularity** but are not bad.

C. Some argument exemplifies **Defeater Circularity** and is not bad.

Undermining

1. Any reason to think that exemplifying **Defeater Circularity** is bad is really, on closer reflection, a reason to think that exemplifying **Defeater Circularity** entails some other less than desirable feature.

C. We have no reason to deny that exemplifying **Defeater Circularity** need not be bad.

Indirect Method

1. Liberalism is true about some method of forming beliefs or other, though we aren't necessarily in a position to know which method it is.
2. As argued above, if liberalism is true about some method of forming beliefs or other, then some arguments exemplify **Defeater Circularity** but are not bad.

———————————

C. Some argument exemplifies **Defeater Circularity** and is not bad.

I think if you look closely enough, you can find all four arguments in Pryor, though in a way he has less to say about the fourth. So I'll concentrate my remarks here on that argument, after a brief comment on each of the first three.

I think the prospects of coming up with a particularly compelling instance of an argument that exemplifies **Defeater Circularity** but which is not bad are poor. I certainly wasn't greatly moved by Pryor's own purported examples, though of course I do believe that some examples exist. The problem with trying to find such examples is that we are faced with a squeeze from two directions. On the one hand, if the method used to form belief in the premises is too complex, then it will be arguable that rebutting the conclusion will defeat the premises. On the other hand, if the method is too simple, it will be arguable that the argument doesn't really *generate* warrant for the conclusion, but more reveals that you were already warranted. On this point, see the arguments in, for example, White (2006). The position White ends up with is, in effect, conservativism about perception, combined with it being a priori that perception generally works. I do think that there are arguments that exemplify **Defeater Circularity** and which are not bad, but I think finding one involves squeezing an example into a fairly tight window. Even if you succeed, I think intuitions around here are too coarse to tell that you really have squeezed into the window.

The same will probably be true of the second kind of argument, the one I called **Direct Method**. It's very hard to show that we should be liberals about any particular kind of method. It would take us way too far afield to go through all of the arguments that Pryor gives for liberalism about perception, or which, say, C. A. J. Coady (1995) gives for liberalism about testimony. (Coady doesn't use that language, because his work pre-dates Pryor's, but I think describing his view as liberal is fair.) It would be much too big a digression from our main topic here to seriously investigate whether Pryor's theory of perception, or Coady's theory of testimony is correct. In any case, we should be able to figure out foundational questions about the nature of **Defeater Circularity** without first solving all the problems concerning perception or testimony. So I'll set this approach to one side.

Pryor spends a little bit of time at the end of his paper on the third kind of argument, arguing that once we distinguish persuasiveness from other goodness in the sense relevant here, we'll see that there is no intuitive support for saying that arguments which exemplify **Defeater Circularity** are always bad. The short version of what he's saying is that there's an important distinction between learning and persuasion. If an argument exemplifies **Defeater Circularity**, it might not be particularly *persuasive*, particularly to a person who antecedently believes its conclusion to be false. But that doesn't mean that it isn't a good argument.

There are some interesting distinctions here about ways in which an argument might or might not be persuasive.[2] One way an argument could be persuasive is if it could reasonably convince someone who antecedently believed that its conclusion was false. That's not an impossible standard to meet. Gödel and Gettier produced such arguments, for instance. But it is a very high standard. A more moderate standard is that an argument is persuasive if it could be used to convince an open-minded person of the truth of its conclusion. And we can imagine yet more classifications; for instance, whether the argument could convince someone who started with antecedent doubts about the conclusion that the conclusion is, indeed, true.

Pryor thinks that Moore's argument has at least some dialectical flaws. (But note he also thinks, rightly, that the considerations in his sections 5 and 6 show that the argument also has some under-appreciated dialectical virtues.) In some interesting senses, it isn't a persuasive argument. If we were doing a longer seminar on dialectics, we might be interested in going over just which categories these are. But I think the general point that Moore's argument isn't particularly *persuasive* in the way that many good philosophical arguments are should be clear. The point Pryor wants to make is that we shouldn't confuse dialectical effectiveness in just this sense with epistemological value.

This connects to the discussion of evidence neutrality in Williamson (2007). Williamson thinks that too many philosophers are tacitly committed to the view that when we are in a philosophical dispute, we should only rely on propositions that are accepted by all sides. And he thinks that is false. He agrees that if we are trying to convince someone that p, we should start with premises they accept. But when we are trying to simply figure out whether p, perhaps so we can decide whether it is a good idea to convince others that p, we should start with anything we know, whether or not our friends accept it. Williamson thinks that much of the attention paid to psychological evidence, e.g., it seems very plausible that p, is due to a mistaken commitment to evidence neutrality. Even if we can't all agree, for example, that mountains exist, we can all agree that mountains *seem*

---

[2]There is a much more detailed discussion of how to evaluate arguments in Oppy (2006).

to exist, so that's what we start with. And he thinks this encourages scepticism, since it is hard to get from these psychological premises to claims about the mind-independent world.

These are impressive arguments, but I think there are reasons to be cautious of attempts to cleave epistemological and dialectical standards so far apart. I think there's a plausible case to be made that we shouldn't rely on arguments that can't meet some minimal standards of persuasiveness. My reason for thinking this is essentially a balance of risks consideration. It's true that sometimes if we simply take as given some things that people around us reject, we will come to new knowledge. Indeed, some of the best advances in the history of science come from just this route. But most people who take as given things that many people around them reject aren't super-scientists, they are kooks. At least when it comes to our *premises*, our friends are a useful epistemic check on our potentially irrational exuberance.[3] So I'm not convinced we can explain away intuitions that Moore's argument is a bad argument quite this easily. Perhaps it is true that our intuitions really track properties that are, in the first instance, dialectical not epistemological. And perhaps that could undermine our initial reluctance to classify Moore's argument as a good argument. But if there are connections between dialectical virtues and epistemological virtues, as I suspect there are, that won't get us out of the worry that Moore's argument is still a bad argument.

In contrast to all this, I think the case for **Indirect Method** being a good argument is strong. The argument for its key premise is a scepticism-avoidance argument. Call someone an **extremist** if they are anti-liberal about all methods. The Pyhronnhians are one kind of extremist, since they are conservative about every method. But extremists come in more varieties. Someone who thought that for any method, either radicalism or conservatism is true of that method is an extremist in my sense. Here's the key premise in the argument for liberalism:

- Extremism implies external world scepticism

Since external world scepticism is false, if the key premise is true, extremism is false.

Why might we believe the key premise? One strong reason, I think, is the long-running failure of anyone to come up with an extremist anti-sceptical response to external world scepticism. To my mind, the only really viable option here is the inference to the best explanation approach that we see in Russell

---

[3]As an aside, I might mention that I don't think the same is true of rules. If, say, *reductio* is a good rule, we can appeal to it in reasoning even if our friends disagree. This is a big part of how I think we can avoid some of the sceptical conclusions Williamson thinks follow from evidence neutrality. But this is taking us too far afield.

(1912/1997, Ch. 2), Jackson (1977), Vogel (1990), BonJour (BonJour and Sosa, 2003) and other internalists. In my terminology, that theory is radical about introspection, logic and inference to the best explanation (IBE), and conservative about everything else. But we then argue that we can use IBE to justify lots of the methods we regularly use, such as sense perception, memory and testimony.

Perhaps you think this kind of view can be made to work; my hopes for this project are dim. Let's just note one problem, one boldly conceded by BonJour. Since most humans have not justified their use of perception, etc by IBE, it follows that most people do not have (doxastically) justified beliefs. I think that's implausible on its face, and I think it's symptomatic of a deeper problem. I think figuring out, or even being sensitive to, the quality of different explanations of the way the world appears is cognitively downstream from the kind of simple engagement with the world that we get in perception.

Here's a quick recap of what I've argued here. Anti-scepticism implies we should be liberals about some method or other. If the only methods we could use are those about which radicalism is true, we wouldn't have enough resources to argue that methods of learning about the external world *work*, in the sense that's relevant here. So if conservatism were true about all other methods, we could only ever use the methods about which radicalism is true, and that leads to external world scepticism. Since external world scepticism is false, we have to be liberals about *something*. But liberalism implies that some good arguments exemplify **Defeater Circularity**. So whatever's wrong with Moore's argument, or with arguments which purportedly provide 'easy knowledge', it isn't that they exemplify **Defeater Circularity**.

Chapter 10

Wright on Entitlement

Crispin Wright (2004) offers a number of suggestions for how to respond to classical sceptical arguments. As Wright notes, this isn't going to be a totally *happy* response to scepticism. Indeed, he's going to concede several points to the sceptic. But ultimately, he wants to argue that part of the appeal of scepticism turns on a subtle conceptual confusion. We must distinguish, says Wright, between:

- having **justification** to **believe** a proposition; and
- having **warrant** to **accept** a proposition.

A number of sceptical arguments – Wright stresses the Cartesian and Humean arguements – makes heavy use of the fact that it is hard to see how to *justify* various 'cornerstone' propositions. The argument then continues that it is hard to see how to justify various everyday beliefs with justifying the cornerstones. So the argument concludes that our everyday beliefs are not justified after all.

Wright's response is that it isn't really true that the cornerstones need to be justified. All that matters is that they are warranted. Indeed, we don't even need to be warranted in believing them, we just need to be warranted in accepting them. And this warrant can be grounded in considerations that might be different to what classical epistemologists have typically focussed on.

10.1 Belief and Acceptance

A crucial feature in Wright's theory is the distinction between belief and acceptance. It's very easy to see what this difference comes to on a functional theory of belief. Hopefully the functional characterisation will carry across to non-functionalist theories of belief.

I think Wright thinks that his distinction mirrors in some sense the ordinary uses of the terms 'belief' and 'acceptance'. At the very least, he responds to a suggestion that his usage and the ordinary usage come apart. If he is making such a claim, I think it's pretty implausible. Personally, I'm sceptical that philosophers'

use of 'belief' runs particularly close to the ordinary language use of 'belief'. Whether that's true or not, 'acceptance' seems to be even further from ordinary usage. So I think it's safest to stick to the technical, functional characterisation of these states.

In textbook presentations of functionalism such as Braddon-Mitchell and Jackson (2007), mental states are individuated in terms of three kinds of functional roles:

- Input conditions;
- Internal connections; and
- Output conditions

The point of the second clause is to rule out the possibility that a giant lookup table, like Ned Block's 'Blockhead' (Block, 1978), could properly have beliefs or desires. In Wright's way of setting things up, we distribute the 'internal connections' over other inputs and outputs. So instead of saying that a belief that p is something that's typically produced by a belief that $q \wedge p$, and typically produces, in conjunction with a belief that $p \to r$, a belief that r, we say that one possible input condition for having a belief that p is antecedently having a belief that $p \wedge q$, and one possible output condition for that belief is that, combined with a belief that $p \to r$, a belief that r is produced. In other words, we insist that the inputs and outputs are not just worldly inputs and behavioural outputs, but can include other mental states.

Having said that, we can now say what acceptance is. To **accept** that p is to be in a state which fills the functional role associated with the *output* conditions for believing that p. So acceptance requires acting as if p is true, it requires inferring things that you'd naturally infer from p and so on. But it doesn't require, even in the weak, general way that beliefs require filling this functional role, anything by way of input conditions.

As Wright stresses, this is still meant to be a fairly holistic characterisation of a state. Just because you do *some* things that you would do if p were true, doesn't mean you accept that p. You've got to do *all*, or at least the vast majority, of things that you'd do if p were true in order to accept that p. This is why we don't count as accepting that a plane will crash when we buy insurance on the flight, for example.

I don't think Wright says this explicitly, but thinking about acceptance this way suggests a reason that the norms for acceptance could be very different to the norms for belief. You might think that the norms for belief are tied very closely to the input conditions for belief. In general, a good belief is one whose actual input conditions are close to the canonical input conditions for being a belief

(with that content). Like for many other things (e.g., jokes), there is a continuum from being a bad belief to being a non-belief. Presumably the output conditions for belief also matter for the appropriateness of belief, but they are a smaller consideration.[1] And they're smaller because in general, having appropriate input conditions will imply having appropriate output conditions.

Now acceptances, by definition, have no input conditions. That is, whether a state is an acceptance that p is independent of how the state came to be formed. So perhaps we should also say that whether a state is an appropriate acceptance that p is independent of how it is formed. What it takes to appropriately accept that p is to be such that it's appropriate to act (and infer) as if p is true. And the conditions for that are, most of us think, more broadly pragmatic than evidential. So that's why we can have non-evidential warrant, or what Wright calls 'entitlement'.

