Epistemic Norms, A Priority, and Self-Knowledge

Richard Flockemann

PhD Thesis

University of York

Department of Philosophy

March 2011

Abstract

This thesis is primarily focussed on developing a novel characterisation of the distinction between a priori and a posteriori justification. My working hypothesis is that we can make a surprising amount of progress in this field by paying attention to the structure of *epistemic norms.*

I argue that direct a priori beliefs are governed by a structurally different kind of epistemic norm to the one that governs perceptual beliefs. That, I argue, is where the fundamental epistemological difference between the two categories lies. I call this view 'Seeming-Independence'.

Seeming-Independence holds that while a posteriori beliefs depend epistemically on how it *perceptually* seems to us, there is no corresponding dependence relation between a priori beliefs and how it *intellectually* seems to us. Intellectual seemings, or intuitions, simply do not play the kind of epistemological role that perceptual experiences play.

The central contention of this thesis is that Seeming-Independence is a *theoretically fruitful* view of the a priori. The arguments I marshal in favour of Seeming-Independence are in this way primarily focussed on the explanatory power and flexibility of the view.

In effect, what I suggest is that Seeming-Independence, unlike some of its rivals, is a particularly clear way of dividing the a priori from the a posteriori, and it allows us to neatly bypass some of influential criticisms of a priority.

Contents

Acknowledgements

I would like to thank first and foremost my supervisor, Professor Tom Stoneham, for his extremely helpful advice, his insightful observations, and his encouragement.

I am also grateful for the overwhelming generosity and emotional support my family provided throughout this four year period: particular mention must go to my parents, Janet and Richard; my siblings, Tanya and Peter; and my grandmother, Joyce.

Declaration

I declare that I am the sole author of this work. None of this material has been published elsewhere.

Introduction

What is the difference between a priori justification or warrant and a posteriori warrant? Can we plausibly construe this distinction such that it is both *genuinely epistemological* and philosophically important? What does it mean to say that a priori warrant is independent of experience? What sort of mental state counts as experience in this sense? What kinds of propositions can we know a priori?

These are the sort of questions that this thesis is devoted to exploring. My working hypothesis is that we can make a surprising amount of progress on questions like these by paying attention to the structure of *epistemic norms.*

In essence, the view I defend is that direct a priori beliefs are governed by a structurally different kind of epistemic norm to the one that governs perceptual beliefs. That, I argue, is where the fundamental epistemological difference between the two categories lies.

This project as a whole is very firmly grounded in a particular understanding of epistemic norms. I begin by drawing attention to the fact that, amongst those that take a normative approach to epistemology, questions of whether a believer is *rational* in holding her beliefs are often settled by considering whether it would be appropriate to epistemically reproach, or criticise, or in any sense *blame* her for how she went about forming and maintaining her beliefs. If there is nothing to criticise or find fault with in her epistemic behaviour then this establishes that her belief is *rational*.

A key contention of this thesis is that this is a mistake. What I take to be the central error here is that discussions of epistemic normativity tend to uncritically model epistemic norms on moral norms. There is a strong tradition in moral philosophy to take *normative moral judgements* to be closely connected to 'reactive attitudes' like praise and blame. The assumption being made in normative epistemology is that there is a similarly close connection between analogous epistemic judgements and analogous epistemic sentiments.

The presumption that epistemic norms have much in common with moral norms is a long-standing one, dating all the back to Clifford's classic *The Ethics of Belief* (1877).

However, in my view, it is an assumption that should be challenged. Of late there has been considerable and increasing discussion of non-moral norms. Philosophers frequently debate about what the *norms* of assertion are, or whether mental content and beliefs are things that properly have their own distinct norms. Now the picture of normativity that emerges from these debates is importantly different from moral normativity in that there is nothing like the close connection between normative judgements and blame-like sentiments.

This is because exactly what these non-moral norms require from us is not always *transparent* to us. For example, consider the norm of assertion. The two most influential accounts of the norms of assertion, are the Knowledge Account ('assert only what you know') and the Truth Account ('assert only what is true').

Now since it is not always transparent to us *whether* some proposition *p* is true or known, it follows for both accounts that the fundamental norm of assertion is structured such that it is possible to misunderstand, through no fault of our own, what it is that we may and may not assert. If we are fooled into thinking we know *p* when *p* is false, then we will have violated the norm of assertion if we assert that *p*, even if we had excellent reasons for thinking we knew that *p* (and hence that *p* was true).

I claim that some epistemic norms are non-transparent in exactly this way: we may, through no fault of our own, be mistaken about what epistemic norms require from us in some specific context. As such, we may also blamelessly fail to comply with those norms. The resulting beliefs, I argue, are *both* irrational and blameless.

Now the use I put this account of epistemic norms to, as I have mentioned, is to develop and defend a new characterisation of the a priori. I call the account 'Seeming-Independence'.

Seeming-Independence might be described as follows: while a posteriori beliefs depend epistemically on how it *perceptually* seems to us, there is no

2

corresponding dependence relation between a priori beliefs and how it *intellectually* seems to us. Intellectual seemings, or intuitions, simply do not play the kind of epistemic role that perceptual experiences play.

The fact that a priori beliefs are not epistemically dependent on intuitions, I argue, indicates that the norms governing non-inferential a priori beliefs are—unlike those governing perceptual belief—*truth guaranteeing.* Insofar as one really does correctly follow those particular epistemic norms, one is guaranteed to arrive at true belief.

Now the major contention of this thesis is that Seeming-Independence is a *theoretically fruitful* view of the a priori. The arguments I marshal in favour of Seeming-Independence are in this way primarily focussed on the explanatory power and flexibility of the view.

In effect, what I suggest is that Seeming-Independence, unlike some of its rivals, is a particularly clear way of dividing the a priori from the a posteriori, and it allows us to neatly bypass some of influential criticisms of a priority.

Since Seeming-Independence is fundamentally a claim about epistemic norms, one advantage it has is that it uses a *clearly* epistemological criterion in dividing the a priori from the a posteriori. It is well positioned to respond to criticisms that the a priori/a posteriori distinction is ad hoc or incoherent.

Another advantage of Seeming-Independence is that it does not wed itself to an implausible philosophy of mind. Seeming-Independence does not commit us to the existence of any kind of strange, perception-like faculty of priori insight. Seeming-Independence is also perfectly consistent with the thought that humans are limited and fallible reasoners.

The plausibility of Seeming-Independence, however, depends crucially on whether the idea of a truth-guaranteeing norm is intelligible. The latter part of the thesis is devoted to exploring whether it is.

This exploration leads us relatively far afield: into discussions of *analyticity* and discussions of *self-knowledge*.

Essentially, the plausibility of Seeming-Independence depends crucially on whether or not we can make sense of an unusual property; a property that I call *self-evidence.* Seeming-Independence ultimately presupposes that self-evidence is actually a property that some propositions have. And this presupposition would be illegitimate without any explanation of what self-evidence is.

I suggest that there are two readily available, paradigmatic examples of self-evident propositions: those propositions that are analytic and those propositions that we can know *introspectively.*

Now neither suggestion is entirely unproblematic. While Paul Boghossian has developed a new account of analyticity that appears capable of avoiding the problems that Quine famously raises, this version of analyticity—epistemic analyticity—has recently encountered a major set-back. Timothy Williamson has developed a recipe for generating decisive-looking counter-examples to it.

The account of analyticity Boghossian advocates depends crucially on the claim that there are links between understanding analytic propositions, and assenting to them. Failures to assent entail failures of understanding.

Williamson offers several very plausible-looking counter-examples to this suggestion: examples of highly competent language users who, as a result of suitably bizarre background beliefs, steadfastly refuse to assent to a number of paradigmatically analytic propositions.

The claim that our own beliefs, desires, intentions, and so forth are self-evident to us likewise faces challenges. While it is widely agreed that self-knowledge of the sort I am interested in seems to have many of the same epistemological characteristics that self-evident beliefs do, there is decidedly less agreement over whether it really does have those characteristics. There are a number of *prima facie* plausible construals of self-knowledge that deny that it genuinely does have any special epistemic qualities.

Moreover, the suggestion that self-knowledge is self-evident would imply that Seeming-Independence is committed also to the view that our knowledge of our own contingent intentions, beliefs, etc., counts as genuinely and substantively a priori, a view which some might find unacceptably implausible.

4

I think that both sets of challenges can be overcome. I argue that the account of normativity developed here allows us to *legitimately* limit, in a non-ad hoc way, the proposed links between understanding and assent so that they do not apply to examples like those that Williamson develops. This allows us to acknowledge that Williamson's examples are indeed examples of people who understand analytic propositions without assenting to them, but yet nonetheless insist that they are not counter-examples to a *more plausible* conception of the links between understanding analytic propositions and assenting to them.

I also claim that attention to the way in which our higher-order judgements about our own present-tense attitudes interact with the attitudes themselves provides us with excellent reasons to think that the attitudes in question are most plausibly understood as self-evident when viewed (so to speak) from the first-person perspective.

I further suggest that the idea that self-knowledge of this sort is a priori is not nearly as unacceptable a conclusion as we might think. On the contrary, being able to appeal to the a priori provides us with an especially clear solution to what I call the Problem of Self-Knowledge. The Problem of Self-Knowledge arises from the fact we seem to treat people as if they know what they think in a special way, without any clear idea of *how* they could have this special kind of knowledge.

I argue that the claim that this special kind of knowledge is a priori knowledge is a surprisingly satisfactory answer to this kind of 'how'-question.

Ultimately, then, there is nothing untoward in Seeming-Independence's reliance on the notion of self-evidence. It remains a plausible-looking and potentially fruitful theoretical model.

1: Epistemic Normativity

Any epistemological theory might be taken to fall into one of two camps—normative and non-normative. A normative epistemological theory holds that whether one knows that *p,* or is epistemically warranted in believing that *p,* depends in part on whether one *ought* to believe it. A non-normative theory is any theory that dispenses with talk of 'oughts'. Typically, a non-normative theory holds that whether a belief is warranted depends exclusively on whether particular kinds of causal conditions obtain.

Amongst those who take a normative approach to epistemology, one's epistemic standing, or warrant, depends crucially on whether one has formed the belief in question in a manner that is 'epistemically rational or responsible' (BonJour 1985, p. 42) or 'reasonable' (Cohen 1984, p. 283).