10.2   Strategic Warrant

One of the central ideas in decision theory (or game theory) is that some options are better than others no matter how everything else turns out. These options are called 'dominating options', and it's a central tenet of decision/game theory that you should choose dominating options. Wright's first idea for how to get a notion of entitlement is that accepting some anti-sceptical claims is a dominating option.

Orthodox decision theory is fairly subjectivist about welfare. So when a decision theorist says that φ is better than ψ no matter how the world turns out, what she means is that φ does better by the agent's own lights. Typically, by that they mean that no matter how the world turns out, the agent *prefers* to have done φ to having done ψ.

This can't be exactly what Wright means. Imagine I strongly prefer to accept the proposition that I'm the handsomest man alive. Accepting this makes me incredibly happy, and I prefer that to any other negative effects that might come from accepting an implausible proposition. Then relative to my preferences, accepting that I'm the handsomest man alive will be a dominating option. But I don't think Wright wants to say that I have strategic warrant for accepting that I'm the handsomest man alive, independent of whatever evidence I have for or against this.

Now this isn't a problem for Wright's actual view, because he (a) doesn't endorse the kind of subjectivism about value that drives the counterxample, and

---

[1]My view is that the reason there is some interest-relativity in epistemology is because a justified belief must have a justified output condition, and the output conditions are interest-relative. But that's a story for a very different time.

(b) he puts restrictions on what kinds of projects 'count' for generating strategic warrant. (It's probable that (a) and (b) are really versions of the same point, but I wanted to make both of them explicit.) Wright says that we get a strategic warrant to accept *p* if we have no evidence against *p*, and accepting *p* is a dominating option relative to an *essential* project. Pretty clearly, my project of making myself feel better by accepting claims about my looks is not *essential*. Still, what makes a project essential is not in a traditional sense an epistemological matter. The Cartesian Meditator perhaps need not engage in the inductive project. It's in part because our lives are not like those of the Meditator that we are entitled to get going on that project.

The main application of this idea is to induction. The first thought is that the project of making predictions about the world is essential to living the kind of lives we live. That sounds plausible. The second thought is that accepting the Uniformity of Nature hypothesis (as Mill put it) is a dominating option with respect to that project. If the thesis is true, then accepting it gives us a good foundation for later inductive activities. If the thesis is false, then induction is hopeless, so it's no real loss to have accepted it. So we have a strategic warrant to accept the Uniformity of Nature hypothesis.

I've simplified a fair bit here, but I think I've said enough to let us see a couple of reasons for doubting this approach to induction will work. Both the project under discussion, and the Uniformity of Nature hypothesis, are under-described. Filling in the details creates difficulties for Wright's view.

First, we might wonder what exactly the inductive project is. If it's the project of forming true beliefs, then it isn't clear that accepting the Uniformity of Nature hypothesis is dominating. In some worlds, we form more true beliefs by studying chicken entrails than by using scientific methods. Thankfully our world is not like that, but that doesn't matter for the purpose of determining what a *dominating* option is. If the project is the 'Jamesian' project of forming true beliefs and avoiding false beliefs, then it is even less clear that accepting the Uniformity of Nature hypothesis is dominating. If the world is anything like the way Hume feared it to be, then the approach of forming no inductive beliefs is preferable to accepting the Uniformity of Nature hypothesis and moving on from there. Neither approach yields much by way of true beliefs, but the cautious approach at least avoids error.

I think the best thing to say here is that the project is the 'Williamsonian' project of forming knowledge about the unobserved world. Accepting the Uniformity of Nature hypothesis could lead to such knowledge, arguably no other approach could, so we should accept the Uniformity of Nature hypothesis. I'm not sure if we can really read that 'Williamsonian' project into Wright's text, but

it seems the best way to make sense of the idea that going in for induction is dominating.

Second, as Wright notes, the 'grue' problem is really challenging here. The issue is that there simply isn't any such thing as *the* Uniformity of Nature hypothesis. Indeed, nature is bound to be uniform in some respect or other. The question is just which predicates it is uniform with respect to. But this suggests there can't be any option which is really dominating. The best approach, from the point of view of induction, is to accept that nature is uniform with respect to some class C of predicates, where C is the class with respect to which nature is actually uniform. But we can't *a priori* know what that class is, and it differs in different worlds. So I don't really see how strategic warrant can really help with induction.

10.3   Entitlement of Cognitive Project

The next form of entitlement is 'Entitlement of Cognitive Project'. The entitlements of a cognitive project include p if two conditions are met:

1. There is no reason to believe that $\neg p$; and
2. Inquiring into p would require a new project with a further presupposition that is no more secure than p, and inquiring into that would launch a new project with yet another presupposition, and so on ad infinitum.

Wright says that these entitlements are meant to be different to Wittgensteinian hinges because they are project-relative. I'm not exactly sure how that is supposed to play out. Let's work through an example to see what we can find.

I'm trying to work out when the Battle of Hastings was. That's my project. (It isn't an essential project, so it isn't clear I'm entitled to its presuppositions, but set that aside.) I look up Wikipedia, and it says 1066. Now a presupposition of that project is that Wikipedia is realiable. But that doesn't look like a presupposition that sets off an infinite regress. I can investigate that by looking up reports on Wikipedia's reliability, and arguably the presuppositions of that inquiry are more secure than the presupposition that Wikipedia is reliable.

So I don't quite see what Wright is aiming for here. If the presupposition of a particular project is not a Wittgensteinian hinge, then it seems in general to me that it will be possible to investigate that project's presuppositions using a more secure footing. To be sure, once we get to the hinges, it seems that we must stop. But I'd like to see more of an argument that there is some presupposition of a project that is both more narrow, i.e., more project-relative, than one of the hinges, and just as secure as the hinges are.

I also confess that I didn't understand exactly why this kind of entitlement is supposed to not extend to propositions about the external world. For one thing, I'm not sure what's at issue here. If we're entitled to accept procedural claims like *My senses are generally reliable*, and such procedures immediately issue in claims about the external world, then it's not at all clear why it matters whether we're really entitled to these ontologically loaded claims.

And I don't at all see the force of the reason Wright gives for saying entitlement of cognitive project does not extend to the external world. I think there's meant to be a disanalogy between these cases:

- We start a project with a 'procedural' presupposition, e.g, that our senses are reliable. We could investigate that, but the investigation would use a similar, or at least similarly secure, presupposition.
- We start a project with an 'ontological' presupposition, e.g., that the external world exists. We have, says Wright, no idea how we could justify that, so the project of investigating it doesn't launch anything, let alone an infinite regress.

But the second claim seems to be simply false. It's easy to imagine a way to investigate such an ontological presupposition. We just use one of the 'procedural' presuppositions. For instance, we investigate whether the external world exists by opening our eyes and looking around. Of course, that project has other presuppositions, which are (to put it mildly) no more secure than the original presupposition that the external world exists. But that seems to count *in favour* of the idea that the existence of the external world could be a presupposition of a cognitive project.

I think I must be missing something here, because I really can't see how to generate the distinction between these kinds of presuppositions that seems so important to Wright.

10.4   Entitlement of Rational Deliberation

This section of Wright's paper goes by very quickly. This is odd, since the conclusions it seems to draw are considerably stronger than what he draws elsewhere in the paper. Or at least they seem much stronger to me. If I'm reading it right, the argument is as follows:

1. The project of rational (practical) deliberation requires having reasonable beliefs about one's values and about what will happen if you perform various actions.

2. The project of rational (practical) deliberation is an essential project for living the kinds of lives we live.
3. We are entitled to trust in any presupposition of any essential project.
4. So we are entitled to trust our introspection, and in our ability to form reasonable beliefs about subjunctive conditionals about what will happen if we do various things.
5. We could only form reasonable beliefs about these subjunctive conditionals if inductive reasoning, broadly construed, is rational.
6. So we're entitled to believe that inductive reasoning, broadly construed, is rational.

The same kind of worries that we raised about 'grue' earlier will apply here. At most this shows us that some or other inductive practice is rational. But it might be completely unknown, and unknowable, which one that is, in advance of knowing which predicates are projectible. And there's no obvious argument along these lines that a particular set of claims about projectability is among our entitlements. (Indeed, concluding that something like this is among the entitlements seems like the kind of thing that Wright wants to deny.)

I'm also not convinced that step 5 is well supported. Imagine a creature that was capable of instrumental reasoning, and which was born with an innate stock of true beliefs about various subjunctive conditionals of the form *If I do this, that will happen*. That kind of creature would, I think, be capable of performing the kind of practical deliberation of the kind necessary for living something like the kind of life we live. But it isn't clear that the world needs to be arranged so that inductive reasoning will be reliable. That's because the subjunctive conditionals need not come from reasoning; they could be innate.

The example I sketched in the last paragraph isn't meant to be completely absurd. I think a lot of creatures have a (weak) capacity for instrumental reasoning, and a lot of innate knowledge about act → outcome subjunctive conditionals. So I don't think we can get directly from the possibility of instrumental reasoning to the rationality of induction, in the way Wright seems to suggest. (Ironically, this is because Wright seems to conclude that we must be able to make a certain kind of rational inference, when all his evidence shows is that we must be able to rationally act as if we had rationally drawn that conclusion.)

But maybe there's a more roundabout argument that Wright could give. It might be arguable that it is impossible for there to be much innate knowledge except in a world where induction is rational. So if the first four steps of the argument show that either induction is rational or we have a lot of innate knowledge, then we can argue by cases that either way we're entitled to inductive reasoning.

The problem with this argument is that it isn't altogether clear why we should believe the claim about innate knowledge. It's surely a nomological necessity that innate knowledge presupposes the kind of regularities in nature that also support induction. But nomological necessities are not the kind of thing we should be in the business of relying on in sceptical scenarios. It's *logically* possible that we have innate knowledge because a not-so-evil demon gave it to us, even though no Uniformity of Nature hypothesis that we could come up with would be true. So I don't exactly see how to make the last step of this argument work.

10.5   Liberalism or Radicalism?

Last chapter we distinguished between radical and liberal approaches to epistemology. Both radicals and liberals think that we can get justified beliefs via some method without having antecedent justification for thinking the method works. Wright thinks this too, though he has a qualified version of it. In a lot of the cases that liberals and radicals care about, Wright thinks we have antecedent **warrant** to accept that the method works, but this will be based on entitlement, not justification. So is he, in our terminology, a radical or a liberal?

A first reading of the paper would suggest that he is very much a liberal. Every one of the kinds of entitlement that Wright posits is qualified by a clause like "as long as there is no evidence to the contrary". So it looks like he thinks entitlements are defeasible, just like liberals accept and radicals deny.

But on second look, things aren't so clear. If we look at the kinds of things that Wright says we are warranted in accepting, it isn't clear what *could* count as evidence against them. It's not clear that we really could get evidence, for example, that we are bad at processing evidence. If it seemed to us that we had such evidence, it would be self-defeating. So it's tempting to think that there is no real way that the entitlement could be defeated. And that makes it look like a more radical position.

Or consider the position he takes on induction. Again, it's always part of the official position that entitlements to inductive reasoning are defeasible. But I think there's less to this in practice than there seems. What the entitlement arguments show, if anything, is that we're entitled to use some inductive method or other. They don't, I think, show that one particular method (i.e., one particular choice of projectible predicates) is what we're entitled to. But once we note that the entitlement is so weak, it's not clear it is defeasible. It isn't clear that there is anything that could count against the rationality of some inductive method or other, as opposed to being evidence that some predicates are not really projectible. So I'm inclined to think that, despite the care taken to note that all the

entitlements should be defeasible, to think that some of the positions will end up being more radical than liberal.

10.6   Closure Arguments

You might think there is a simple argument against Wright's position on entitlement. It seems to be a consequence of his view that we can have evidential warrant for 1 and 2, but not for 3, which is clearly entailed by 1 and 2.

1. It appears to me that there's a desk here.
2. There is a desk here.
3. Things are as they appear here.

That is, there could be a situation where we are *entitled* to 3, but have actual evidential justification for 1 and 2. But isn't this absurd, since 1 and 2 together entail 3?

It isn't absurd, because there is a reason to think that *evidential* justification is not closed under entailment. Note that particular kinds of justification frequently aren't closed under entailment. For example, I could have testimonial justification for 4, but not have *testimonial* justification for 5.