The intuition behind this thought is most clearly captured by the well-known series of thought experiments Laurence BonJour runs in *The Structure of Empirical Knowledge* (1985)*.* BonJour puts forward a series of scenarios designed to suggest that the reliability of a cognitive faculty cannot *by itself* justify the beliefs that faculty produces. The most compelling of these cases is as follows:

> Norman, under certain conditions which usually obtain, is a
> completely reliable clairvoyant with respect to certain kinds of
> subject matter. He possesses no evidence or reasons of any kind for
> or against the general possibility of such a cognitive power or for or
> against the thesis that he possesses it. One day Norman comes to
> believe that the President is in New York City, though he has no
> evidence for or against this belief. In fact the belief is true and results
> from his clairvoyant power under which it is completely reliable.
> (BonJour, 1985, p. 41)

It is intuitively clear that Norman is unjustified in uncritically accepting this belief, despite the fact that his belief is formed via a reliable faculty. BonJour argues, persuasively, that this is because Norman was not properly rational in the manner in which he acquired his belief. Given his background beliefs, he is not rationally entitled to rely upon his faculty of clairvoyance in the way he has.

A great many philosophers have taken this to conclusively demonstrate that necessarily, if one is warranted (justified) in believing that *p,* then one has formed and maintained the belief that *p* in a way that *is fully rational* or *epistemically responsible*.

What I want to call attention to in this chapter is how the nebulous terms 'epistemically rational', 'responsible', and 'reasonable' are understood in this context. Interestingly, philosophers within this normative tradition frequently seem to understand being epistemically rational as equivalent to being *epistemically blameless.* In this chapter I will argue that this is a mistake: there is good reason to keep epistemic blamelessness distinct from epistemic rationality.

Let me make one final terminological point. Careful readers may have already noticed that I sometimes use the terms 'warrant' and 'justification' interchangeably. However, I do not take them to mean exactly the same thing: I loosely follow Burge (1993) in understanding justification to be a subspecies of epistemic warrant. In my view a belief is *warranted* if it possesses the sort of positive epistemic status that is usually sufficient for knowledge.[1]

A justification is the kind of epistemic warrant that is cognitively accessible to us. A person has a justification for believing that *p* if she believes that *p* on the basis of some *articulable* reasons: a paradigmatic example of a justification for the belief that *p* would be an argument of some sort for *p*. I mark the distinction between warrant and justification because I leave open the possibility that sometimes a believer may be epistemically *entitled* to hold some belief even if she cannot think of a justification for it. A belief that we were entitled to in this sense would count as a warranted belief.

---

[1] The *unusual* cases where warrant is not sufficient for knowledge I take to be the kind of cases Gettier famously called attention to: cases where the belief in question is warranted but is only accidentally or luckily true.

1.1 Epistemic rationality and epistemic blame

Let us begin by setting up what is meant by 'epistemically rational' (or 'responsible') and 'epistemically blameless'. Epistemologists who use expressions like 'rational' 'responsible' or 'reasonable' tend to use them interchangeably, and seldom seem to take there to be any important distinctions between the terms. Stewart Cohen, for instance, explains that he thinks 'rational' and 'reasonable' are both *virtually synonymous* with 'justified' (Cohen 1984, p. 283). In the context of this normative tradition, then, both 'rational' and 'reasonable' are used to pick out a person who has fully complied with the relevant epistemic norms.

'Epistemically responsible' is likewise seldom used in a way that indicates it means anything other than this. I take it that the key idea is that being properly responsible in this epistemic sense is much like being properly responsible in the everyday, conversational sense of being *a responsible person*. When we describe a person as being a responsible person, what we mean to indicate is that she is someone who reliably meets her responsibilities. If she is entrusted with certain duties, she will typically discharge those duties.

Someone behaving in an epistemically responsible way, then, is someone who is successfully meeting, or complying with, the relevant epistemic norms.

Before moving on to discuss epistemic blamelessness, let me properly articulate what it is to comply with an epistemic norm. I take the notion of an epistemic norm to be a fundamentally deontological notion. When believing rationally, we:

> [R]ely on a set of epistemic rules that tell us in some general way
> what it would be most rational to believe under various epistemic
> conditions. We reason about what to believe, and we do so by
> relying on a set of epistemic rules. (Boghossian 2008, p. 1)

An epistemically rational belief, then, is fundamentally a belief we arrived *via* reasoning that (in some sense) relies on a special set of rules. This is what it is for a belief to be in compliance with an epistemic norm.

Now, my use of the term 'epistemically blameless' originates from Paul Boghossian. Boghossian explicitly takes epistemic blamelessness to be equivalent to being epistemically responsible. He says:

> Our robust response to [BonJour's thought experiment] is that Norman is not justified. And a plausible and widely-accepted diagnosis of our response is that we are reluctant to regard someone as justified in holding a given belief if they are being epistemically irresponsible in holding that belief. Being justified is, at least in part, a matter of being *epistemically blameless*. (Boghossian, 2003a: p. 228. Emphasis mine)

Here Boghossian treats 'epistemically irresponsible' and 'epistemically blameless' as antonyms. In the very next paragraph, Boghossian indicates that the lesson he draws from BonJour's example is that 'being justified excludes being *epistemically blameworthy*' (Emphasis mine). So Boghossian, it seems, does not take there to be an important distinction between being epistemically rational and being epistemically blameless.

But how are we to understand epistemic blame? Boghossian does not offer much elaboration on this. He sets up epistemic blamelessness simply as follows: 'If someone is epistemically blameless in believing something, then it makes no sense to criticize him for believing it' (Boghossian, 2001, p. 18). When we talk of *blaming* people, then, we are talking about adopting— or being disposed to adopt— a kind of critical attitude or sentiment towards them.

Yet I take it that the notion of blame Boghossian has in mind is a little more sophisticated than that brief analysis suggests. I take it that if we *blame* a person for φ-ing, then it follows that we think—or are disposed to think— that she *ought*

not to have φ-ed. That is certainly the sense of blame that the above quotation generates.

But not all criticism has that normative implication. Consider an art critic who criticises the artist for having produced an ugly painting. This sort of evaluative criticism might have *some sort* of normative implication; for instance, it seems to follow from the painting being *ugly* that it is a *bad* work art.

However, it does not obviously follow from a painting's being ugly (or bad) to it being a painting that the artist *ought not* to have painted. This is because it is not obvious that the painter is under any obligation to produce only good art. And without that obligation, it is difficult to understand how it could be that she ought not to have painted what she did. 'Ought' is a fundamentally deontic notion, while 'bad' is fundamentally evaluative. Deontic and evaluative notions might well both be *normative* notions, but it is not obvious that there are any straightforward logical implications between them.

Blame, we can conclude, involves only a specific type of criticism: what we might call deontic criticism as opposed to evaluative criticism. So a charitable reading of Boghossian would be to understand him as claiming that to blame a person for φ-ing just is being disposed to *deontically criticise* her for doing so: taking her to have done something that she *ought* not to do.

And I consider taking someone to have done something that she ought not to have done to be fundamentally a matter of taking her action to reflect on the person in a particular sense. That is to say, deontically criticising a person for φ-ing is more than simply considering her having done so to be *unfortunate*: it involves taking her to be what we might call *deeply responsible* for the unfortunate result. The clearest, most discussed, example of deep responsibility of this sort is moral responsibility. Consider the following quote from R. Jay Wallace:

> People who are morally responsible are not seen as acting in ways
> that happen to be good or bad; they are not just causally responsible
> for certain welcome or unwelcome happenings... Rather, the actions
> of morally responsible people are thought to reflect specially on

10

them as agents, opening them to a kind of moral appraisal that does more than record a causal connection between them and the consequences of their actions. (Wallace 1994 p. 52)

To morally blame a person for something, then, is to hold her responsible for it in this deep sense: it is to consider what she did, or brought about, to *reflect poorly* on her as a moral agent.

Epistemic blame is the epistemic equivalent of this stance. To epistemically blame a person is to hold her responsible for her beliefs: to take these beliefs to reflect poorly on her as an epistemic agent in this 'deep' way.

Above I indicated that Boghossian—quite rightly—takes being epistemically blameworthy in this sense to entail that one has failed to comply with the relevant epistemic norms: the norms germane to one's epistemic situation. However, given his treatment of 'epistemically blameless' and 'epistemically irresponsible' as antonyms, we might also read him as committed to a stronger view. This would be the view that being epistemically *blameless* likewise entails that one is epistemically rational: that one really has complied with the relevant epistemic norms.

Now while Boghossian is unusual in framing this point in terms of blame, I take the idea here to be not uncommon. For instance, consider Cohen's 'New Evil Problem' for reliabilism. Cohen (1984) argues that it is deeply implausible to suppose that a victim of a Cartesian evil demon, who goes about forming his beliefs in a perfectly sensible manner, is not *justified* in holding his perceptual beliefs. While it is clear that such a person would not have perceptual *knowledge*, Cohen insists that the manner in which he forms his beliefs is sufficient for his being fully justified, despite the fact that his perceptual faculties are—thanks to the demon's interference—entirely unreliable. Cohen sets up the argument for this conclusion as follows:

My argument hinges on viewing justification as a normative notion. Intuitively, if S's belief is appropriate to the available evidence, he is

> *not to be held responsible* for circumstances beyond his ken. (Cohen 1984, p. 282, my emphasis)

The intuition Cohen is trying to exploit here is the intuition that there is a connection between not being able to hold a believer responsible for his (false) belief, and his being justified. Given we cannot hold S in any way responsible for his belief being false, Cohen urges, he cannot plausibly be said to lack justification.

Now how should we understand the relevant notion of 'holding responsible'? I take it that this is clearly to be understood in exactly the way I have suggested we understand Boghossian's notion of epistemic blame. I characterised epistemic blame as involving holding a believer 'deeply responsible' for her beliefs. It is difficult to see what other idea Cohen could have in mind here if not that.

So I suggest that Cohen is claiming the same thing that Boghossian is claiming: that if believers go about forming and maintaining their beliefs in a manner for which they cannot be blamed, criticised or held responsible, then they cannot plausibly be thought to be unwarranted. It is *only* epistemically blameworthy believers, then, that are epistemically irrational or irresponsible. If a believer cannot be blamed, then she has not formed her belief in a way that can be considered epistemically irrational or irresponsible.

The thought that there is no significant distinction to be drawn between epistemic irresponsibility and epistemic blameworthiness runs parallel to a venerable tradition in moral philosophy, and it is this parallel, I suggest, that provides the support for the view that to be epistemically blameless *just is* to be epistemically responsible. Now, according to this tradition, moral judgements are grounded in 'reactive attitudes' like praise and blame. All it is to judge a person to be guilty of moral wrong-doing, for example, is to be disposed to blame them for their actions. Blame, in this way, is a constitutively moral sentiment.

So I take it that the assumption at play here is simply that the same sort of relationship holds between epistemic blame and epistemic rationality: such that to

judge someone epistemically irrational *just is* to epistemically blame them for those beliefs.