4. Barack Obama is in Indonesia.
5. Either Barack Obama is in Indonesia, or Barack Obama is not in Indonesia.

If I can simply deduce 5 as a matter of logic, then I don't have testimonial justification for it. Indeed, perhaps I don't really have evidential justification for it. I am warranted in believing it, but the warrant derives from the proof, not from any evidence I have.

In short, what's really plausible is closure under (known) entailment for warrant, not closure under (known) entailment for any kind of warrant. To the extent that evidential justification is a kind of warrant, we shouldn't assume that it is closed under entailment.

10.7 Why Only Entitlement?

Having said all that, it isn't clear just why Wright thinks that we don't get *justification* for the propositions he thinks we're entitled to accept. It's clear why he thinks we don't get *evidential* justification for these propositions. The nature of the propositions is such that accepting them is a necessary precondition to getting evidence, or at least to getting evidence in the way that we're currently investigating the world. But unless we think only evidential justification amounts to genuine justification, that won't really settle the question.

Indeed, you might think that there are a whole bunch of beliefs we have that are justified, but not on the basis of evidence. It isn't clear what our evidence is that, for example, Modus Ponens is valid, or that 2+2=4. So you might think these propositions are things that can be justified but not on the basis of evidence. Wright's reply, I take it, is to extend the notion of entitlement to cover these as well.

Once we have distinguished justification from evidential justification, or at least raised the possibility that something other than evidence could justify, the argument that justification need not be closed under competent deduction looks a little weaker. It is clearly true that deductions from things that are evidentially justified need not be evidentially justified. But it isn't clear why we aren't justified in believing things we competently deduce (and know that we've competently deduced) from things that we are justified in believing?

There are two things we might say in reply to that somewhat rhetorical question. The first is that we might put a restriction on competent deduction that the premises are not just known, but that we're in a position to claim knowledge of them. In that case, since Wright thinks that we can't claim knowledge of various everyday propositions, we couldn't deduce from them that, say, our perceptions are reliable. But on reflection, Wright probably wouldn't want to say this. The problem is that this restriction on inferential knowledge is too strong - it would rule out almost all inferential knowledge!

The other thing we might say, and one more in keeping with Wright's views, is that justification does not always 'transmit' across inferences. In particular, although we can know that it appears to us that p, and that p, and that therefore things are as they appear, the justification for the premises could not transmit to the conclusion. To really assess this would require re-arguing all the material from the previous chapter. My feeling is that Wright is too quick to restrict transmission across inferences, and Pryor's objections to Wright here are correct. But if Wright is correct about his dispute with Pryor, then he would have grounds for saying that we're not justified in believing everything we're entitled to accept.

Chapter 11

Contingent A Priori

11.1 The Humean Argument for Scepticism

Assume, for *reductio*, that we can have knowledge that goes beyond its basis. Assume also that in at least one such case, the basis consists of propositional evidence. Let H be such a piece of knowledge, and let E be the evidence the agent has, immediately before she comes to believe H. By hypothesis, she's at least in a position to know H; it is just a matter of forming the belief on the right basis. We'll use $E \supset H$ to stand for the material conditional *If E, then H*, i.e., for $\neg(E \wedge \neg H)$. Now there's a case to be made for each premise in this argument.

Humean Argument for Scepticism

1. The agent is not in a position to know $E \supset H$ *a priori*.
2. The agent is not in a position to know $E \supset H$ *a posteriori*.
3. It's not the case that the agent is in a position to know $E \supset H$, but not in a position to know if either *a priori* or *a posteriori*.
4. If the agent is in a position to know H, she's in a position to know $E \supset H$.

———————————

C. The agent is not in a position to know H.

But this contradicts our original assumption, that the agent can immediately infer H given what she antecedently knows.

I call this a Humean argument because of its resemblance to Hume's argument for inductive scepticism. Hume offers, or at least has suggested to posterity, a dilemma. The reliability of our inductive practices can't be known either deductively or experimentally. As we'd put it, it can't be known either *a priori* or *a posteriori*. But it must be known. On the standard interpretation, Hume makes this step because he is an inductive conservative, and thinks that if we reason inductively, we must have antecedent reason to believe that this reasoning is sound. But it doesn't matter for the Humean argument whether this conservatism about

induction is either true or Hume's view. Because all we need to motivate premise four is, roughly, that the reliability of induction is know *ex post facto*, not that it is antecedently known. In particular, we just need to know after the fact that $E \supset H$ is true, and we do assuming closure. So we have a real challenge for the anti-sceptic.

I've tried to frame the argument here so that it poses a threat even to the person who accepts E=K. I've set up the puzzle so that the agent doesn't yet believe, and hence doesn't yet know, H. But they have everything they need to know H, and hence presumably to know $E \supset H$. Since the agent doesn't know H, even the E=K theorist should deny that H is part of their evidence. So we can't say that premise 1 fails because $E \supset H$ is a simple tautology. Really anyone who thinks we can know something without having evidence which entails it should be moved by premise 1, and that includes the E=K theorist.

It might not be obvious now why premise 2 is true, but we will see a number of arguments for it as we go along. Many of these can be profitably thought of as refinements, or developments, of the anti-circularity arguments we find in Hume. The argument we'll spend the most time on, the probabilistic argument, may look quite different to anything actually in Hume. But I think, on reflection, we'll see it has some similarities.

There are a few reasons why the Humean argument is a more interesting argument than the other sceptical arguments we've looked at so far. I'll briefly mention three.

First, it simply isn't obvious which premise is false. In particular, it isn't obvious whether our knowledge of things like $E \supset H$ is *a priori* or *a posteriori*. Several epistemologists think it is obvious what the right answer is; not all of those epistemologists agree on the answer. This isn't a sufficient condition for a sceptical argument being interesting, but it helps.

Second, there are various philosophical positions that seem to imply each of the premises. I think any kind of theory that combines hostility to the contingent a priori with a broadly Bayesian theory of empirical learning can imply both of the first two premises. I also think the kind of theory we see in Chalmers (2002); Chalmers and Jackson (2001), which ties a priority closely to modality could, at least with some plausible extra assumptions, be used to motivate each premise of the Humean argument. We'l return to that issue toward the end of these notes. Since the conclusion of the Humean argument is false, I take that to be a problem for those philosophical theories.

Third, as Pryor (2000) notes, this argument seems to generalise to an argument for scepticism about justified belief. Many sceptical arguments do not. It's

not very plausible that justified belief must be sensitive, for instance, so a sceptical argument based on sensitivity won't give us much reason to be sceptical about justified belief. But the best arguments that $E \supset H$ can't be known either *a posteriori* or *a priori* seem to generalise to arguments that $E \supset H$ can't be justifiably believed, either *a posteriori* or *a priori*. So we can conclude, if those arguments work, that the agent can't even justifiably believe H, let alone know it. That's a really interesting conclusion I think, and a reason to spend some time on this argument.

11.2   The Probabilistic Argument for Premise 2

You can take a lot of the arguments that Roger White (2006) offers against dogmatism to really be arguments for premise 2 of the Humean argument. Indeed, in many ways they are better arguments for premise 2 than they are as anti-dogmatist arguments I think. (I think as arguments against dogmatism, the probabilistic examples White uses run afoul of the dogmatists' insistence that the default entitlement to believe things are as they appear is very easily defeated.) Here's how I'd present the argument. Start with three assumptions.

Credences are Probabilities The agent's credence function Cr is a classical probability function at all times.

Updating is Conditionalisation If we write $Cr(H)$ for the old credence in H, and $Cr_E(H)$ for the new credence in H, then $Cr_E(H) = Cr(H|E)$.

Learning requires Credence Raising If an agent learns H by getting E, then $Cr_E(H) > Cr(H)$.

Now we rely on an easy to prove fact about the probability calculus.

$$\Pr(E \supset H|E) \leq \Pr(E \supset H)$$

There are a bunch of ways to prove this. Note first that if $\Pr(E) = 0$, then $\Pr(E \supset H) = 1$, and since all probabilities are less than or equal to 1, then $Pr_E(E \supset H|E) \leq \Pr(E \supset H)$. Note also that if $\Pr(E) = 1$, then conditionalising on E doesn't change any probabilities, so the inequality obviously holds there. So the only case remaining is when $\Pr(E)$ is strictly between 0 and 1. But in that case we can express both $\Pr(E \supset H|E)$ and $\Pr(E \supset H|\neg E)$ as a ratio. Now the proof makes use of the following easy (and important) theorem of the classical probability calculus, which I'll state without proof here.

$$\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|\neg B)\Pr(\neg B)$$

Let's apply that, substituting $E \supset H$ for A, and E for B.

$$\Pr(E \supset H) = \Pr(E \supset H|E)\Pr(E) + \Pr(E \supset H|\neg E)\Pr(\neg E)$$

Note that $\neg E$ entails $E \supset H$, so $\Pr(E \supset H|\neg E) = 1$ so $\Pr(E \supset H|\neg E) \geq \Pr(E \supset H|E)$. Hence

$$\begin{aligned}
\Pr(E \supset H) &\geq \Pr(E \supset H|E)\Pr(E) + \Pr(E \supset H|E)\Pr(\neg E) \\
&= \Pr(E \supset H|E)(\Pr(E) + \Pr(\neg E) \\
&= \Pr(E \supset H|E)
\end{aligned}$$

There's another proof of this which is a little less direct, but hopefully more explanatory. Note first one equivalence, which we can easily prove on the assumption that $\Pr(E) > 0$.

$$\begin{aligned}
\Pr(E \supset H|E) &= \frac{\Pr((E \supset H) \wedge E)}{\Pr(E)} \\
&= \frac{\Pr(E \wedge H)}{\Pr(E)} \\
&= \Pr(H|E)
\end{aligned}$$

The key move is noting that $(E \supset H) \wedge E$ is logically equivalent to $E \wedge H$, and logically equivalent propositions have the same probability. Now note these two obvious equations, which follow form the fact that the probabilities of exclusive, exhaustive propositions sum to 1, either unconditionally or conditional on any (consistent) proposition.

$$\begin{aligned}
1 - \Pr(H|E) &= \Pr(\neg H|E) \\
1 - \Pr(E \supset H) &= \Pr(E \wedge \neg H)
\end{aligned}$$

Finally, note these two equations, which will end up having the same numerator and different denominators.

$$\begin{aligned}
\Pr(\neg H|E) &= \frac{\Pr(E \wedge \neg H)}{\Pr(E)} \\
\Pr(E \wedge \neg H) &= \frac{\Pr(E \wedge \neg H)}{1}
\end{aligned}$$

Since $\Pr(E) < 1$ (by assumption), $\Pr(\neg H|E) > \Pr(E \wedge \neg H)$, so $1 - \Pr(\neg H|E) < 1 - \Pr(E \wedge \neg H)$, so $\Pr(H|E) < \Pr(E \supset H)$, so $\Pr(E \supset H|E) < \Pr(E \supset H)$, as required.

There is a more general result here, which I'll simply state without proving. In general, you can't raise the probability of a disjunction by conditionalising on the falsity of one disjunct. This shouldn't be too surprising a result. Perhaps oddly, this constraint doesn't hold in intuitionist probability theory; it matters that we are reasoning classically here. In intuitionist probability theory you certainly can raise the probability of $p \vee \neg p$ by conditionalising on the falsity of the first disjunct.

11.3   Reflections on Bayesian Proof

That $\Pr(E \supset H | E) < \Pr(E \supset H)$ is simply a theorem of the probability calculus. There's a closely related philosophical claim; namely that you can't learn H on the basis of E unless you antecedently know $E \supset H$. I think the philosophical result simply isn't true, and we can see it isn't true by thinking a little about the philosophical underpinnings of Bayesian theory.

It's helpful, I think, when dealing with probability based philosophical arguments to see what happens when we restrict all of the probabilities to 0, 0.5 and 1. (Sometimes it is even better to restrict everything to 0 and 1, but I think it helps to have the 0.5 option here.) And let's consider the case where E is simply the evidence that a certain system (perhaps a perceptual system, or a testifier), represents that H. Now let's see what happens when we add in the assumption that learning is always a matter of conditionalisation.