However, whatever the merits of this view as a thesis about moral responsibility, there is no reason to think it holds as a thesis about epistemic rationality. On the contrary, as I shall argue, there are strong reasons for thinking that it does *not* hold. That is to say, there are strong reasons to suppose that it is logically possible for one to be simultaneously epistemically irrational, and epistemically blameless.

1.2   Non-moral norms and blame

Epistemic rationality, we have seen, is a matter of complying with epistemic norms. One believes in an epistemically rational manner when one is fully in keeping with the relevant epistemic norms or demands. Being epistemically blameless is a matter of being epistemically unimpeachable or uncriticisable. You are blameless in this sense when you cannot appropriately be faulted for having the beliefs that you have: when your belief does not reflect poorly upon you as an epistemic agent. But there is no reason to suppose that people are *always* blameworthy in this sense for failing to meet a normative demand.

In the past decade, there has been an increasing amount of discussion of *non-moral* norms: for example the norms governing assertion, or the norms generated by linguistic meaning, by mental content or by the nature of belief itself. One of the most interesting developments in these debates is that it has become quite clear that many of the most influential of the theories proposed explicitly allow that a person may, in a variety of ways, *fail* to meet a normative demand and yet still be entirely faultless for this failure. In other words, a person's failure to meet a normative requirement does not entail that she deserves to be blamed for her failure.

So if it is indeed true that judging someone to have failed to meet a moral norm entails blaming that person, that entailment is to be most plausibly understood as arising from features specific to *moral* judgements, or *moral* blame.

There is no incoherence in judging someone has failed to meet some other normative requirement and yet not adopting anything resembling a blame-like stance towards that person.

This is especially clear in the debate surrounding the norms of assertion. The most discussed view in this field is Timothy Williamson's Knowledge Account of Assertion.[2] This is the view that all assertions are governed by the norm 'assert only what you know'.[3] A prominent competitor to this view—indeed the one it sets itself up as an alternative to—is what we might call the Truth Account: the view that assertions are governed by the norm 'assert only what is true'.[4] While the debate is primarily conducted between the Knowledge and Truth Accounts, a third option in the literature is what we might call the Reason To Believe Account: which holds that the norm governing assertion is 'assert only what you have reason to believe'.[5] All three of these accounts maintain that a person might well make an assertion that fails to meet to the fundamental norm of assertion without deserving of any sort of blame, reproach or criticism.

The Knowledge and Truth Accounts both allow that this may occur in the following way. Suppose a speaker asserts that *p* when she has excellent reasons to think that *p* is true, but *p* is in fact false. To return to a familiar thought experiment, we may suppose that the speaker's perceptual faculties are being manipulated by an evil demon in such a way that nothing she seems to perceive actually exists. Since almost all her beliefs are false, almost everything she asserts will also be false (and not known). Hence, for both Knowledge and Truth Accounts, she will almost always be in violation of the norms of assertion. Yet, nonetheless, as Cohen pointed out, it is strongly intuitive that the speaker does not deserve any *blame* for this. We can imagine her being an extremely honest and epistemically careful person, making sure to only make assertions after close attendance to the

---

[2] See Williamson (2000).
[3] Other notable defenders of this view are Keith DeRose (2002) and John Hawthorne (2004).
[4] Williamson attributes this view to Grice (1989). For a more recent defender, see Matthew Weiner (2005).
[5] Note that having *reason to* believe something need not entail *actually* believing it. Jennifer Lackey (2007) is a prominent defender of this account and she maintains—interestingly—that one might properly assert something if one has good reason to believe it, even when one does not in fact believe it.

evidence. Given her exemplary epistemic character, it is plainly implausible to suggest she can be held responsible for her incorrect or inappropriate assertions.

And we do sometimes adopt a blame-like stance towards asserters. For example, consider our stance to *gossips*: people who habitually spread unsubstantiated rumours. Not only is it conversationally appropriate to criticise gossips, there is a strong sense in which the criticism involved is deontic criticism: we regard such people as behaving in a way that they *ought* not to. Our critical stance towards gossips constitutes a blame-like stance.

The Reason To Believe Account might be able to sidestep this sort of example by following Cohen in claiming that such a person would still have reason to believe what she is asserting in this scenario, and hence that her assertions violate no norm. Yet there are other cases that fail to meet the norms of assertion on all three accounts despite the speaker being undeserving of criticism.

The clearest example of this would be a case of justified lying. Consider a Rwandan harbouring Tutsis during the Rwandan Genocide, who regularly lies to roaming bands of Hutu extremists in order to protect the Tutsis. There is nothing *blameworthy* about telling these sorts of lies. If anything, we would be more inclined to criticise anybody who *did not* lie in circumstances like those. In the imagined scenario, not only is the statement not true (and hence not known either) but we can safely assume that the speaker has no reasons whatsoever to believe it. On whichever account of assertion we favour, she must be understood as failing to conform to the norm of assertion.

So on all the available accounts of the norms of assertion, not only is it true that people may blamelessly fail to meet the norm of assertion, but sometimes speakers can even be understood as deserving criticism for *following* the norms of assertion to the letter. So if any one of these three accounts is correct, there is a substantial gap between judging someone to fail to meet a normative requirement, and criticising her for that failure. The two positions are distinct.

The same conclusion follows for anybody who follows Kripke (1982) in holding that linguistic meaning is normative.[6] For anyone sympathetic to that view, both the justified liar and the justified but false asserter will be examples of a person who has failed to meet a normative requirement, but yet is not blameworthy or criticisable for the failure.

To hold that linguistic meaning is normative is typically to hold that linguistic rules are in fact norms: failing to use some term as these rules dictate is to fail to meet a normative requirement. Now any false assertion, according to the standard view, is an instance of an utterance that is linguistically incorrect. The linguistic rules that apply to the term 'cat' indicate that the term is appropriate only when applied to cats. So, if I mistake a cleverly disguised dog for a cat and utter 'there is a cat', then I am using the term incorrectly. Likewise if I deliberately say something untrue: all lies are cases where I am using linguistic expressions incorrectly. So all instances of reasonable but false beliefs, and all instances of rational or moral lying, will count as using language in a way contrary to a normative demand.

It is also widely held that a *belief* is an attitude that generates its own specific norms as well. While the orthodox position here is that belief is fundamentally governed by some sort of truth norm[7], it is not uncommon to think that belief is fundamentally governed by what we might call a knowledge norm like 'believe only what you know'.[8]

Now while the person lying to the Hutu extremists is not in violation of either the truth norm or the knowledge norm, since where she believes the Tutsis to be hidden is where they in fact are, it is clear that the reasonable but false believer is in violation of the norm of belief, on both accounts.

The upshot of this is that if we take *any* of these accounts of the norms of assertion, meaning or belief seriously, then we must conclude that taking a person to be in violation of a norm does not in itself obviously necessitate or entail

---

[6] This debate is slightly different to the debate about the norms of assertion in that here the debate is exclusively about *whether* meaning is normative and not at all about what these norms might be. If it is normative, the norms themselves will obvious.

[7] Philosophers vary in how the details of this claim are to be understood. For example, see Boghossian 2003b, and Wedgwood 2002.

[8] For example, see Williamson 2000.

adopting a blame-like stance towards that person. Which raises the question: why should we assume that this entailment holds for epistemic norms and epistemic blame?

1.3 Normative conflict

At this point, what is required is to make sense of how it is possible to fail to meet some norm while remaining blameless for that failure. In particular, what we need to examine is whether there is any reason for thinking that while blameless normative failure is possible for some kinds of activity, blamelessly failing to meet an epistemic norm is not possible. That is to say, we need to question whether the considerations that suggest that a person might sometimes be beyond reproach despite failing to meet (for example) the norm of assertion, admit of no parallel considerations with regard to epistemic norms.

Here is Williamson's explanation of how what I've been calling blameless normative failure is possible for assertion:

> It is not denied that false assertions are sometimes warranted in the everyday sense that they are sometimes reasonable; the claim is rather that the reasonableness of such assertions is explicable as the joint outcome of the knowledge rule [i.e.: assert that *p* only if you know that *p*] and cognitive considerations not specific to assertion. (Williamson, 2000: 243)

I take it that concluding that an assertion is reasonable in this 'everyday sense' that Williamson talks about is, in effect, to conclude that the asserter cannot be held responsible for asserting falsely. Now the suggestion here is that it isn't *only* whether the assertion accords with the norms of assertion that determines whether an assertion is blameless. Other factors must also be taken into account.

What sort of factors might be relevant here? One obvious answer would be any other normative requirements that might apply to that assertion. In addition to

the norms of assertion, there are also more general norms that our assertions might be answerable to. An assertion is still a speech *act,* and so will still fall under the mandate of any norms that apply to action generally: prudential and moral norms, for example. And it is possible that normative conflict might arise. That is to say, it is possible that a speech act that is entirely correct according to the norms of assertion is also an *immoral* or *imprudent* act. Moral or prudential norms can pull apart from the norms of assertion.

The Good Samaritan lying to the Hutu extremists is a clear example of this. By Williamson's lights, that person is blameless despite deliberately failing to meet the norms of assertion because the decision to assert incorrectly was informed by moral considerations. She judged that it was more important to keep the Tutsis hidden than it was to assert what she knew to be the case. Moral norms in this sense *overruled* the norm of assertion: the relevant moral norms generated reasons to lie that outweighed the reasons to tell the truth that were generated by the norm of assertion.

So one way in which 'other cognitive considerations not specific to assertion' can play a role in determining whether an assertion is blameworthy or not is by providing reasons to act contrary to the norms of assertion that are stronger than the reasons generated by the norms of assertion.

The implication here, then, would be that at least one way of determining whether a speaker is blameworthy or blameless would be to look at *all* her reasons, both for and against, asserting as she did. If she had reasons that trump her reasons for asserting what she knows, then she is blameless. If she does not, then she is blameworthy.

One problem with this suggestion, however, is that it relies on the thought that the reasons generated by one set of norms are *comparable* with the reasons generated by another set of norms. The way I set up the example suggested that the reasons generated by the norms of assertion could be measured against the reasons generated by the relevant moral norms, such that we could decide that, in light of all these reasons, the Good Samaritan had, all things considered, stronger reasons to lie than to tell the truth.

18

This thought is intuitive in some cases, but yet not in others. It is plausible that moral norms take precedence over the norms of assertion. It seems unambiguous what the right thing to do is in the situation described above. Anybody who took the norm of assertion to be stronger than the relevant moral norms would have been blameworthy: the reasons in favour of telling the Hutus what he knows seem to pale in comparison to the reasons in favour of lying.