We want this to be a case of *learning* that H. So we want it to be a case where the posterior probability of H is 1. (Presumably a probability of 0 or 0.5 is not compatible with having learned that H.) Well the only way that can happen is if the prior probability of $E \supset H$ is 1. That's to say, we can only learn H on the evidence that a certain method outputs H if we antecedently knew that the method was reliable, at least with respect to H. That is, we have an argument for conservatism about all methods, or at least all empirical methods. (Even Bayesians note that the theory doesn't work super well for non-empirical methods, so let's assume it is silent on that.) But across-the-board conservatism is implausible. So the assumption of universal conditionalisation is implausible.

More generally, the idea that when we learn E, we should replace our old probability for H with the old probability for H given E is, in a good sense, the probabilistic equivalent of the idea that when we learn E, we should replace our attitude towards H with our old attitude towards $E \supset H$. That is, it is the probabilistic equivalent of the idea that learning always goes by modus ponens. We always learn that H by combining evidence E with a prior knowledge that $E \supset H$. But that's really implausible, especially for fundamental methods of inquiry.

The points I've been making in this section are, I think, just versions of points that Pryor makes in response to Bayesian criticisms of dogmatism. Bayesianism has a lot of assumptions built in to it, often very conservative assumptions. Those assumptions are plausible in some of the cases that Bayesianism handles well; they aren't very plausible in foundational epistemology.

11.4   Fallible A Priori Knowledge

We'll come back to other arguments about the *a posteriori* premise. But I wanted to look at arguments about the *a priori* premise in the Humean argument for now. Of course, it's a little hard to pull the two premises clearly apart, since some of the arguments for one or other premises are arguments from the falsity of the other premise. But we'll focus for a little on whether $E \supset H$ could be knowable *a priori* for a while.

I suggested this solution to the problem in Weatherson (2005). This wasn't an original suggestion; I learned about the idea from BonJour (1997) and Hawthorne (2002). But when I first presented that paper, I got as many blank stares in response as if I were defending modal realism. That happens to assistant professors making outlandish claims! But of course I wasn't the only one who thought there is a good argument for $E \supset H$ being knowable *a priori*In fact, as I went around asking people about the Humean argument, it turned out a lot of epistemologists thought that the *a priori* premise is the false one. So why the blank stares?

I think there's a simple explanation for this. $E \supset H$ could be false. Things look *exactly* like they would look if they were false. We're not used to the idea that *a priori* knowledge is fallible. If we assume infallibilism about the *a priori*, then presumably we can't know $E \supset H$ *a priori*.

But of course fallibilism is not very plausible in general for knowledge, and it isn't any more plausible, I'd think, for the *a priori*. Once we note that a belief is fallible it can seem counterintuitive to insist that it amounts to knowledge. Hawthorne notes this point, and also notes that it is true for *a posteriori* as well as *a priori* knowledge. And we don't respond to this by believing fallibilism in general.

11.5   Williamson's Method M

Hawthorne's discussion of the contingent *a priori* starts with a particularly complicated case, due to Timothy Williamson (1986). The case is of someone who reasons by Rule M:

M Given a proof from *Someone believes that p* to p, infer p.

Anyone who reasons (correctly) using M cannot form a false belief. To see this, assume that the person uses M to form the belief that p. Then someone believes that p. And there's a proof from that to p. So the belief must be true.

Hawthorne calls this property of methods 'hyper-reliability'. It feels more apt to describe it as a kind of infallibility to me, because that suggests the doxastic aspect of it. But either way, what's at issue is whether this property suffices for knowledge. And it is reasonably clear that it doesn't, at least as stated.

Hawthorne compares M to the method M3, described as follows.

M3 Given that p, infer p.

That's hyper-reliable, but clearly not in general knowledge producing. After all, it's not the case that whenever someone forms a true belief, they have knowledge. But this is a slightly unfair comparison. The point of M is not just that it is hyper-reliable, it is that it is *a priori* knowable that the condition in the antecedent of M is met. So it seems we could use M in *a priori* reasoning, in a way that we couldn't be in a position to use M3 (unless we had independent reason to think p was *a priori*).

I think what's going on with M is that two conceptions of basicness come apart. One conception is what we might call **normative basicness**. A transition rule is normatively basic if it transmits justification, but we don't need any further explanation in terms of justified beliefs about why it transmits justification. In particular, it doesn't transmit justification because we have a prior proof of its reliability. A different conception is what we might call **psychological basicness**. A transition rule is psychologically basic if it is a rule people follow 'directly', and not by following another rule. As Wittgenstein (1953) famously pointed out, at least some rules must be psychologically basic. Rule-following can't always consist in applying another rule.

Now it isn't always clear how closely these two concepts are tied. It might be that we are built to treat some rules as psychologically basic, even though their justification turns on the availability of justified belief in their reliability. Someone who thought that there was an *a priori* proof of the reliability of sense-perception, and who thought that availability of this proof licenced basic transitions from *Appears that p* to p, would think such a case existed. And it might be that there are transitions which are as a matter of fact normatively basic, but which humans usually require extra steps to carry out. Perhaps reasoning by *modus tollens* is like that.

It is very hard, I think, to make the case that M is normatively basic. If reasoning by M produces knowledge, it is because of the provability of M's reliability. On the other hand, Williamson's example requires, I think, the agent

to treat M as psychologically basic. If they reason to a conclusion using M, such as the conclusion *Someone believes something*, by explicitly stepping through the proof of M's reliability, they will be using indexical beliefs in their justification, and defeating Williamson's argument that there are non-indexical examples of the contingent *a priori*.

I don't think that there's anything wrong with the idea that there are psychologically but not normatively basic transition rules. But I'm a little doubtful that M is one of them. I think M, at least if it is understood as psychologically basic, is similar to cases where an agent has a prior guarantee in the truth of her beliefs. (C.S. Jenkins calls these 'truth fairy' cases; see for instance her Jenkins (2007), though there the truth fairy is a 'quirky goddess'.) So imagine that an agent is told that whichever horse she believes will win the next race will, in fact, be made to win that race. In fact, she comes to know that a genie will fix the race so that whoever she believes will win will in fact win. Knowing that, but without any evidence about the horses, she forms the belief that Horse 7 will win. What should we say about the agent?

One thing we should clearly say is that she can eventually come to knowledge this way that Horse 7 will win. If she observes that she believes that Horse 7 will win, and remembers the existence of the genie, she can knowledgeably infer that Horse 7 will win. What about before she makes that inference, i.e., when she simply forms the belief that Horse 7 will win, knowing that whatever belief she forms will be true? I think in that case she doesn't know that Horse 7 will win. Knowledge requires evidence (or basicness) and she doesn't have any evidence. I know from experience that people's opinions on this kind of case differ wildly, but it seems like non-knowledge to me.

And beliefs formed using M, where M is treated as basic, are the same way. It's true that once the agent comes to believe that, say, someone believes something, she can verify that that belief must be true. But the verification will involve some introspection, and so is not *a priori*. And before she does the verification, it isn't knowledge. So it is never a case of contingent *a priori* knowledge.

11.6 Hawthorne's Explainer

Hawthorne's 'Explainer' example is closer in spirit to other parts of the literature. In short, the issue in the Explainer case is whether ampliative inferences that we make in real life cases can also be made 'off-line', in suppositional reasoning. If they can, then we have an easy example of contingent *a priori* knowledge. Let *E therefore H* be an ampliative argument that preserves knowledge. Imagine that the agent supposes that E, infers H in just the way that she would if her actual evidence were E, and then discharges the asumption that E to derive $E \supset H$.

Since every step is sound, she comes to know $E \supset H$. Since the inference from E to H is ampliative, $E \supset H$ is contingent. And since the argument for $E \supset H$ involved no non-discharged empirical premises, the knowledge of $E \supset H$ is *a priori*. So it's a nice case of contingent *a priori* knowledge.

As Hawthorne says, the big question about this case is why it should be that the inference from E to H is justified in the scope of suppositional reasoning. Hawthorne suggests one answer which I think clearly does not work. He suggests, following Strawson (1952), that perhaps it is analytic that the inference from E to H is rational. That is, it is part of the meaning of 'rational' that an inference from E to H is licenced. But I can't see how this gets the desired result. The natural conclusion to draw here is that from E the agent can infer that H is rational to believe. But it isn't obvious that the conditional *If my evidence is E, then H is rational to believe* is contingent. So I'm not sure how this helps.

Schematically, we'd like to say that what justifies the inference from E to H in the suppositional reasoning is just whatever justifies it in regular reasoning. If you're a conservative about induction, that presumably will be the prior justification for $E \supset H$. (In which case the reasoning above to the conclusion that $E \supset H$ will not be particularly enlightening.) If you're a certain kind of rationalist about induction, then it will be the fact that E rationally supports H in the right kind of way. And given this kind of rational support exists, perhaps you'll find it plausible to conclude that $E \supset H$ is a piece of contingent *a priori* knowledge.

An interesting case is what to say if you're a reliabilist about induction. Let's say the inference from E to H is justified by the actually existing correlation between E and H. (Or, more likely, between two types of propositions, of which E and H are examples.) Then even in *a priori* reasoning, the correlation between E and H will still actually obtain. It won't be knowable *a priori* that the correlation obtains, but since the justifiedness of the inference is independent of the knowability of the correlation, that doesn't block the agent from inferring this way. So perhaps in this case $E \supset H$ will be a piece of contingent *a priori* knowledge, and it will be contingent that $E \supset H$ is knowable. Of course, for any contingent truth, it is contingent that it is knowable because no falsehood is knowable. But in this case there will be a second source of contingency, namely the contingency of the correlation which justifies inferences from E to H.

All of this is to say that it isn't completely obvious why explanations of when ampliative reasoning works, can't generalise to explanations of how we get contingent *a priori* knowledge. But nor is it obvious that these arguments will generalise. I'm personally inclined to think that at least some inferential transitions are only licenced in 'on-line' reasoning, because I think the alternative is a kind

of paradox. But we'll get to that point later. The initial point is that there is a *prima facie* plausibility to the Explainer. So we shouldn't think it is obvious that this response to the Humean fails.

This leaves two further questions, which we'll take up over the following chapters. One is whether ampliative reasoning really can work 'off-line' as Hawthorne suggests. Another is whether the argument for premise 2 of the Humean argument works, so we are compelled to accept the contingent *a priori*. I prefer negative answers to both questions, for reasons that will hopefully become clear.

Chapter 12

Justification, Knowledge and Probability

12.1 Probability and Gettier Conditions

Here's one way in which we might think that evidence could produce knowledge while lowering probabilities. (I learned about it from Martin Smith.) Consider a case where all the following obtain.

- S starts with testimony from a very trusted source T that all Fs are Gs.
- S trusts this source, comes to believe that all Fs are Gs, and hence that a, which is known to be F, is G.
- As a matter of fact, a is G, though some other Fs are not.
- At a later time, S gets experimental evidence that entails the commonly held theory of Fs is false, and entails that some hitherto unobserved Fs are in fact not G, but which doesn't give any particular reason to think that previously observed Fs are not G. Indeed, if the theory suggested by Fs evidence is true, then a is G.

Plausibly, the probability of Fa falls a little when S gets this evidence. After all, her grounds for believing Fa are now a new experimental theory, and those are very fallible. But assuming the way that S finds out about a is reasonably secure, it could be that she still knows (defeasibly) that Fa. Presumably she didn't antecedently know that Fa. After all, Fa was an inference from a false premise. So this could be a case where S comes to know Fa through a process whereby the probability of Fa falls.

The structure of the case above is as follows.

- The initial evidence provides strong justification for p, but that justification is blocked from being the basis for knowledge that p because of some Gettier condition.

- The new evidence both undermines the initial evidence for p, but generates a new justification for p which, despite being less strong in some respects, does suffice for knowledge.

Here is a very different kind of case that fits that profile, while supporting perhaps a stronger conclusion.

- S initially starts with evidence that p is true iff a particular lottery ticket loses.
- This is misleading, p is true, but its truth is not related to the lottery.
- Then S gets evidence of a more familiar kind for p, evidence that at once undermines any connection between p and the lottery, but suffices for knowledge that p.

If the lottery is big enough, and the normal evidence for p sufficiently weak, this new evidence will lower the probability of p. But it still might be enough to get knowledge that p despite all that. If we assume (a) that lottery propositions are never knowable, and (b) that evidence can suffice for knowledge without raising the probability of p to 1, then evidence which shows p is not a lottery proposition, but is true, could lower the probability of p while making it a candidate for knowledge.

12.2 Normalcy and Justification

There's an interesting difference between the two examples I just presented. The first case was one where the evidence that lowered probability was unnecessary for justification. By hypothesis, S was justified in believing p before and after getting the evidence. The second case was one where the evidence was necessary for justification. At least if you think lottery propositions cannot be justified, let alone known, then you'll think that initially S was not justified in believing p. She was justified in believing p to be probable, but not in believing it to be true. So perhaps there are cases where probability lowering evidence introduces not just knowledge, but also justification.

That's the lesson Martin Smith (2010) takes away from these lottery cases. And his response is to come up with a theory of justification that does not tie justification to probability. Here's my way of formally representing his theory. For simplicity, we'll identify propositions with sets of possible worlds.[1]

---

[1]This leads to complications with applying the theory to the epistemology of mathematics and modality. I'll leave that problem aside here, noting that Smith's aim is to come up with a more plausible model for justification than classical probability, and that he doesn't need to have anything plausible to say about mathematics and modality to do *that*.

Start with a function from worlds to $[0,1]$, intuitively the 'abnormality' of any given world.[2] The 'worlds' in question are all the possible worlds, and the (one) impossible world, which has abnormality 1. We'll call this function N. The degree of justification that some evidence E gives to a hypothesis H is the minimum value of $N(w)$ for $w \in E \wedge \neg H$.[3] We'll write this as $J(H,E) = x$, where J is the justification function, and x is the minimum value of $N(w)$ for $w \in E \wedge \neg H$.

Here's one nice consequence of this definition of justification.

$$J(A \wedge B, E) = min(J(A,E), J(B,E))$$

Proof: Let $J(A \wedge B, E) = x$. Let w be a world such that $w \in E \wedge \neg(A \wedge B)$ and $N(w) = x$. Then either $w \in \neg A$ or $w \in \neg B$. If the first, then $J(A \wedge B, E) \geq J(A,E)$. If the second, then $J(A \wedge B, E) \geq J(B,E)$. Now w be any world such that $w \in E$ and $N(w) < x$. If $w \in \neg A$ then $w \in \neg(A \wedge B)$, and so $J(A \wedge B, E) \leq N(w) < x$, contradicting our assumption that $J(A \wedge B, E) = x$. So $w \notin \neg A$. So $J(A,E) > N(w)$. By similar reasoning, $J(B,E) > N(w)$. Putting all this together, we get that $J(A \wedge B, E) = min(J(A,E), J(B,E))$, since both $J(A,E)$ and $J(B,E)$ are greater than anything less than $J(A \wedge B, E)$, and one of $J(A,E)$ and $J(B,E)$ is less than or equal to $J(A \wedge B, E)$.

And the upshot of that is that reasoning by $\wedge$-introduction does not reduce justification. This is a striking difference from what happens with reasoning probabilistically. In general, it is possible for $Pr(A \wedge B)$ to be less than both $Pr(A)$ and $Pr(B)$. Indeed, that will happen unless $Pr(A \wedge \neg B)$ or $Pr(\neg A \wedge B)$ equals zero.

But intuitively, reasoning by $\wedge$-introduction does not reduce justification. If it is settled, for the purposes of the current inquiry, that Jack was at the party, and it was settled (for those purposes) that Jill was at the party, no further inquiry is needed into whether Jack *and* Jill were at the party. That immediately suggests that standards for inquiry are not probabilistic thresholds. (Or, at least, that they aren't thresholds less than 1.) Let's say that what it takes to settle something for the purposes of a given inquiry is to show that its probability is greater than 0.95. Then if $Pr(\text{Jack was at the party}) = Pr(\text{Jill was at the party}) = 0.96$, and $Pr(\text{Jack and Jill were both at the party}) = 0.92$, then we'll settle each conjunct

---

[2]The way I'm presenting Smith's story here makes it seem a whole lot like the 'ranking functions' theory that Wolfgang Spohn (2009) and Franz Huber (2007) have written about in recent years. I'll leave the comparisons between those theories as an exercise though.

[3]Strictly speaking, we should fret a bit here about what Lewis (1973) calls the 'Limit Assumption'. Consider an E, H such that for any $x > 0.9$, there is a $w \in E \wedge \neg H$ such that $N(w) = x$, but for no $w \in E \wedge \neg H$ does $N(w) = 0.9$. I'm going to assume the Limit Assumption holds; I'll leave questions about what relaxing the Limit Assumption does to the logic of Smith's theory as an exercise.

without settling the conjunction. That's absurd, so there's something wrong with the probabilistic picture.

12.3   Updating on the Normalcy Model

One of the nice things about the probabilistic model of justification is that it doesn't just tell us a story about justification at a time (i.e., a *synchronic* story about justification), it also gives us a dynamic, diachronic story about how justification evolves over time. Now we've seen reasons to be suspicious of that story. In a sense, the story looked a lot like the story which says you work out all the conditionals you want to believe at the start of time, then when any of their antecedents come true, you believe the consequents. That's not a very plausible story in general, though it works well for a lot of everyday cases. How does the 'normalcy' story that Martin Smith suggests compare to the conditionalisation story that the probabilists like?

In the first instance, it looks fairly similar. It seems that we can prove the following theorem.

$$J_E(A,K) = J(E \supset A, K)$$

Think of the K here as a 'background' evidence. If we are reasoning *a priori*, then we can let K be the conjunction of all *a priori* knowable truths.

The proof of this claim is usefully divided up into two sub-theorems.

- $J_E(A,K) = J(A, K \wedge E)$
- $J(A, K \wedge E) = J(E \supset A, K)$

The second claim is pretty easy. By definition, $J(A, K \wedge E)$ is the normalcy of the most normal world in $K \wedge E \wedge \neg A$. And $J(E \supset A, K)$ is the normalcy of the most normal world in $K \wedge \neg(E \supset A)$. But $\neg(E \supset A)$ is true iff $E \wedge \neg A$ is true. So $K \wedge E \wedge \neg A$ is the exact same set as $K \wedge \neg(E \supset A)$, hence $J(A, K \wedge E) = J(E \supset A, K)$, as required.

What about the first claim? It would follow given two assumptions. The first is that adding evidence has the effect of adding a conjunction to the evidence proposition. This is also in effect what most Bayesians believe. The second is that the function N doesn't change when we update with E. This is more questionable, and perhaps something that isn't very plausible for some understandings of 'normalcy'. (We'll come back to this in the next section.) But first let's look at one other theorem that follows if we assume N is constant over changes in evidence, and that we add evidence by taking conjunctions. Those assumptions entail that,

$$J_E(E \supset A, K) = J(E \supset A, K)$$

Remember, we're assuming in effect that $J_E(E \supset A, K) = J(E \supset A, K \wedge E)$. That is, we are assuming that updating is by conjoining the old evidence to the new, and that the N function doesn't change. Now $J(E \supset A, K \wedge E)$ is the 'normalcy' measure of the most normal world in which $K \wedge E$ is true and $E \supset A$ is false. But if $K \wedge E$ is true in a world, then E is true, so $E \supset A$ is false iff A is false. So the worlds where $K \wedge E$ is true and $E \supset A$ is false are just the worlds where $K \wedge E$ is true and A is false. That is, they are the worlds where K is true and $E \supset A$ is false. So the most normal world where K is true and $E \supset A$ is false will be the most normal world where $K \wedge E$ is true and $E \supset A$ is false, from which $J_E(E \supset A, K) = J(E \supset A, K)$ follows.

This has consequences for the argument we discussed last week that we can't get *a posteriori* justification for certain conditionals of the form $E \supset H$. The reason is that the level of justification for those conditionals does not change when we get E. This is slightly different to what happens on the proababilistic picture when the justification for $E \supset H$ goes *down* when we get E. But the general picture, that E doesn't seem to be the *basis* for $E \supset H$ remains.

12.4   Relativity of Normalcy to Evidence

In the arguments of the previous section it was assumed that the N function is, in a sense, absolute. In particular, it was assumed that getting new evidence did not change how 'normal' various worlds are. Now I think the way Smith writes, that's the most natural way to interpret him. The talk is always about how normal worlds are, not about how normal we know they are, or how normal we have reason to believe they are.

But let's set aside exegetical issues, and think instead about how we might get a theory to work. Say we start instead not with a notion of the normalcy of worlds, but with a notion of resemblence between a world and an evidence set. This world resembles our evidence. It doesn't actually resemble it perfectly, because in many ways the unobserved is unlike the observed, but it resembles it well. Very different worlds are much more unlike our evidence, and so are less normal relative to our evidence.

On this kind of picture, evidence has two roles. First, it establishes a baseline of normality. Second, it rules out some worlds, even some worlds that are normal given the baseline. These two roles are distinct. A world in which last night's lottery numbers were a little different to what they actually were is perfectly normal, given our evidence, but is excluded by the evidence.

Note that given the null evidence, any world is equally normal. Presumably any world whatsoever resembles a tautology. So if our evidence is just T, the conjunction of all tautologies, then for any world, $N(w, T) = 0$. We've now made N a two-place function; it takes worlds and evidence as input, and returns an abnormality as output.

A consequence of that is that gathering evidence E might increase the justification of $E \supset H$. Let E be all the evidence we have, and let H be that Times Square will have people in it this New Years' Eve. Let w be a world where $E \supset H$ is false. That is, it is a world where everything we've seen and heard about the world (to date) is true, but for some reason Times Square is unoccupied this New Years' Eve. Indeed, let w be a maximally normal world satisfying this condition.

Now on the model we've put forward here, $N(w, T) = 0$, since relative to T, no world is abnormal. Abnormality, remember, is difference from evidence, and in w things are just like T. But $N(w, E)$ is presumably very high. A world where our evidence is all true, but Times Square isn't occupied by hundreds of thousands of people on New Years Eve is a world where a form of inductive generalisation that our evidence shows has worked a lot, spectacularly breaks down. So $N(w, E)$ is high. And since $J_E(E \supset H, E) = N(w, E)$, it follows that $J_E(E \supset H, E)$ is high. But $J(E \supset H, T)$, the prior justification of $E \supset H$, is equal to $N(w, T)$. That is, it is zero. So this is a model on which evidence can raise the justification of $E \supset H$.

There's something fairly general about this kind of way we made E be evidence for $E \supset H$. If the role of evidence is purely eliminative, then it is very hard for E to justify $E \supset H$. After all, eliminating $\neg E$ worlds doesn't help at all with justifying a proposition that is entailed by $\neg E$. If N is a constant function, then on Smith's view the role of evidence is purely eliminative. On the Bayesian model, evidence more or less only eliminates, though there is a normalising step after that. Conditionalisation on E, remember, is equivalent to setting the probability of all $\neg E$ possibilities to 0, and then multiplying the probabilities of all possibilities by a constant so that the probabilities sum to 1. So it's not just eliminating $\neg E$ possibilities, but setting all their probabilities to zero comes to much the same thing.

On the model described in this section, evidence plays two roles, not just one. We do eliminate the $\neg E$ worlds when we get evidence E. But we also have the normalcy of worlds be determined by evidence. And since the structure of the justification relation is determined by the normalcy of worlds, in a deeper sense we have the evidence determining the structure of the justification relation.

That's similar to what I was trying to do in Weatherson (2007). The structure of the model is pretty similar to what I've described here. Evidence does play an

eliminative role, since the model includes conditionalisation. But evidence also plays a structuring role. It isn't given *a priori* which are the 'right' probability functions. Rather, what the function is from evidence to hypotheses is something that can only be determined with evidence.

Now one glaring weakness of the model in that paper is that I say more or less nothing about how evidence might licence or endorse a probability function. The model I sketched in this section is a little better on that front. We can see, dimly perhaps, how it is that evidence might show that some worlds are more normal than others. So we can see how an evidence-relative notion of normalcy might be usable in a theory of justification. And, interestingly, using such a notion we can get a theory where we can get empirical justification for propositions like $E \supset H$. So perhaps this is a plausible approach for looking at more going forward.

12.5   Consistency of Normalcy

One issue with the 'normalcy' approach, at least as Smith develops it, is that it isn't obviously consistent. Note that we're meant to have the following all be true:

- In a fair lottery, the closest worlds in which each ticket wins are equally normal.
- Worlds in which odd things happen that are consistent with statistical mechanical laws (e.g., hot and cold water separating in a pan) are not as normal as worlds in which more macroscopically common things happen.
- Our best theory of statistical mechanics is probabilistic.

If we treat the outcome of statistical mechanical processes as literally a lottery, then the first two points seem inconsistent. But the fact that our best theory of statistical mechanics is probabilistic suggests that we should treat these outcomes as 'lottery-like'.