But even if we allow that this is an instance where the requirements of one set of norms can be unproblematically measured against the requirements of another set of norms, it is not obvious we can do this in all cases. Sometimes in cases of normative conflict the reasons generated by conflicting sets of norms cannot straightforwardly be compared.

Dramatic fiction is rife with characters who must decide between conflicting sets of reasons for acting. In the most interesting versions, what the all-things-considered best thing to do is far from obvious. Sometimes these reasons are generated from plainly distinct sets of normative requirements. For instance, consider Willoughby's decision to rescue Marianne in Jane Austin's *Sense and Sensibility.*[9] While Willoughby had excellent reasons to do as he did—Marianne did need his aid—he nonetheless was in clear breach of etiquette; by the standards of the time, the rescue constituted an inappropriately intimate contact between an unmarried young lady and a bachelor. Did Willoughby act rightly in terms of his best overall reasons? Did the moral reasons entirely trump those generated by the norms of etiquette? This is not so clear, given what happens after this. More interestingly, it is not obvious that Austin herself considers the act to be the best decision in terms of overall reasons. While the act is initially assumed to be unambiguously the best available option, the novel goes on to problematize it. This very act turns out to serve as an example of why those specific standards governing male and female interaction were considered important in the first place. The rescue sets in motion a chain of events that results in disastrous consequences for Marianne. The overall best thing to do in Willoughby's situation, then, is arguably presented by the author as being *indeterminate.*

---

[9] I am grateful to Tom Stoneham for drawing my attention to this example.

Nonetheless, while readers may be inclined to blame Willoughby for a great many of his subsequent actions, I think it fair to say we cannot reasonably blame him for stopping to save a young lady in need. For that act he remains blameless. Given that there aren't any choices that are presented as clearly the right ones to make in a situation like that, Willoughby cannot be blamed despite his serious breach of protocol.

Interestingly then, it looks like it is quite possible to view Willoughby as being blameless, but nonetheless to hold that the choice he faced was an impossible one: that the reasons he had for acting as he did are not in any interesting way comparable to the reasons he had to act differently.

What this sort of case illustrates is that sometimes a person may be blameless for failing to meet a normative requirement even when she doesn't have contrasting reasons that *overrule* her reasons for adhering to that norm. Conflicting normative demands need not generate reasons that are easily comparable. But people caught in such conflicts might plausibly be considered blameless for failing to meet the normative demand they do fail to meet. That is to say, people making decisions in impossible situations are not blameworthy for deciding to comply with one normative demand at the expense of another.

However, is it possible for another set of norms to come into conflict with epistemic norms? And if it is, is it possible that these other norms might either *overrule* the relevant epistemic norms or generate reasons that are incomparable with the reasons epistemic norms generate? It is almost universally held that this sort of normative failure is *not* possible for epistemic norms. While we might successfully justify an assertion by appealing to moral reasons, we cannot successfully justify a belief by appealing to non-epistemic reasons.

Consider a person who acknowledges being in violation of an epistemic norm, but insists on keeping her belief nonetheless. Assuming for the sake of argument that it is even *possible* to adopt that sort of stance, what sort of reason could she offer that might plausibly make her epistemically blameless? There are no obvious, unproblematic candidates, and to attempt to generate such examples is to invite serious controversy. Beliefs, it is nearly universally held, are answerable only

to considerations related to their truth or falsity: other considerations may not be legitimately employed in support of them. Belief, the popular metaphor goes, aims at truth.

A possible example would be a mother who insists on believing her fighter-pilot son is alive despite his plane having crashed in the ocean, and the prospects of his survival being astoundingly slim. We might imagine a mother in such a plight appealing to *prudential* reasons for thinking he is still alive; for example, claiming that she simply cannot bear to face the possibility that he died.

While we might sympathise with her plight, there remains a clear sense in which this belief is unacceptable. A case like this is strongly disanalogous with the case of the Good Samaritan lying to Hutu extremists. In that case we might fully condone his decision to ignore the norms of assertion in favour of more pressing moral obligations. But in this case we cannot condone her ignoring epistemic reasons in favour of prudential ones. These prudential reasons are simply not good enough to overrule the epistemic reasons in favour of believing her son has died.

A better illustration of this point might be found in the standard response to Pascal's Wager.[10] Pascal urges his readers to believe in God by offering reasons to think that believing in God is the best bet. Betting that God does not exist, he argues, is to risk eternal damnation for the guaranteed reward of a slightly more convenient life. Whereas betting God does exist is to guarantee ourselves some unnecessary inconvenience, but we stand to win an eternity in heaven as a result.

It is widely agreed that Pascal offers the *wrong sort* of reasons in this argument: even if we were to agree that his appraisal of the possible risks and rewards is accurate, these reasons would nonetheless not suffice to counter any possible epistemic reasons against the existence God. Even very strong prudential reasons simply cannot outweigh epistemic reasons the way moral reasons can outweigh our reasons to comply with the norm of assertion.

---

[10] I must hasten to point out that the following interpretation of Pascal is something of a straw man, and should not be considered historically accurate. I employ the argument for the purposes of illustration only.

So while we might appeal to normative conflict to explain how it is possible to blamelessly fail to comply with the norm of assertion, then, it cannot so easily be employed to explain blamelessly failing to meet an epistemic norm.

1.4 Subjective and objective norms

Normative conflict, however, is not the only sort of situation in which one might blamelessly violate a norm. It is also possible to blamelessly fail to meet an *objective* norm by misidentifying its antecedent conditions.

An objective norm, in this particular sense, is a norm that specifies what we ought to do *when some objective state of affairs obtains*. So an objective norm will be structured roughly like 'if x obtains, then φ'. By contrast, a subjective norm is a norm that specifies what we ought to do when certain *subjective* conditions obtain. That is to say, a subjective norm will have a structure like 'if it subjectively seems to you that *x*, then φ'.

Paul Boghossian takes the crucial difference between an objective norm and a subjective norm to be that the antecedent conditions of an objective norm are non-transparent. Consider the following comment he makes about a paradigmatically objective norm, the norm 'believe that *p* only if *p* is true':

> What is true is that it will not always be *transparent* how one is to obey the norm [believe that *p* only if *p* is true]. Subjectively speaking, one might well be required by the evidence at one's disposal to believe *p* even if (unbeknownst to one) it is not the case that *p*… But the mere fact that [it] is a norm whose satisfaction isn't transparent doesn't mean that it isn't important, or that it's not a real norm. (Boghossian 2003b, p. 38 my emphasis)

So as far as Boghossian is concerned, what makes this norm objective *in this sense* is that its antecedent conditions are not transparent to us: that we may sometimes be unable to tell whether or not the conditions in which we ought to φ obtain.

Now as the above quote indicates, Boghossian does not take the lack of transparency to indicate that that an objective norm is not interesting or a genuine norm. The example Boghossian uses to illustrate this point is the maxim 'buy low, sell high'. This is straightforwardly an objective norm: it stipulates what stock traders ought to do when the stock price actually is low, and what they ought to do when it actually is high. However this norm is nonetheless the fundamental norm of stock trading: this norm underpins and orientates all activity on the stock market. Not only is it perfectly intelligible to understand this as a norm, but it is also a norm that is centrally important to the enterprise of buying and selling stock.

So there is nothing in principle wrong with insisting that there are indeed objective norms.

Now it is perfectly possible to blamelessly fail to comply with an objective norm. Given that an objective norm's antecedent conditions are non-transparent to us, it is possible that we might sometimes be unable to tell whether they obtain. It is even logically possible that we might have excellent reasons for thinking that they *do* obtain. Objective norms, in virtue of being focussed on how things are rather than how things subjectively seem to us, introduce the possibility that even our best attempts to meet those norms may not succeed. We might thus be simply *unlucky* for failing to comply with objective norms.

Consider a stoke-broker who buys when all the available evidence suggests that the stock is at its lowest, or close to its lowest. If the stock were then to fall significantly further—due perhaps to some unforeseeable event like a natural disaster—the stock broker would have failed to have complied with the norm 'buy low, sell high'.

Yet we cannot *blame* the stock broker for this: his failure to do so is not the result of shoddy research on his part but is rather due to circumstances beyond what we could reasonably expect him to predict. We would consider this stoke broker to be unlucky, not incompetent, in his failure to meet the norm of stock-trading. We cannot hold him responsible for failing to predict that the stock price would fall as low as it did. His normative failure is a blameless failure.

This kind of blameless failure is made possible by the fact that the norm of stock trading is objective.

I take it that this kind of error is not possible with regard to subjective norms. A subjective norm is a norm that stipulates what we ought to do when in certain transparent subjective states: when things seem to us to be a particular way, or when we have evidence in favour of something being the case.

And to say that a state is transparent, I take it, just is to say that it is immediately obvious to us: we are always in a position to tell whether we are in this state or not. Simply considering whether it obtains is sufficient to decisively settle the question, one way or the other.

I take it that we could not blamelessly misidentify the antecedent conditions of a subjective norm. It follows from the transparency of its antecedent conditions that even the most cursory consideration of the question would make it clear whether such states do obtain. So we could only be unaware of being in such a state through negligent inattentiveness: inattentiveness for which we would be *blameworthy*.

This provides a clear explanation of another way in which the non-moral norms mentioned in section two may be blamelessly violated.

As we have seen, belief is commonly claimed to be governed by either some version of the truth norm, or the knowledge norm. Any false belief is in violation of the norm of belief on both of these accounts.

But—as Boghossian made explicit—there are many possible cases where all the available evidence suggests that p is true (and that we are in a position to know that p) when p is false. Suppose that a believer comes to believe that p after she has proportioned her belief to the available evidence, and has generally behaved in an epistemically cautious manner. She is epistemically blameless for this belief, but is nonetheless in violation of the truth norm and the knowledge norm.

But since the norm of belief is an objective norm, then we have an easy explanation for how this is so: she had every reason to think that the antecedent conditions of the norm of belief obtained even though they did not. She carefully

complied with belief-forming procedures that are generally excellent ways of arriving at truth (and knowledge). She had no way of knowing that those methods would be ineffective in those specific circumstances. Like the stock-broker, she is simply epistemically unlucky for her false belief, not epistemically incompetent or inattentive. Since we may not appropriately hold people responsible for failures due entirely to bad luck, we have a clear explanation of why she is not deserving of any epistemic blame, reproach or criticism for her false belief.

The same would be true of the norm of assertion, at least on either the Knowledge or Truth Accounts of assertion.[11] It is not always transparent to us whether *p* is true, and given that factive mental states like knowledge[12] are not transparent either, the norms 'assert only what is known' and 'assert only what is true' are not transparent. So according to these accounts, the norm of assertion is an objective norm, not a subjective one. As such, it is possible for a failure to meet those norms of assertion to be due purely to bad luck, not incompetence.