Now there isn't a formal inconsistency here, and perhaps there is a natural way around the puzzle. If we can motivate the idea that 'normalcy' is determined by macroscopic outcomes, then perhaps we can avoid any problem. And there is hard work being done by various philosophers of physics on showing how to make the details of such a theory work. But Smith doesn't really do any of that work; he simply assumes the existence of a solution.

It seems crucial to Smith's story that we take intuitions people have about what's normal to be closely connected with what is actually normal. We do think any lottery ticket winning is equally normal, so they are. We don't think each

statistical mechanical outcome is equally normal, so they are not. What could justify such faith in our intuitions about normalcy?

One option is to assume that we're really very good at working out what's normal. But that seems very implausible. Our intuitions about normalcy are not, for instance, sensitive to facts about the underlying workings of the everyday world. But presumably, even by our own lights, how normal certain outcomes are is sensitive to facts about the underlying workings of the everyday world.

A better option might be to posit a constitutive, rather than an evidential, connection between intuitions about normalcy and what is really normal. This would require making normalcy relative to a community, since different communities could presumably find different things normal. Or it might require rigidly fixing normalcy by the standards of what actual humans have. Either approach would have some similarities to Strawson's idea that induction is rational because it is part of what we mean by 'rational'. I'm not sure how plausible it would be, but it would provide some justification for Smith's use of intuitions about normalcy.

12.6   Philosophical Knowledge and the A Priori

Let's end by looking in more detail at the model I suggested above for how normalcy could be relative to evidence. The idea was that the normalcy of world w relative to evidence set E was given by how much w resembles E. Let's also assume the correctness of this theory is *a priori*. We might reject that last assumption for a couple of reasons. First, we might think that even very broad philosophical theories like this are still *a posteriori*. Second, we might think that what resembles what is really an *a posteriori* matter. The difficulties in coming up with a purely *a priori* theory of resemblence make that last claim fairly plausible. But set those aside, and let's investigate what would follow if our theory was entirely *a priori*.

You might think something very odd would follow. As usual, let H be some hypothesis supported but not entailed by E. As shown above, on this model $E \supset H$ is not knowable *a priori*. But it is knowable *a priori* that E supports H. When E comes along, we don't learn anything new about the relationship between E and H. It's true that we now learn $E \supset H$, which we didn't previously know. But it's a mystery just why we have evidence for $E \supset H$ now, when we didn't previously. Or, at least, it is intuitively mysterious to me why one's attitudes to $E \supset H$ should change in this way.

One way to respond to this intuition would be to tighten it up into an argument that $E \supset H$ really must be knowable *a priori*. But as we've seen, the main

ways to really tighten this argument up don't seem to work. Still, that is consistent with the idea that the intuition is right, and we just haven't got the argument right yet.

In Weatherson (2007), I relied on this intuition a little too heavily. As I noted above, a weakness of that paper was that I don't say exactly *how* evidence could determine what the right evidential probability function is, as opposed to moving us about in a given probability function. And what I said in the paper was that this was a *good* thing, since on the theory being put forward, we shouldn't be able to tell *a priori* how evidence determined what evidence justified. I'd like to be able to do better than that now.

Let's look back at the simple argument by conditional proof for $E \supset H$. We suppose E, infer H by whatever lets us infer H from E in real life, and then infer $E \supset H$, discharging the assumption that E. If we don't like that argument, we need to explain why it fails.

One place to start looking is to think a bit harder about just what E represents. There is a difference between the following two propositions: ⟨There are three chairs in Brian's office⟩; ⟨It is part of Brian's evidence that there are three chairs in Brian's office⟩. When I walk into my office and see three chairs there, presumably the first of these propositions becomes part of my evidence. (I assume seeing that p is sufficient for p being part of one's evidence.) But something else happens when I walk into my office; the second proposition becomes true. That second proposition might or might not become part of my evidence, but it is true, and this might be relevant.

Something else becomes true when I look around my office. I *see* that there are three chairs in the office. I can do this without immediately believing that I see that there are three chairs in the office. Depending on your theory of perception, perhaps I have to believe p in order to see p. But I certainly don't have to believe that I see p in order to see p.

Given all that, how could we make a supposition from which we could (conditionally) draw the same inferences that we can draw from seeing that p. We can't merely suppose p. It could be that seeing p lets us infer more things than p itself does. But we can't suppose that we see that p. From that we can clearly infer that we see that p. But it isn't always true that when we see that p, we are even in a position to know that we see that p. So it isn't obvious that any supposition is quite like our observation.

This idea is reinforced by the following somewhat Williamsonian thought experiment. Imagine you're in a world with a countable infinity of testifiers. There are infinitely many truth-tellers, and infinitely many liars. But in a good sense, the truth-tellers are thicker on the ground than the liars. Perhaps it's true that for

all points x, and for some not-too-large distances d, then for any distance $d* > d$, the overwhelming majority of testifiers within distance $d*$ of x are truth-tellers. Plausibly, in that world, testimony from strangers is a source of justification in the absence of reason to think that they are lying. After all, truth-tellers are thicker on the ground than they are in this world, and in this world testimony from strangers is a source of justification in the absence of reason to think that they are lying. But it could also be true in the world that for any false, expressible p, there is a liar who testifies that p.

Now let p be some proposition that you have reason to think is less likely true than not, but which you could be convinced to believe by testimony from an unchecked source. Somewhere out in the world, is a testifier who testifies that p. If they gave that testimony to you, you would infer p. But they don't. Still, you can suppose they do. And supposing they do, do you infer that p, conditional on your being told that p by someone?

This would be absurd. It seems like you'd be led to inconsistent views, since you know someone testifies that p, you believe of any person that if they testify that p, then p, but you don't believe p. (And the same is true replacing p with $\neg p$ throughout.) What has gone wrong?

It seems to me that there's a difference between hearing p, and supposing that p, a difference that isn't entirely captured by making p a mere supposition. When you hear p, it isn't just p goes into your 'evidence box'; it goes in a certain way. And that could be the start of an explanation for why we can do different things with evidence we have than with 'evidence' we merely suppose to be true.

Chapter 13

Exhaustive Arguments

13.1   What is Dogmatism?

We're mostly interested this week in arguments of the following form: Whatever evidence the agent gets, she'll be justified in believing p, so she's justified in believing p *a priori*. But before I get to that, I want to note one quick point about how it relates to dogmatism.

In both White (2006) and Weatherson (2005), it is assumed that dogmatism about perception is in conflict with a theory of justification which says that the reliability of perception is, at least in some worlds, knowable *a priori*. But it isn't obvious that this is right, and indeed it is denied by Pryor. What's crucial to dogmatism is the possibility that an agent can rationally transition from *Appears that p* to p, without a prior justification to believe the conditional *If it appears that p, then p*. What isn't crucial is that this transition takes place 'on-line', when the agent actually has an appearance that p. It is consistent with dogmatism that an agent could make this transition in suppositional reasoning, and come to *a priori* knowledge of the conditional *If it appears that p, then p* by the following simple proof.

1. Assume, for conditional proof, that it appears that p.
2. So p, by the transition licenced by dogmatism.
3. So if it appears that p, then p, discharging the assumption at line 1.

Note that the truth, or justifiedness, of *If it appears that p, then p* is not actually an input to that proof. The input is the dogmatist transition rule, and we get various assumptions out without assuming the conditional.

If we can do all this *a priori*, you might wonder what is significant about dogmatism. Well one thing that remains significant is that we are taking a transition rule, not a belief as basic. And that seems like a very important distinction to me, since the epistemology of transition rules is plausibly very very different from

the epistemology of beliefs. So the view that says the conclusion of the above argument is a piece of inferential (albeit *a priori*) knowledge is a very interesting theory.

One upshot of this is that a lot of the arguments in White (2006) and Weatherson (2005) are not so much arguments against dogmatism, as arguments that dogmatists are committed to the contingent *a priori*. But let's see whether that's really true.

13.2   White's Exhaustive Argument

White's version of a dominance argument, or perhaps better an exhaustive argument, starts with this definition.

> A **super-fake-hand** is a non-hand which not only looks just like a hand, but has magical powers that prevent observers from gaining independent evidence that it is not really a hand.

He then says that if we look at someone, say G. E. Moore, who may or may not have a hand, there are three possibilities.

A: He will not appear to have hands.

B: He will appear to have hands, but we will have some reason to suspect that this appearance is deceptive.

C: He will appear to have hands, and we will have no reason to suspect that this appearance is deceptive.

And we want to consider these propositions. I've relabelled them in hopefully easier to remember ways.

AH He appears to have hands.

H He has hands.

NSF He does not have super-fake hands.

White argues that in each of the three cases, we'll be justified in believing **NSF**. Here's the reasoning in case A.

> Now suppose A obtains. We will clearly be justified in believing that he is not super-fake-handed, as he doesn't even appear to have a hand. (White, 2006, 538)

I'm dubious about this reasoning, and not on the general grounds that when a philosopher says something is 'clearly' the case you should check where your wallet is. It's true that if the agent is justified in believing ¬**AH**, then they are justified in inferring **NSF**, since ¬**AH** entails **NSF**. But that's not what White is supposing. What he is supposing is simply that ¬**AH** is true. We need something extra to get that one could justifiedly believe ¬**AH**. Maybe there is meant to be a tacit assumption that appearances are luminous. Given that dogmatism is an internalist theory, that's not a completely crazy tacit assumption, but it should be expressly noted..

A similar kind of point arises with case B.

> If B obtains then we will likewise be justified in denying that he has a super-fake-hand. For we know that if he did, we would not have obtained any evidence that we were subject to an illusion. (White, 2006, 538)

This argument is subject to the same objection as the previous argument. What's true in this scenario is that the agent has a reason to suspect that appearances are deceptive. It doesn't follow, except on an implausible view about the luminosity of reasons, that the agent knows they have such a reason, or are even justified in believing they have such a reason.

It's also worth thinking about this in a broader context. What matters for scenario B is that one of the conditions which block an inference from **AH** to **H** is triggered. On an externalist version of dogmatism, there's really no reason to suspect that whether one of these conditions obtains will be luminous. Consider, for example, a kind of dogmatist reliabilism which says that an agent can infer **H** from **AH** unless appearances in general are unreliable. (Perhaps, they might say, this unreliability is a reason not to make the inferences, though not necessarily a reason available to the agent.) Call scenario B* the scenario where the way things appear to the agent are in general unreliable. There's no plausibility at all to the conclusion that in scenario B*, the agent would be justified in believing **NSF** (at least by the route White suggests). The fact that their appearances are unreliable does not on its own justify the belief that appearances are unreliable. So the agent can't infer that way that **NSF**.

Things are a little different with scenario C.

> If C obtains, then according to dogmatism, we will be justified in believing that it is a hand on the end of his arm, and hence in turn that it is not a super-fake-hand. (White, 2006, 538)

This seems right. And, perhaps significantly, it does not involve using dogmatist reasoning under the scope of a supposition. The reasoning isn't from a supposition of **AH** to **H**. Rather, the reasoning is from a supposition of **AH** to the claim that in such a scenario, belief in **H** is justified. And that's something the dogmatist is really committed to, even on the assumption that only 'on-line' uses of dogmatist reasoning is justified. So I think this step is fine.

13.3 The Exhaustive Argument in 'Scepticism, Rationalism and Externalism'

The argument in Weatherson (2005) is not meant to show that everyone is committed to the existence of the contingent *a priori*. It's just meant to show that anti-sceptic internalists are so committed. In the paper, I use a somewhat idiosyncratic definition of internalism, that was intended at least to be weak enough to capture what's going on with actual paradigm internalists. The definition starts with a technical term.

> Say that a class of properties (intuitively, a determinable) is *introspective* iff any beliefs an agent has about which property in the class (which determinate) she instantiates are guaranteed to not be too badly mistaken. (Weatherson, 2005, ??)

Annoyingly, despite defining introspectiveness to be a property of *classes* of properties, I immediately go on to talking about particular properties being introspective or not. This seems like a category mistake. Let's say that a particular property F is introspective iff the class $\{F, \neg F\}$ is introspective. Surprisingly, the obvious vagueness in 'too badly' doesn't make as much of a difference as you'd expect. The vagueness is going to 'wash out' in a certain way.

Given those definitions, my internalist is someone who accepts the following two claims.

- Which propositions an agent can justifiably believe supervenes on which introspective properties she instantiates, and this is knowable a priori.
- There exist some introspective properties and some deeply contingent propositions about the future such that it's a priori that whoever instantiates those properties can justifiably believe those propositions.

It's really the first clause that does all the work; it's hard to imagine a non-sceptic who would accept the first and reject the second. So the core of internalism, as I'm construing it, is the idea that it is *a priori* that justification supervenes on

properties about which one can't be too badly mistaken. Since one can be massively mistaken about things like reliability, this rules out reliability (and anything similar) being part of the subvenient base for justification. And that in turn rules out most forms of externalism. The theories that remain are, intuitively, internalist theories.

If F is introspective, then I argued that the following principle is plausible, where $\Box$ and $\Diamond$ are epistemic operators with their usual meaning, i.e., $\Box p$ means that the agent, here a, is in a position to know that p, and $\Diamond$ is the dual of $\Box$.

$$Fa \rightarrow \Box \Diamond \Diamond Fa$$

That is, if a is F, then a is in a position to know that $\Diamond \Diamond Fa$. And $\Diamond \Diamond Fa$ means it is consistent with all a is in a position to know that it's consistent with all a is in a position to know that Fa. This principle is true on a reasonable 'margin-of-error' model for knowledge, though notably it isn't true on Williamson's preferred version. (As I note in the paper, Delia Graff Fara (2002) originally pointed out that to get $Fa \rightarrow \Box \Diamond \Diamond Fa$ out of the margin-of-error model, you simply needed to add a very weak, and very plausible, density assumption.)

This isn't a *luminosity* assumption, since there's nothing that is said to be knowable when true. But it is a principle that says that whenever something is true, something related is known (or at least knowable). And that's all we need to get the argument running. If a property isn't introspective, and one could in principle be arbitrarily badly wrong about how and when it applies, the margin-of-error model doesn't seem appropriate. But for introspective properties, even given the vagueness of their definition, the margin-of-error model does seem appropriate. That's why I thought $Fa \rightarrow \Box \Diamond \Diamond Fa$ was plausible when F is introspective.

Now we simply apply the second clause of the definition of internalism. Let G be an introspective property and p be a proposition such that it is knowable *a priori* that anyone whose evidence is G is justified in believing p. And define q as follows.

$$q =_{df} p \lor \neg \Box \Box Ga$$
$$= \Box \Box Ga \subset p$$

The conditional form is there to show the resemblence to earlier attempts to find a contingent *a priori* proposition, but the disjunction is easier to think about.

The claim I made in Weatherson (2005) was that q is knowable *a priori* for a, and then only if the mode of presentation of a in q is a distinctively first-personal mode. Or, at least, that's the claim I should have made. Instead I made things more complicated than they should have been by presenting everything in a first-personal mode. This is bad practice, and I wish I hadn't engaged in it here. Hopefully splitting up the agent of the example from the person writing about the example will make it clearer what is being written!

So why is q knowable *a priori*, at least by a? The argument in Weatherson (2005) is fairly complicated, and the presentation of it doesn't help matters. Here's the core argument. It seems a can reason to herself, entirely *a priori*, as follows.

1. Either my evidence is G or it isn't. Take these in turn.
2. If my evidence is G, then I'm justified in believing p. For all I know, I'm justified *a posteriori* by my evidence being G, but I am justified.
3. Hence I'm justified in believing q, by a simple closure principle.
4. If my evidence is $\neg G$, then by the luminosity principle I know $\diamond \diamond \neg Ga$. And that's equivalent to $\neg \Box \Box Ga$.
5. So I'm again justified in believing q.
6. Since either disjunct implies I'm justified in believing q, I am justified in believing q.

This much I think clearly follows from the assumptions that are made in Weatherson (2005). The next step is tricky though.

Note that all we've done so far is show that a knows *a priori* that she's justified in believing q. We haven't shown that she knows, or even is justified in believing q *a priori*. By the factivity of knowledge, it follows that she's justified in believing q, but this justification may, for all we've said so far, be *a posteriori*. How might we close this gap?

Here's one way, not the way I tried in Weatherson (2005). Consider the inference *Belief in q is justified (for me), therefore q*. That doesn't correspond to a valid argument, unless you believe that justification is factive. But it seems like a plausible form of inference to make, at least 'on-line'. (If you can make it in suppositional reasoning, then there is a simple argument for the existence of the contingent *a priori*. There is an easy proof of *If belief in q is justified, q*.) Now our agent a knows, and indeed knows *a priori* that belief in q is justified for her. Assume that the inference from *q is justified* to q can transmit knowledge. Then a can, simply by using this inference, come to know q. And that knowledge will be *a priori*.

For better or worse, that wasn't the strategy I used in Weatherson (2005). Instead I did something a little more roundabout. What I argued was that since *a* knows she'd be justified in believing *q* come what may, the actual empirical evidence she has for that belief can't really be the *basis* for the justification. After all, even if that evidence was somehow defeated, she could recreate a justification for *q*, and she knows this. She knows that *q* is justified, so it is justified. And the justification doesn't have any empirical evidence as its basis. Any belief that's justified, but not on the basis of empirical evidence, seems to me to be properly called *a priori*. So I think there's an *a priori* justification for *q*.

Unlike the previous approach, this approach leaves us with some work to do in getting from an *a priori* justification for *q* to *a priori* knowledge of *q*. In Weatherson (2005) I argued, not particularly convincingly, that since there was no Gettier-like 'funny business' going on, it's reasonable to assume that this particular justified true belief is indeed knowledge. But I no longer believe that kind of move is legitimate. There are, after all, a lot of distinct ways in which a justified true belief can fail to be known.

13.4   Exhaustive Arguments in General

I hopefully was a little clear about this in the earlier paper, but I wanted to say a little bit about why the strategy of proving that there is contingent *a priori* knowledge won't work given externalism about justification. The strategy is fairly easy in outline. Partition the possible scenarios. For each cell of the partition, find a proposition that is justified given that the agent is in that cell. Make sure one of the propositions is not entailed by being in that cell. (That is, use the fact that theories of justification allow for ampliative justification in at least one cell.) Take the disjunction of all these propositions. Hand wave a little about how the justification in each cell entails *a priori* justification. *Voila!* A proof, a constructive proof even, of the existence of the contingent *a priori*.

It's easy to see how to get started on this, given any non-sceptical theory of justification. Let E and H be such that the theory of justification being targeted says that E justifies H without entailing it. Then let one of the cells be that the agent has evidence E, and let the proposition that's paired with that cell be H. Great; one step down.

The problem is dealing with the cells where the agent does not have evidence E. Given internalism, this is an easy problem. Whether the agent has evidence E or not is an internal matter. Internal matters are knowable, or at least are roughly knowable. So if the agent doesn't have evidence E, some proposition like that will be justified. Maybe, if the epistemologist doing the proof is sloppy, or buys a crazy strong luminosity assumption, the proposition will simply be that the

agent doesn't have E. Or perhaps, as in my proof, it will be some complicated proposition, such as that it is possibly possible that they lack evidence E. There are technical challenges here, but they seem managable.

The harder case is when we are looking at an externalist theory. Consider the following three possibilities.

1. S has an appearance of a hand, and her appearances are a reliable indicator of reality.
2. S has no appearance of a hand.
3. S has an appearance of a hand, and her appearances are not a reliable indicator of reality.

The proposition that goes with case 1 is *I have a hand*. The proposition that goes with case 2 is *I don't have a hand appearance*, or perhaps something more complicated. But what could possibly go with case 3? What proposition could the agent be justified in believing in virtue of having a misleading appearance of a hand? Not that she has a hand; her appearances are unreliable. Not that she has no hand appearance; it seems (correctly) to her that she does. What else? Nothing springs to mind.

That's not a full argument, to be sure. But I really don't see how one could finish this step of the proof. So I don't see how *this* kind of argument can be used to prove that there are contingent *a priori* knowable propositions, if the underlying theory of justification is externalist.

Things are different if we are allowed to use ampliative inferences inside the scope of suppositional reasoning. You might well think, given reliabilism, that the agent can infer from being in case 1 to her having a hand. Reliabilists typically say that the agent doesn't *need* to know her appearances are reliable to use them in gaining knowledge, but there doesn't seem much harm in her doing so. Then she could just reason as follows.

1. Assume, for conditional proof, that I have an appearance of a hand, and my appearances are a reliable indicator of reality.
2. Then I have a hand, by reliabilism.
3. So if I have an appearance of a hand, and my appearances are a reliable indicator of reality, then I have a hand, discharging the initial premise.

Since reliability is meant to be weaker than infallibility, that conclusion is contingent. But it seems to be knowable *a priori*. What, if anything has gone wrong? We'll end this week on that question.

13.5 Ampliative Suppositional Reasoning

I don't think it is legitimate to use ampliative reasoning in natural deduction systems. It seems to me to involve confusion between inference and implication. (To be honest, I think natural deduction systems encourage just that kind of confusion.) So I don't think 'proofs' like the one immediately above work, because I don't think the step from 1 to 2 actually works. I think inside suppositional reasoning, one can only use deductively valid rules.

Other philosophers, such as Stewart Cohen (2010) and Sinan Dogramaci (2010), think we can use ampliative rules inside suppositional reasoning. I think a construction similar to one Dogramaci uses can refute this. Here's how. If any ampliative inference is justified, I think the following rule, called 'IR', is justified, since this is a very weak form of an inductive inference.

IR From *There are infinitely many Fs, and at most one is not G* and *x is F* infer *x is G* unless there is some *H* such that it is provable from the undischarged assumptions that *x is F and H* and *There are finitely many things that are both F and H, and one of them is not G.*

Note that the rule doesn't say that merely one $F \wedge \neg G$ has been observed; it requires that just one such thing exists. So this seems like a very plausible inference; it really is just making an inference within a known distribution, not outside it. And it is explicitly qualified to deal with defeaters. And yet even this rule, when applied inside the scope of suppositions, can lead to absurdity.

In the following proof, we'll let *N* be the predicate 'is a natural number', and *P* be the predicate 'is the predecessor of', and I'll appeal to the fact that there are infinitely many natural numbers, and each number has at most one predecessor. I'll use a version of the proof system in E. J. Lemmon's *Beginning Logic*, but it should be easy to transform the proof into any other proof system.

1	(1)	Na	assumption
2	(2)	Nb	assumption
1,2	(3)	$\neg Pab$	(1), (2), IR
1	(4)	$Nb \rightarrow \neg Pab$	(2), (3), CP
1	(5)	$\forall y(Ny \rightarrow \neg Pay)$	(4), UI
	(6)	$Na \rightarrow \forall y(Ny \rightarrow \neg Pay)$	(1), (5), CP
	(7)	$\forall x(Nx \rightarrow \forall y(Ny \rightarrow \neg Pxy))$	(6), UI
	(8)	$N2 \rightarrow \forall y(Ny \rightarrow \neg P2y)$	(7), UE

So we get the absurd result that if 2 is a number (which it is!), then it is the predecessor of no number. But that's absurd, since obviously 3 is a number and 2 is the predecessor of it. Note that at step 3 we use rule IR with F being the predicate *is a natural number*, G being the predicate *does not have a as a predecessor*, and b being x.

What could have gone wrong? I think the problem is using IR in the context of a suppositional proof, as we've done here. But let's check if there is another guilty suspect.

If the problem is Conditional Proof (CP in Lemmon's system), then that's about as bad for the proof in the first paragraph that there are contingent a priori truths as if the problem is IR. Since we're interested in whether that proof works, we won't investigate this option further. In any case, if $\rightarrow$ is material implication, that rule seems unobjectionable. A referee suggested that if we've used an ampliative rule earlier, then $\rightarrow$ should be weaker than material implication, and under that interpretation (5) through (8) may be plausible. I think that claim is basically right, but note that if we do this the argument for contingent a priori knowledge with which I started will fail, since the contingency of $A \supset B$ will not imply the contingency of $A \rightarrow B$ if $\rightarrow$ is weaker than $\supset$.

It is hard to imagine that Universal Elimination (UE) is the problem. In any case, line (7) is obviously bad anyway, so something must have gone wrong in the proof before that.