Thus far we have established that there are at least two ways in which a person might blamelessly fail to meet a normative requirement. A person might blameless fail to comply with a norm if she is caught within a normative conflict, in the ways sketched in the previous section. Alternatively, it might be that the norm in question is *objective,* and she was, through no fault of her own, unable to properly discern whether its antecedent conditions had been met.

We have already seen that normative conflict cannot unproblematically be used to explain blameless failure to comply with an epistemic norm. Even if we allow that there might be other, non-epistemic norms that apply to beliefs, there are no plausible scenarios where epistemic norms would not straightforwardly trump these other norms. Non-epistemic reasons for believing that *p* are always outweighed by purely epistemic reasons for *not* believing it. So, what needs to be established at this point is whether epistemic norms are best understood as

---

[11] We are not yet in a position to judge whether the Justified To Believe account of assertion takes assertion to be governed by a subjective or objective norm. Since I understand a justified belief to be one that complies with the relevant epistemic norms, this will depend on whether we think epistemic norms are themselves subjective or objective.
[12] Assuming for the sake of argument that knowledge is a mental state. If it is not, then it is not transparent for more obvious reasons.

objective norms or subjective norms. So, if we conclude that epistemic norms must be subjective, then we will have good reasons to suppose that epistemic norms are importantly different from norms like the norm of assertion, and the norms generated by the nature of belief, or by mental content. So, the fact that we might blamelessly fail to meet those sorts of norms would not give us reason to think that it is possible to blamelessly fail to comply with an epistemic norm.

1.5 Objective epistemic norms

There does seem to be a train of thought that suggests epistemic norms must be subjective. Boghossian appeals to the subjective/objective distinction to explain how epistemic norms are related to the norm of belief in general. This is what he says about the norm of belief:

> [I]t is because belief is governed by this objective ought ['believe only what is true'] that the less controversial *subjective oughts* hold of it as well. For example: that we ought to believe that which is supported by the evidence and not believe that which has no support; that we ought not to believe p if some alternative proposition incompatible with p has a higher degree of support; that we ought to believe p only if its degree of support is high enough, given the sort of proposition that it is, and so on. All of these familiar epistemic norms are grounded in the objective norm of truth. It is that ought that supplies their rationale, even if it has proven extremely difficult to say… exactly how. (Boghossian 2003b, p. 39 my emphasis)

So Boghossian's view is that epistemic norms are (somehow) derived from the truth norm. The truth norm sets the target for beliefs to aim at, and epistemic norms provide us procedures that will reliably get our beliefs to this target. The analogy used to illustrate how this works is once again the analogy with the norm of stock trading. Boghossian says:

We are often in the position of attempting to comply with some non-transparent norms by following other more transparent ones. Traders on the stock market are attempting to comply with the rule: Buy low, sell high. But there is no direct way to recognize when one's stock price is low relative to the price for which one will be able to sell it. So traders follow certain other rules as means of attempting to comply with the non-transparent rule that truly captures the aim of their trading activity… These are rules that may be followed directly, by doing what the rules call for when their input conditions obtain (Boghossian, 2005, p. 211)

Yet even if we grant for the sake of argument that epistemic norms are followed as an attempt to comply with the objective norm 'believe only what is true', why does it follow that epistemic norms *must* be transparent?

The account Boghossian seems to have in mind is one in which objective and subjective norms perform different functions. Objective norms stipulate the conditions under which the activity is *correct* or *appropriate* while subjective norms serve as a set of guiding principles that illustrate how to meet those conditions.

As a generalisation, that sounds like a highly plausible suggestion. But, I argue, there is no reason to think that objective and subjective norms are *strictly* divided into these two functional roles. It is perfectly possible that certain objective norms could serve as guiding principles for other objective norms.

I take it that a guiding principle must meet two minimal conditions. First, it must reliably lead to the stipulated correctness conditions. Epistemic norms could not be guiding principles for the norm 'believe only what is true' if complying with them was not a reliable way of avoiding false beliefs.[13] Second, it must be easier to

---

[13] This assumption for many will be far too quick. For instance, readers with strongly internalist intuitions might insist that I am illegitimately smuggling in an externalist view of justification by implying that an epistemic norm must *actually* be a reliable guide to truth. An acceptable guiding principle, the suggestion might go, will simply be one that we have *good reason to think* is a reliable guide to truth. This, I think, is too inclusive a requirement: it allows just about *any* inferential rule to count as a proper epistemic norm. This is an unacceptable consequence. I will press this point more thoroughly in chapter two.

see how to comply with a guiding principle than it is to see how to comply with the corresponding fundamental norm. If a norm's antecedent conditions were as difficult to identify as the fundamental, stipulative norm's antecedent conditions, then it would not be of any use as a guiding principle. A guiding principle's purpose is to allow us to make quick and accurate decisions about whether to perform one action or another. Guiding principles must be helpful, not only reliable.

But this does not establish that objective norms *cannot* serve as guiding principles. Some objective states of affairs are significantly easier to accurately identify than others. Suppose there was an objective norm that stipulated that some action, φ, was correct whenever it was performed upon a mammal (if *x* is a mammal, then φ). Yet it is not always easy to see whether some animal is a mammal or not: there are many different species of mammal, many of which look very unlike one another. In order to comply with this norm, we would need to develop a set of guiding principles. But objective norms, like 'if *x* is hairy, then φ', 'if *x* has mammary glands, then φ', 'if *x* is a dolphin, then φ', can plausibly be understood as fulfilling this role. The antecedent conditions identify specific characteristics and instantiations of mammals. Following a set of norms like those would make it significantly easier to recognise when exactly one ought to φ.

But these norms are still objective: their antecedent conditions specify objective states of affairs, and they are not entirely transparent. These antecedent conditions are much easier to recognise than the original norm, but they are not entirely transparent. Conceivably, even a very careful observer could mistake a cleverly disguised shark for a dolphin, and could end up φ-ing when it was not appropriate. But this remote possibility does not make the norm 'if *x* is a dolphin, φ' unhelpful. Just because it is *possible* to think a creature is a dolphin when it is not does not mean that we are not generally reliable at distinguishing dolphins from non-dolphins. The logical possibility of cleverly disguised sharks is no reason to rule out the possibility that this objective norm could serve as a guiding principle.

So Boghossian's claim that epistemic norms are grounded in the truth norm does not rule out the possibility that epistemic norms are best understood as being

objective. In the remains of this section, I shall argue that some of them in fact are best understood as being objective, not subjective.

Here are some possible examples of epistemic norms, all of which have, at one point or another, been suggested by Boghossian himself.

(1) Believe that which is supported by the evidence and do not believe that which has no support (Boghossian, 2003b, p. 39)

(2) Do not believe that *p* if some alternative proposition incompatible with *p* has a higher degree of support (*Ibid*.)

(3) Believe that *p* only if its degree of support is high enough, given the sort of proposition that *p* is (*Ibid*.)

(4) If it perceptually seems to you that *p*, then, all things being equal, you may believe that *p* (Boghossian, 2008: p. 1)

(5) For appropriate Fs and Gs, if you have observed *n* (for some sufficiently large *n*) Fs and they have all been Gs, then, all things being equal, you may believe that all Fs are Gs (Boghossian 2008: p. 1)

(6) If you are justified in believing that *p*, and are justified in believing that if *p* then *q*, then believe *q* or give up one of the other beliefs (Boghossian 2001, p. 2)

Of these six, (4) has the strongest case for being taken as a subjective norm. Its antecedent conditions explicitly refer to subjective states of mind that are transparent. If we interpret 'evidence' and 'support' as *evidence that is available to you* then (1), is arguably a subjective norm as well. But the others are better construed as objective norms.

Consider (2). Even if we grant for the sake of argument that it is transparent to us whether proposition *p* enjoys a higher degree of support than proposition *q*, it is not always clear whether *p* and *q* are incompatible. Is the thesis that persons possess free will incompatible with Determinism? Is Moral Realism incompatible with Naturalism? Is the hypothesis that there are mental states incompatible with

Naturalism? Regardless of how we answer these questions, it is clear that the answers are not transparently obvious. Through the centuries, great philosophers, marshalling carefully considered arguments, have disagreed vociferously on all of those questions. As long as the propositions *p* and *q* are sufficiently rich and complex, it is certainly very possible to be mistaken about whether they are incompatible, no matter how carefully one considers the question. The norm's antecedent conditions are not transparent.

Let us turn to (3). For this to be subjective, the *degree* of epistemic support some proposition *p* enjoys would need to be entirely transparent, and it would also need to be transparent to us what sort of proposition *p* is. The thought behind this norm is that different sorts of propositions require different degrees of evidential support, so in order to judge whether or not the evidence we have for *p* is sufficient for believing *p* we would need to know what sort of proposition *p* is. For example, is it a proposition that requires a high degree of support or a low level?

The answer to that question is not transparent. Answering this question is fundamentally a matter of calculating how *believable* propositions of that nature are: if it is a highly unbelievable proposition, it would require a higher degree of evidential support than a proposition that is highly believable. But working out the believability of a proposition requires us to perform probability calculations, and people can be easily mistaken when doing so. For example, if we were to work out the probability of some proposition by performing a Bayesian calculation, a normally reliable source might have fed us incorrect base rates, or we might have misremembered the base rates we were given, or we might simply have made an error when performing the Bayesian calculation.

And, standardly, people do not perform Bayesian calculations every time they need to make a judgement about the probability of some proposition. Rather, we rely on a number of heuristics. And while these heuristics are largely reliable, they nonetheless can lead to systematic errors (or cognitive biases). For example, one heuristic is the representativeness heuristic, which leads us to assume that a sample will be representative of a larger population. This heuristic is used to explain how Gambler's Fallacy is so widespread. Because people know that the probability

of a random coin toss landing heads is 0.5, we tend to assume that five out of any sample of ten coin tosses will be heads. This leads them to think that a long sequence of tails would increases the probability of the next toss landing heads, when in fact the probability always remains 0.5.

As a result, the question of whether or not the degree of support a proposition enjoys is high enough *relative to the sort of proposition it is*, is not transparent: there are a number of ways we might be drawn into making false conclusions about what sort of proposition *p* is. So the norm is not subjective.

(5) is likewise not subjective. The problem with (5) is the question of whether *n* is sufficiently large is another question that people might well disagree about. Standardly, when we make inductive generalisations we do not actually count the number of Fs we have observed to determine whether our sample size is large enough, but rather rely on various heuristics. And once again these heuristics are systematically unreliable.

Take the availability heuristic. We have a bias towards thinking that events that we *can vividly recall* are more common than they really are: that is, we assume that if we remember something, it must be because it is statistically relevant. So it is quite possible that people might take the number of Fs they have observed to be Gs to be significantly higher than they really are: assuming a G is a sufficiently memorable property it is quite possible for a person to misrepresent to herself the number of Fs she has observed to be Gs.