Perhaps the problem is with Universal Introduction (UI); this is what Dogramaci suggests. One objection he offers is that although we can prove every instance of the universal quantifier, inferring the universal version creates an undue aggregation of risks. Even if line (4) is very probable, and it would still be probable if a were replaced with c, d or any other name, it doesn't follow that the universal at line (5) is very probable. But I think this is to confuse defeasible reasoning with probabilistic reasoning. The only way to implement this restriction on making inferences that aggregate risk would be to prevent us making any inference where the conclusion was less probable than the premises. That will rule out uses of $\forall$-introduction as at (5). But it will also rule $\land$-introduction, and indeed any other inference with more than one input step. To impose such a restriction would be to cripple natural deduction.

Another objection he offers (UI) is simply that it is the least plausible, or least intuitive, of the rules used here. But in fact (UI) is extremely intuitive. If we can prove every instance of a schema, we should be able to prove its universal closure. On the other hand, allowing ampliative rules to be used inside the scope of a supposition allows a quick proof of contingent a priori knowledge, as shown

in the first paragraph. Now maybe there is such knowledge, but its existence is hardly intuitive.

So I conclude the weakest link in the argument is step (3). Although IR is a good rule, it can't be used inside the scope of a supposition. And since IR is about as weak an inductive rule as we can imagine, I conclude that ampliative inference rules can't in general be used inside the scope of suppositions.

The general lesson here is that, as was made clear many years ago by Gilbert Harman (1986) is that there is a difference between rules of inference and rules of implication. The quick proof that there's contingent a priori knowledge uses a rule of inference as if it is a rule of implication. Not respecting this distinction between inference and implication leads to disaster, as we've shown here, and should be shunned.

Bibliography

Block, Ned. 1978. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9:261–325.

BonJour, Laurence. 1997. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

BonJour, Laurence and Sosa, Ernest. 2003. *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Great Debates in Philosophy. Malden, MA: Blackwell.

Braddon-Mitchell, David and Jackson, Frank. 2007. *The Philosophy of Mind and Cognition, second edition*. Malden, MA: Blackwell.

Brewer, Bill. 1999. *Perception and Reason*. Oxford: Oxford University Press.

Brueckner, Anthony. 1994. "The Structure of the Skeptical Argument." *Philosophy and Phenomenological Research* 54:827–835.

Byrne, Alex. 2004. "How Hard are the Sceptical Paradoxes?" *Noûs* 38:299–325, doi:10.1111/j.1468-0068.2004.00471.x.

Campbell, John. 2002. *Reference and Consciousness*. Oxford: Oxford University Press.

Carroll, Lewis. 1895. "What the Tortoise Said to Achilles." *Mind* 4:278–280.

Chalmers, David. 2002. "Does Conceivability Entail Possibility?" 145–200.

Chalmers, David J. and Jackson, Frank. 2001. "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110:315–360, doi:10.1215/00318108-110-3-315.

Chisholm, Roderick. 1942. "The Problem of the Speckled Hen." *Mind* 51:368–373, doi:10.1093/mind/LI.204.368.

Christensen, David. 2005. *Putting Logic in Its Place.* Oxford: Oxford University Press.

—. 2010. "Disagreement, Question-Begging and Epistemic Self-Criticism." *Philosophers' Imprint* 11:1–22.

Coady, C. A. J. 1995. *Testimony: A Philosophical Study.* Oxford: Clarendon Press.

Cohen, Stewart. 1998. "Two Kinds of Skeptical Argument." *Philosophy and Phenomenological Research* 58:143–159.

—. 2002. "Basic Knowledge and the Problem of Easy Knowledge." *Philosophy and Phenomenological Research* 65:309–329.

—. 2010. "Bootstrapping, Defeasible Reasoning and *A Priori* Justification." *Philosophical Perspectives* 24:141–159, doi:10.1111/j.1520-8583.2010.00188.x.

Descartes, René. 1641/1996. *Meditations on First Philosophy, tr. John Cottingham.* Cambridge: Cambridge University Press.

Dogramaci, Sinan. 2010. "Knowledge of Validity." *Noûs* Forthcoming.

Dretske, Fred. 2005. "Is Knowledge Closed Under Known Entailment? The Case Against Closure." 13–26.

Dudman, V. H. 1991. "Interpretations of 'If'-Sentences." In Frank Jackson (ed.), *Conditionals*, Oxford Readings in Philosophy, 202–232. Oxford: Oxford University Press.

Elga, Adam. 2007. "Reflection and disagreement." *Noûs* 41:478–502, doi:10.1111/j.1468-0068.2007.00656.x.

Fara, Delia Graff. 2002. "An Anti-Epistemicist Consequence of Margin for Error Semantics for Knowledge." *Philosophy and Phenomenological Research* 64:127–142. This paper was first published under the name 'Delia Graff'.

Fodor, Jerry A. 1998. *Concepts: Where Cognitive Science Went Wrong.* Oxford: Oxford University Press.

Goldman, Alvin. 2009. "Williamson on Knowledge and Evidence." In *Williamson on Knowledge*, 73–91. Oxford: Oxford University Press.

Harman, Gilbert. 1986. *Change in View.* Cambridge, MA: Bradford.

Hawthorne, John. 2002. "Deeply Contingent A Priori Knowledge." *Philosophy and Phenomenological Research* 65:247–269.

—. 2005. "Knowledge and Evidence." *Philosophy and Phenomenological Research* 70:452–458, doi:10.1111/j.1933-1592.2005.tb00540.x.

Huber, Franz. 2007. "The Consistency Argument for Ranking Functions." *Studia Logica* 86:299–329, doi:10.1007/s11225-007-9062-9.

Hume, David. 1739/1978. *A Treatise on Human Nature*. Oxford: Clarendon Press, second edition.

—. 1748/2004. *An Enquiry Concerning Human Understanding*. New York: Barnes and Noble.

Ichikawa, Jonathan. 2008. "Scepticism and the Imagination Model of Dreaming." *The Philosophical Quarterly* 58:519–527, doi:10.1111/j.1467-9213.2007.546.x.

—. 2009. "Dreaming and Imagination." *Mind and Language* 24:103–121, doi:10.1111/j.1468-0017.2008.01355.x.

Jackson, Frank. 1977. *Perception: A Reprsentative Theory*. Cambridge: Cambridge University Press.

Jenkins, C. S. 2007. "Entitlement and rationality." *Synthese* 157:25–45, doi:10.1007/s11229-006-0012-2.

Joyce, Jim. 2004. "Williamson on Evidence and Knowledge." *Philosophical Books* 45:296–305, doi:10.1111/j.1468-0149.2004.0356c.x.

Kelly, Thomas. 2010. "Peer disagreement and higher order evidence." In *Disagreement*, 111–174. Oxford University Press.

Kennett, Jeanette and Smith, Michael. 1996. "Philosophy and Commonsense: The Case of Weakness of Will." In Michaelis Michael and John O'Leary-Hawthorne (eds.), *The Place of Philosophy in the Study of Mind*, 141–157. Norwell, MA: Kluwer, doi:10.1017/CBO9780511606977.005.

Keynes, John Maynard. 1921. *Treatise on Probability*. London: Macmillan.

Klein, Peter. 2011. "Skepticism." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2011 edition.

Larmore, Charles. 1984. "Descartes' Psychologistic Theory of Assent." *History of Philosophy Quarterly* 1:61–74.

Lemmon, E. J. 1978. *Beginning Logic*. Indianapolis: Hackett.

Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell Publishers.

—. 1982. "Logic for Equivocators." *Noûs* 16:431–441. Reprinted in *Papers in Philosophical Logic*, pp. 97-110.

—. 1997. "Finkish Dispositions." *Philosophical Quarterly* 47:143–158. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 133-151.

—. 1998. *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.

Lipsey, R. G. and Lancaster, Kelvin. 1956-1957. "The General Theory of Second Best." *Review of Economic Studies* 24:11–32, doi:10.2307/2296233.

Loeb, Louis E. 1990. "The Priority of Reason in Descartes." *Philosophical Review* 99:3–43.

Mackie, John. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

Moore, G. E. 1959. "A Defence of Common Sense." In *Philosophical Papers*, 32–59. London: Allen and Unwin.

Newman, Lex. 2010. "Descartes' Epistemology." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Palo Alto: CSLI, fall 2010 edition. http://plato.stanford.edu/archives/fall2010/entries/descartes-epistemology/.

Newman, Lex and Nelson, Alan. 1999. "Circumventing Cartesian Circles." *Noûs* 33:370–404, doi:10.1111/0029-4624.00160.

Nozick, Robert. 1981. *Philosophical Explorations*. Cambridge, MA: Harvard University Press.

Okasha, Samir. 2001. "What did Hume Really Show about Induction?" *The Philosophical Quarterly* 51:307–327, doi:10.1111/1467-9213.00231.

Oppy, Graham. 2006. *Arguing About Gods*. Cambridge: Cambridge University Press.

Pargetter, Robert and Bigelow, John. 1997. "The Validation of Induction." *Australasian Journal of Philosophy* 75:62–76, doi:10.1080/00048409712347671.

Priest, Graham. 2000. "Could Everything be True?" *Australasian Journal of Philosophy* 78:189–195, doi:10.1080/00048400012349471.

Pryor, James. ???? "Problems for Credulism." Retrieved from http://www.jimpryor.net/research/papers/Credulism.pdf.

—. 2000. "The Sceptic and the Dogmatist." *Noûs* 34:517–549, doi:10.1111/0029-4624.00277.

—. 2004. "What's Wrong with Moore's Argument?" *Philosophical Issues* 14:349–378.

Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA.: MIT Press.

Rickless, Samuel. 2005. "The Cartesian Fallacy Fallacy." *Noûs* 39:309–336, doi:10.1111/j.0029-4624.2005.00503.x.

Rubin, Ronald. 1976. "Descartes's Validation of Clear and Distinct Apprehension." *Philosophical Review* 86:197–208.

Russell, Bertrand. 1912/1997. *The Problems of Philosophy*. Oxford: Oxford University Press.

—. 1921/2008. *The Analysis of Mind*. Allen and Unwin. Retrieved from Project Gutenberg - http://www.gutenberg.org/ebooks/2529.

Smith, Martin. 2010. "What Else Justification Could Be." *Noûs* 44:10–31, doi:10.1111/j.1468-0068.2009.00729.x.

Sober, Elliot. 1988. *Reconstructing the Past – Parsimony, Evolution, and Inference*. Cambridge: MIT Press.

Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.

Spohn, Wolfgang. 2009. "A Survey of Ranking Functions." In Franz Huber and Cristoph Schmidt-Petri (eds.), *Degrees of Belief*, 185–228. New York: Springer.

Strawson, P. F. 1952. *Introduction to Logical Theory*. London: Methuen.

Vogel, Jonathan. 1990. "Cartesian Skepticism and Inference to the Best Explanatio." *Journal of Philosophy* 87:658–666.

—. 2004. "Skeptical Arguments." *Philosophical Issues* 14:426–455.

Weatherson, Brian. 2005. "Scepticism, Rationalism and Externalism." *Oxford Studies in Epistemology* 1:311–331.

—. 2007. "The Bayesian and the Dogmatist." *Proceedings of the Aristotelian Society* 107:169–185, doi:10.1111/j.1467-9264.2007.00217.x.

White, Roger. 2005. "Explanation as a Guide to Induction." *Philosophers' Imprint* 5:1–29. http://www.philosophersimprint.org/005002/.

—. 2006. "Problems for Dogmatism." *Philosophical Studies* 131:525–557.

Williamson, Timothy. 1986. "The Contingent *A Priori*: Has it Anything to do with Indexicals?" *Analysis* 46:113–117.

—. 1996. "Cognitive Homelessness." *Journal of Philosophy* 93:554–573.

—. 1998. "Conditionalizing on Knowledge." *British Journal for the Philosophy of Science* 49:89–121, doi:10.1093/bjps/49.1.89.

—. 2000a. *Knowledge and its Limits.* Oxford University Press.

—. 2000b. "Scepticism and Evidence." *Philosophy and Phenomenological Research* 60:613–628.

—. 2007. *The Philosophy of Philosophy*. Blackwell Pub. Ltd.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. London: Macmillan.

Wright, Crispin. 2000. "Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof." *Philosophical Issues* 10:140–163.

—. 2002. "(Anti-)Sceptics Simple and Subtle: G.E. Moore and John McDowell." *Philosophy and Phenomenological Research* 65:330–348, doi:10.1111/j.1933-1592.2002.tb00205.x.

—. 2004. "Warrant for Nothing (And Foundations for Free)?" *Proceedings of the Aristotelian Society, Supplementary Volume* 78:167–212.