And even those well aware of such biases might be led into error. It seems plausible that whatever number *n* is, it must be of a sufficiently high proportion of the overall number of Fs. If there are billions of Fs, *n* would need to be substantially higher than it would need to be if there were only several hundred Fs. And we can imagine possible cases where a person is, through no fault of her own, vastly mistaken about the overall number of Fs. If F is the property of belonging to a particular species, and almost all members of that species reside on a continent that has not yet been discovered, one could be led into error regarding whether *n* is high enough, even if one keeps careful count of what *n* actually is.

The same is true for (6) as well. The issue here is that people might be systematically mistaken about what would count as evidence for the truth of an 'if... then...' statement. They might plausibly take themselves to have evidence for thinking that an 'if... then...' statement is true when they do not. As such this norm is not subjective.

The results of the Wason Selection Task establish that people are astoundingly poor at establishing what would count as a possible counterexample to the rule 'if a card has an even number on one side, then it is red on the other side', or at least as that rule is interpreted according to classical logic. Wason found that when presented with four cards (for example, odd number, even number, red, purple) participants were overwhelmingly inclined to turn over not only the even-numbered card and the purple card to test the rule, but also the red card as well, even though the rule does not imply that *only* even-numbered cards have red backs. If this sort of error generalises, a scenario in which a person is radically mistaken about whether the evidence available to her supports the statement 'if p then q' is possible. This norm is not subjective either.

So, Boghossian's claim that epistemic norms are subjective is surprisingly at odds with what he takes to be actual examples of epistemic norms: of the six examples selected here, only two of those putative norms had antecedent conditions that could plausibly be construed as transparent. All of the others expressly pick out states of affairs that we could easily be mistaken about.

This gives us a clear recipe for generating examples of blameless failure to comply with an epistemic norm. All we need do is pick an epistemic norm that is objective, and construct a scenario in which a person displaying an admirable degree of epistemic caution and thoughtfulness, is mistaken about whether the antecedent conditions of the norm obtain.

To resist this move, what would need to be established is that the putative epistemic norms examined here are *not* in fact epistemic norms at all, these are not the rules we in fact comply with when we form beliefs in a rational, responsible manner. And this, I will argue, is not plausible.

1.6 Epistemic norms and doxastic rules

Our working hypothesis about epistemic norms is that they are a set of guiding principles designed to help us achieve the fundamental norm of belief. There are a number of candidates as to what this fundamental norm might be: most commonly it is taken to be one or another version of the truth norm ('believe *p* only if *p* is true'; 'if *p* is true, then believe that *p*' 'believe that *p* if and only if *p* is true' etc.) or, less commonly, the knowledge norm ('believe that *p* only if you know that *p*'). Given that knowing that *p* entails that *p* is true, whichever fundamental norm or norms we prefer[14], it is clear that epistemic norms are to be understood as a set of principles designed at least in part to help us arrive at true beliefs rather than false ones. To perform this function, an epistemic norm must be both *reliable* and *instructive*. That is to say, it has to be the case that following an epistemic norm would reliably result in arriving at true belief. But it also has to be the case that what an epistemic norm prescribes is relatively easy to follow: that its antecedent conditions are relatively easy to identify.

It is important to see, however, that describing epistemic norms as guiding principles is to speak metaphorically, not literally. That is to say, the claim that an epistemic norm is a guiding principle is to be understood in roughly the way that the claim 'belief aims at truth' is understood. Beliefs themselves don't literally aim at anything; this is simply a metaphor that many philosophers have found useful. Similarly, I do not mean to suggest that epistemic norms *necessarily* serve as a set of instructions that we turn to whenever we are in doubt as to what to believe (though some people might well be able to employ them like this). And like the metaphor that belief aims at truth, the claim that epistemic norms are guiding principles may be interpreted in a variety of ways. I take this claim to be consistent with a variety of accounts of epistemic norms, and of rule-following generally. As such I intend this claim to be a very broad, basic point, rather than a substantive philosophical view.

---

[14] I am sympathetic towards it being 'believe that *p* only if *p* is true' but the positive claims I make about epistemic norms will not presuppose this.

Let me explain. Tyler Burge has recently argued for a controversial understanding of normativity that does *not* require the creature complying with the norm to be able to 'represent, appreciate, sense, or be at least subliminally guided by the norm' (Burge, 2010, p. 314). In his view, there are norms governing even non-rational creatures that have nothing like the cognitive sophistication required to appreciate what a norm is, or be influenced by one. Normativity, in his view, is conceptually tied to *agency* rather than rationality, and his account of agency is such that even very primitive organisms possess it.

Yet his account of epistemic norms is, in rough outline, very much like the one sketched here. He says:

> Norms for truth and epistemic warrant, which are constitutively associated with belief, further exemplify norms that are apriori associated with representational function, but that do not depend on agent aim or purpose. All these norms are representational natural norms. I believe that neither the psychology of perception, belief, and inference, nor the epistemology of any kind of belief or inference, can be understood without reference to representational natural norms. None of these norms depends on being set, or acceded to, as goals or standards by individuals. (Burge, 2010 p. 313)

Burge takes representation to be a fundamentally teleological notion; representing is conceptually a purposive, goal-oriented activity. Belief, as a type of representational state, inherits its fundamental goals from the goals of representation in general. And epistemic norms, as a set of norms that apply to belief, are designed to help belief achieve one of those goals.[15] An epistemic norm, in his view, is a standard that is 'in some way adequate for fulfilment of [that] function or purpose' (Burge 2010, p. 311).

---

[15] Interestingly, in Burge's view epistemic norms are only one of several sets of norms applying to belief.

To say that a belief is in accordance with an epistemic norm, then, is to say that it has *adequately* fulfilled one of its functions. in Burge's view, belief has several functions; the one epistemic norms are derived from is the function of being as *reliably* veridical as possible, given the cognitive and representational limitations of the believer (Burge, 2010, p. 312). Epistemic norms specify the standards a belief must keep to if it is to adequately fulfil that purpose. In other words, epistemic norms are the norms that specify *how* to go about fulfilling that particular function or purpose of belief.

So, according to this view, if we were to formulate an epistemic norm into a general rule of the form 'if *x* then you may believe *p*', it is still the case that '*x*' would need to be a *reliable indicator* of the truth of *p*, and an indicator that is tailored to the representational and cognitive limitations of the believer. That is to say, an indicator that the believer in question is good at picking up on. So while Burge doesn't understand norms to necessarily be principles that *literally guide* beliefs or actions, introducing the metaphor of a guiding principle is still a useful way of describing what epistemic norms do, on his account. I take the claim that epistemic norms are guiding principles to be consistent with this sort of view as well as with the more standard conception, where norms influence the activities they govern more directly.

It is also worth explicitly distinguishing epistemic norms from what we might call *doxastic rules*. I take some rule, *r,* to count as a doxastic rule just in case it is a rule that we do *in fact* follow when forming or revising our beliefs. To say *r* is a doxastic rule is to make a purely descriptive claim. On the other hand, epistemic norms are those rules or principles that we *ought* to be following. To claim that rule *r* is an epistemic norm is to make a normative claim, not a descriptive one.

As some of the arguments employed in the previous section indicate, many of the doxastic rules we follow are not epistemic norms. Empirical research suggests that we employ a number of heuristics, or rules of thumb, in our garden-variety, everyday reasoning. But these heuristics are systematically unreliable in certain circumstances. So how we ought to be reasoning in those cases comes apart from how we generally are reasoning.

Now I take it that anyone denying that the putative norms put forward in the previous section are genuine epistemic norms would need to argue that these are examples of *mere doxastic rules*, not epistemic norms. So what needs to be addressed is how to distinguish a genuine epistemic norm from a mere doxastic rule.

Complicating this task is the fact that both doxastic rules and epistemic norms are fluid, and are liable to be different from person to person. So we will not easily be able to simply list the epistemic norms at play in any given case, and rule that anything not on that list is merely a doxastic rule.

This fluidity is easiest to see for doxastic rules: there is plenty of empirical evidence suggesting that people approach theoretical problems in different ways, and find certain sorts of reasons more persuasive than other sorts of reasons. What one person might consider an obviously decisive piece of evidence, another might dismiss as irrelevant. One possible explanation for this variance is that some types of background beliefs can affect how one reasons. Consider how you might go about getting someone to stop falling for Gambler's Fallacy. One way to do this would be to get them to think about probabilistic reasoning and heuristics. If you can get a person to believe that the probability of a coin toss landing heads is 0.5 no matter what sample group it belongs to, and alert them to the dangers of relying upon heuristics, then you can change how they actually reason. In other words, by introducing a new set of background beliefs, you can change the doxastic rules they follow. What we believe about probabilistic reasoning and the dangers of relying on heuristics can shape how we *in fact* reason in specific circumstances. Our hypothetical subject, once alerted, might well stop relying on the representativeness heuristic when approaching questions such as whether the next coin toss will land heads or tails: this is a real possibility. If that does happen, then the doxastic rule she follows has changed. Given this, we can not only expect doxastic rules to vary between different people, depending on their background beliefs, but we can also expect them to be evolving, changing in light of newly acquired or discarded background beliefs.

These sorts of changes need not happen as consciously as they do in the above example. A novice birdwatcher might find herself deliberately formulating a set of inferential rules to help her identify a bird's species; for example:

(Robin Rule) 'If bird *b* is plump with an orange breast, then, all things being equal, believe *b* is a Robin'.

These might be highly idiosyncratic, and tailored specifically to her own recognitional abilities. If she finds that her hearing is not very refined, and that she simply cannot hear any difference between, for example, a swallow's song and a sparrow's, her list of rules might have purely visual criteria in their antecedent conditions. (Whereas someone with an excellent ear but poor eyesight might formulate rules that have purely auditory criteria in their antecedent conditions.)

As she spends more time looking carefully at birds, she might find herself able to *just see* that a particular bird is a Robin. That is to say, she may, without actually noticing herself doing so, shift from employing her own original inferential rules to something like following norm:

(4)   If it perceptually seems to you that *p*, then all things being equal you may believe that *p*

Doxastic rules, then, can shift on their own, without us consciously replacing them with a different rule.

This example, interestingly, is not just an example of different doxastic rules changing over time. The Robin Rule is just as much an epistemic norm as (4) is. The Robin Rule is both reliable and informative. Its antecedent conditions are easily identifiable, and employing that rule will reliably lead to true belief about what species bird *b* belongs to.

The point here is that there might be a number of different but equally effective guiding principles. Consider once again Boghossian's analogy with the

norm of stock trading: 'buy low, sell high'. As Boghossian says, stock traders might employ a number of other rules as a means of attempting to comply with that one, but, importantly, he also acknowledges that these guiding principles might not be the same in every instance: 'Some will use rules based on technical indicators, others will use rules based on fundamentals' (Boghossian 2005, p. 211). So if this is the model on which we are to make sense of epistemic norms, there is no reason whatsoever to think that people would not be able to generate their own guiding principles based on their own epistemic strengths and preferences, and stick with those. Just as there are a number of generally reliable methods of complying with the norm 'buy low, sell high', there are a number of generally reliable methods of safeguarding your beliefs against falsity.

So we cannot distinguish a genuine epistemic norm for a mere doxastic rule in terms of its idiosyncrasy. While people might have highly personalised doxastic rules, insofar as these are reliable, these cannot be disqualified out of hand from being an epistemic norm.

A doxastic rule may also have the same aim as an epistemic norm. People particularly prone to self-deception may find themselves following a series of deviant doxastic rules, aimed to avoid painful beliefs rather than false ones ('believe that p if not p is too painful to consider'), but not all mere doxastic rules need be like those. Suppose a person acquires the belief that affirming the consequent is a valid argument structure. (Let us imagine she is a struggling first-year logic student who is told this by a very convincing but mischievous logic lecturer who wants to see what lies he can get away with telling). Suppose as a result she comes to follow the following doxastic rule:

If you are justified in believing that if p then q, and are justified in believing that q, then believe p or give up one of the other beliefs

Now the only reason she follows this rule is because she has been told that affirming the consequent is a deductively valid argument form, and she is poor enough at logic to be convinced that this is right. She is still following it *in order* to

arrive at truth. Yet it is not a genuine epistemic norm, even though it shares the same aim as one.

The fundamental difference between a mere doxastic rule and an epistemic norm, then, is that following an epistemic norm is reliable method of arriving at the fundamental norm of belief (or one of its fundamental norms), whereas following a mere doxastic rule may not be. Just as following Gambler's Fallacy might well still lead you to correctly guess the coin toss (after all, guessing Heads after a long run of Tails is not less likely to be correct than guessing Tails) following a mere doxastic rule might also lead to truth. So essentially whether a doxastic rule you follow is one you ought to be following depends on whether it is a good way to arrive at truth. But whether or not a rule is reliably truth-conducive is something that might not be obvious to the person following the rule.

What this suggests is that the in order to establish that the putative epistemic norms in the previous sections are not genuine epistemic norms, one would need to establish the strong thesis that they *cannot* serve as guiding principles: principles we follow in order to achieve the fundamental norm of belief. And this I take to be implausible. The antecedent conditions of all of those rules give every indication of being both sufficiently easy for cognitively sophisticated creatures like us to follow, and of being eminently reliable.

1.7 Conclusion

I have argued that being epistemically blameless does not entail being properly epistemically rational or warranted. There are good grounds for taking epistemic norms to be set up such that it is possible to fail to comply with them despite forming beliefs in an unimpeachable manner. What follows from this conclusion?

The immediate consequence of this is that we must be careful about *how we go about* evaluating a belief as warranted or unwarranted. In epistemology it is common to rely purely on intuition when we make judgements about whether a belief is warranted or not. Some of the great epistemological arguments appeal to our intuitive response to a number of complex thought experiments, involving evil

demons, mad scientists, fake barns and cleverly disguised mules. However it is far from obvious that our intuitions are fine-grained enough to accurately distinguish *warranted* belief from merely blameless belief. There will be a number of cases that philosophers have taken to be examples of properly justified belief when in fact they are merely cases of blamelessly failing to comply with an epistemic norm.

In the next chapter, I will suggest that some of these cases are examples that philosophers have almost universally agreed are clear counterexamples to the claim that a priori warrant is incorrigible. And reinterpreting these examples as non-decisive, I argue, reopens a range of extremely promising dialectical moves in a number of related philosophical debates about the a priori.

2: The Infallibility Thesis Reconsidered

Recent discussions of the a priori have been strongly shaped by the rejection of the thesis that a priori warrant is necessarily a stronger type of warrant than a posteriori warrant. Even the most prominent of modern rationalists endorse a type of a priori warrant that is capable of being *defeated* by empirical observation.[1] This move is notable for two reasons: first, it marks a rare occasion of near universal philosophical agreement, and second, it marks a decisive break with traditional thinking on the subject. Stretching all the way back to Plato, the overwhelming majority of writers on the a priori took a priori justification to be *infallible*: if a belief was warranted a priori, then the belief was true (henceforth, the Infallibility Thesis). This radical shift in the philosophical consensus may be traced back to a number of influential developments, like, for example, Quine's assault on analyticity (Quine 1951), and Kripke's separating the a priori/a posteriori distinction from the necessary/contingent distinction (Kripke 1980). However, what is taken to truly settle the question is the existence of a number of crucial counter-examples. That is to say, it is widely acknowledged that there are clear and decisive examples of people having had purely a priori warrant to believe that *p,* even though *p* turned out to be false.

Yet, as I suggested in the previous chapter, our intuitions about whether a belief is genuinely warranted are unreliable. They are not fine-grained enough to reliably distinguish those believers that blamelessly fail to comply with epistemic norms, from those whose beliefs are genuinely warranted. So it is worth revisiting these putative counter-examples with this distinction in mind.

In this chapter I will argue that these particular counter-examples are not decisive: close attention to the epistemic norms at play in the examples indicates that these beliefs are not genuinely warranted: or at least not warranted in the way that commentators take them to be. Reflection on this result, I argue, suggests a

---

[1] See BonJour 1998, Peacocke 2004, Burge 1993.

novel and promising way of distinguishing the a priori from the a posteriori. The thought will be that we can distinguish a priori warrant from a posteriori warrant in terms of the *structure* of the relevant epistemic norms. This gives us a clear, purely epistemic distinction. The fruitfulness of this suggestion will be explored in more detail in the following chapter.

2.1 Fallible a priori warrant

BonJour (1998) claims that the examples that can be used to establish that a priori warrant is fallible fall into one of three categories. He says:

> In the first place, there are claims in mathematics and logic which though universally regarded as self-evident by the leading minds in the field in question at a particular time have subsequently proved to be false… Second, there are the various allegedly a priori claims of rationalist metaphysicians… Without pausing to list specific cases, it is obvious that all such claims cannot be true, and thus cannot be infallible, if only because of the great extent to which they conflict with each other… Third, and perhaps most obvious, there are the routine errors in calculation, proof and reasoning that are familiar to anyone who has routinely engaged in such processes. Notoriously, even the most powerful minds are susceptible to such slips… [T]here is no reason to think that a degree of care that would ordinarily be taken to be adequate will make mistake impossible. And even if there was a degree of care and attention that would avoid all such mistakes, there would obviously be no way to be sure that it has in fact been exercised in a particular case and thus no reason to regard any particular case of alleged rational insight as infallible. (BonJour, 1998: 111-12)

While it is not clear that BonJour intended this list to be exhaustive, it does nicely capture a range of possible cases, and provide us with a place to start. Let us begin, then, by fleshing out the kind of examples BonJour has in mind.

The crucial kind of example here is the first one. A paradigmatic example of this category is the belief that Euclidean geometry is true: and this example more than any other is appealed to in motivating the thought that a priori justification is fallible and capable of being empirically defeated. While Euclidean geometry was for centuries widely held by the greatest mathematicians to be the geometric system that best described physical space, it has subsequently not only been decisively overturned, but decisively overturned by empirical observation. The problem with Euclidean geometry can be traced to its reliance on the parallel postulate, which serves as one of its axioms. The parallel postulate holds that if a straight line (A) intersects two other straight lines (B and C) such that the sum of the *interior angles* at the two points of intersection is less than the sum of the exterior angles, then B and C will eventually meet if they are extended indefinitely.

Now this postulate was disproved when it was established empirically that General Relativity is true, and hence that space can be bent by gravity. In this way, if A and B were sufficiently far apart, and were extended through space, there is no guarantee that they would meet. If one of them passed by a large planet or star, it could be bent away from the other line. Since Euclidean geometry as a whole depends crucially on the parallel postulate, it is thus falsified.

And yet, the thought goes, there were excellent reasons to believe in Euclidean geometry up until this point, as evidenced by the fact that so many great rational and mathematical minds *did* believe it. Yet the reasons these mathematicians had were entirely a priori. As a result, the epistemic warrant in favour of Euclidean geometry was both a priori and not only defeasible, but defeated.

In the second category are metaphysical but purely a priori arguments with false conclusions. Assuming for the sake of argument that there isn't a God, let us

take a possible interpretation of Descartes' Ontological argument for the existence of God.[2]

> (P1) I have an idea of a supremely perfect being (i.e. God)

> (P2) Existence is a perfection

> (C) A supremely perfect being exists

Now this argument is widely celebrated as being purely a priori[3] but is also widely thought to fail to establish its conclusion. The most common objection is that (P2) is false. But Descartes provides robust defences of both premises both in the *Meditations* and in his replies. As a result, the thought goes, he clearly has *reasons* for believing (P2), even if that premise is false. So Descartes was not entirely unjustified in believing (P2) despite its falsity. But, whatever justification Descartes had, it was entirely a priori. Hence, Descartes had fallible a priori justification for believing that God exists.

The third sort of case BonJour mentions is a case in which a person makes an error in calculation or reasoning. This happens all the time. For example, I sometimes try to calculate my grocery bill in my head as I shop, but quite often arrive at the wrong answer due to making simple addition error at some point. Now it might not be obvious that I am *warranted* in believing that my groceries will amount to £*m*, given that we can attribute the error to simple inattentiveness on my part, resulting from the fact that I was not concentrating properly at the time. However BonJour hypothises, plausibly, that it is possible to make these sorts of errors even when we are displaying the appropriate level of attentiveness and care. Suppose a very good mathematics student sitting an exam makes a silly calculation error at some point, even though she is concentrating properly and doing her best to get the answers right. The level of concentration she is displaying, let us suppose,

---

[2] Descartes (1641)
[3] This is in spite of the fact that the first premise is arguably justified not on the basis of traditional a priori reasoning, but on the basis of introspection. I will return to the question of whether introspection like this counts as a priori in chapter five.

is consistent throughout. So while this level is sufficient for her correctly performing a number of other equally difficult calculations, she nonetheless slips up on this one. Given that she arrives at the incorrect answer only after performing a calculation over which she has taken an appropriate amount of care, it seems plausible to say that she is *justified* in believing it.

2.2   Following epistemic norms

What these examples do clearly establish is that what seems to us to be obvious a priori is not always true. One may be supremely confident that a belief based on a priori reasoning is true even when it is false. However it has not yet been settled whether these examples establish that the Infallibility Thesis is false. It is important to note that the Infallibility Thesis is a claim about the relation between a priori warrant, and truth: it holds merely that if *p* really is warranted a priori, then it is true. This thesis does *not* entail that if *p* seems a priori obvious to subject S, then *p* is true. For *p* seeming a priori obvious to S, is not sufficient for S having a priori warrant for believing that *p.*

What the previous chapter established is that there is a dialectical move available to the infallibility theorist that has not yet been taken seriously: that is, she might claim that these above examples are simply cases of blameless but yet unwarranted belief. If this move is a plausible one, then we have not yet produced counter-examples to the claim that a priori warrant is infallible, or truth guaranteeing.

Now all three of these examples are cases where a person arrives at a (false) belief on the basis of some sort of a priori reasoning. What we need to establish is whether the kind of reasoning exhibited in these examples is in compliance with the relevant epistemic norms. So what needs to be made clear at this stage is how to distinguish reasoning that is in compliance with epistemic norms from reasoning that is not.

However, as I argued in the previous chapter, it is perfectly possible for us to mistake a doxastic rule for a genuine epistemic norm: that is, while it might *seem* to

us that some belief-forming method or inferential principle is perfectly acceptable, it might turn out that we are wrong about this. The sort of reasoning involved in Gambler's Fallacy strikes many gamblers as eminently reasonable: this does not alter the fact that it is not.

What this indicates, then, is that just because a person takes herself to be reasoning in a perfectly correct manner, it does not follow that she is.

So in determining whether the above examples are indeed examples of a priori warranted but false beliefs, we must determine whether the doxastic rule the believer is following in each of the above examples is an epistemic norm or a mere doxastic rule.

At this point is it worth briefly considering what *following* a rule entails. I do not hope to put forward a thoroughgoing account of rule-following here: such a task would take us too far off-course. All I want to make clear is that there is a distinction between following a rule, and acting in a way that is consistent with a rule. Suppose a chimpanzee is put in front of a chess board, and, purely at random, he picks up a pawn and moves it one square further forward. This action is consistent with the rules of chess but the chimpanzee is not *following* the rules of chess: he is simply moving pieces about at random.

For my action to be following a rule, it must be the case that, at the very least, the rule *explains* the action. While it is far from obvious that this is a sufficient condition for following a rule, it is uncontroversial that it is a necessary condition. To say that a rule explains my action is simply to say that I acted as I did at least partly because of 'some appropriate relation' (Boghossian, 2008, p. 10) that obtains between the rule and my action. This leaves it entirely open-ended what the appropriate relation might consist in. Boghossian, for example, thinks that the appropriate relation in question is one that makes the explanation a rationalizing explanation. He says:

> However the notion of acceptance [of a rule] is understood, what is
> important is that, in any given case of rule-following, we have
> something with the following structure: a state that can play the role

of rule acceptance; and some nondeviant causal chain leading from
that state to a piece of behaviour that would allow us to say that the
accepted rule explains and rationalizes the behaviour. (Boghossian
2008, p. 11)

Burge, on the other hand, would find this view to be an example of 'hyper-
intellectualization' (Burge, 2010, p. 314), since some rules are 'basic, natural norms'
that apply not only to us, but also to non-rational, unreflective creatures.[4] Yet he
still thinks that even basic, natural norms must play some sort of appropriate
explanatory role; just not (necessarily?) a rationalizing one:

An individual need not understand or be guided by the norms, or by
any other general principles, even though *general principles help*
explain the individual's actions. Basic natural norms apply to such
agency even if an individual cannot understand or be guided by them.
(Burge 2010, p. 340 my emphasis)

However, he does not specify what this explanatory role may or may not consist in.

Complications about what sort of explanations suffice for rule-following
aside, it is clear that the first step in assessing whether or not a person is genuinely
following a norm is considering whether that norm could plausibly be construed as

---

[4] It would be interesting to see what Boghossian makes of this account. At the end of Boghossian
2008, he speculates that a primitivist account of rule-following with regard to epistemic norms might
be the best way to avoid an infinite regress problem. A primitivist account is one where our reliance
on epistemic norms is not something that itself requires justification; it would be one where: 'we
take as primitive a general (often conditional) content serving as the reason for which one believes
something; without this being mediated by inference of any kind' (Boghossian, 2008: p. 29). The
inference Boghossian is concerned about here is the one that supposedly takes place when we recall
a rule, and infer from it to what we ought to do (or believe) in our current situation. Boghossian
argues that this inference itself requires justification, which sparks an infinite regress: employing an
epistemic norm involves performing an inference from general rule to particular belief, which itself
would require employing another epistemic norm, which in turn requires another inferential step.
Given Burge is at pains to deny that basic norms inform or guide our actions and beliefs in the way
this picture of rule-following suggests, he is offering precisely the primitivist account Boghossian was
speculating about.

an explanation for the person's action. If it could not, then we have established that the person is certainly *not* following that norm.

As such, when we consider these examples of supposedly a priori warranted, but false beliefs, it is important that we keep in mind not only whether their beliefs were a result of a belief-forming procedure that is *consistent* with a genuine epistemic norm, but also whether that epistemic norm could plausibly explain the belief in question. To be genuinely following an epistemic norm, it cannot, for example, be purely accidental that one is reasoning in the way that one is.

2.3   Miscalculations and reasoning slips

So equipped, let us turn to the three examples of warranted but false a priori beliefs. I take it that the third case, the example where a very good mathematics student makes a calculation error of some sort while sitting an exam, is the least persuasive of the three. The worry with this case is that the sorts of considerations BonJour appeals to in motivating the thought that her belief is justified are related primarily to the concentration, attentiveness and care that she is proposed to be taking. But these are all considerations that relate to the question of whether or not the believer in question is epistemically blameworthy. When we are judging whether or not a person can be held responsible for a false belief, amongst the questions we consider will be questions about whether she has been epistemically careful or sloppy, attentive or inattentive. So by pointing out that the student's concentration levels were at a usually acceptable standard, BonJour is pointing to considerations relating primarily to whether she is *blameworthy* for her belief.

And these are not even decisive considerations. Is a student *really* blameless for this sort of miscalculation? This is not readily apparent. My intuitions diverge with BonJour's on this point. If you make a slip in reasoning, then you have, on that occasion, displayed *poor* reasoning. Holding one responsible for that mistake seems perfectly appropriate. If students give incorrect answers to mathematical problems in exams, they are marked down by their examiners. Marking a person down for

her answer to a question, I take it, *is* an expression of epistemic blame. So not only is epistemic blame *appropriate*, it is also commonly *present*.

Even if we were to set this concern aside, once we abandon the thought that being epistemically blameless entails being epistemically warranted, then we have no reason to take the fact that the mathematics student is concentrating properly, and is exhibiting the sort of level of care and caution that is standardly sufficient for correctly performing these calculations, to be indicative of her possessing genuine epistemic warrant.

So there are no decisive considerations in favour of taking her to be genuinely a priori warranted, which is enough to demonstrate that this particular case is not a counter-example to the Infallibility Thesis. We can, however, go one better than this: reflection on the relevant epistemic norms at play in cases like this provide reasons to think that she is *not* warranted in this sort of case. Let us begin by considering what the relevant epistemic norm would have to look like in order for this to be a belief that is fully in compliance with an epistemic norm. One option would be something like this:

(7)   If it seems to you that *x* plus *y* equals *z*, then all things being equal you may believe that $x + y = z$ is true

However, this rule would generally not be relevant in the cases we are discussing. Making calculation errors is far more common in cases where we have no intuitions whatsoever about what the conclusion is: where it does not 'seem' to us one way or another whether the answer is '*z*'. So in fact the sort of rule we are looking for would have to be something like this:

(8)   If after appropriately careful deliberation you arrive at the conclusion that *x* plus *y* equals *z*, then all things being equal you may believe that $x + y = z$ is true

So could (8) count as a proper epistemic norm? One reason to be wary of this rule is that if we combine (7) and (8)—along with an analogous set of rules for subtraction, multiplication, division, and so on—then *almost any* mathematical belief is warranted. If both (7) and (8) are epistemic norms, then the only occasions in which an addition-belief is unwarranted is when all things are not equal in some way; that is to say, when the believer is in some way cognitively impaired. Whenever a person is in her normal operating conditions, all her addition-beliefs are warranted. This suggests that these rules are too inclusive.[5]

This thought is borne out with regard to (8) when we consider that there are many ways in which a person might try to work out what the sum of *x* and *y* is, and not all of them are reliably truth-conducive. If someone has somehow acquired an improper way of adding numbers, the results of her calculations will not be reliable no matter how careful she is with them. As a result, (8) is far too broad.

Suppose that *x* and *y* are multiple digit numbers and that a student is trying to work out the answer on a piece of paper rather than with a calculator. She sets the two numbers out in rows, such that their digits are appropriately aligned in columns, and works down the columns, digit by digit. However, let's suppose she does this the wrong way around. So she works from left to right rather than right to left. Whenever a column adds up to a double-digit number she carries the extra digit to the column to the right, not to the left. So she ends up with a completely inaccurate answer. Now if (8) were an epistemic norm then she would be properly warranted in believing her answer, insofar as she was sufficiently careful in how she went about it. This is clearly an unacceptable conclusion: beliefs acquired via wildly inaccurate reasoning processes are precisely the sort of beliefs an epistemic norm is supposed to legislate against.

Though perhaps this example does not demonstrate appropriately careful deliberation. The fact that this is an unreliable method of adding numbers, it might be argued, should be obvious to anyone who reflects carefully enough on the method itself.

---

[5] Since (7) is not relevant to the type of example under discussion I will set it to one side for the moment. I will return to rules like this in sections six and seven.

Let us grant for the sake of argument that this is true for this case. However, suppose that we change the mathematical problem being attempted to one that requires a more complicated mathematical formula in order to solve (for example, the quadratic formula). We then simply need to suppose that the student in question has been taught an incorrect version of the formula.[6] Perhaps:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

Instead of:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Following the wrong formula is once again the kind of reasoning that epistemic norms are supposed to legislate against: it is something that will not reliably lead one to having true beliefs. But is this an error that should be *obvious* to anyone reflecting on the quadratic formula? That thought would, I take it, be much harder to motivate. Moreover, motivating that thought would run a very real risk of doing too much. If we strengthen the notion of appropriately careful deliberation such that mistakenly believing an incorrect version of a mathematical formula will always entail not having deliberated carefully enough, then it is very implausible to think that one might be sufficiently careful in one's deliberations *but yet still make a calculation error.*

It is also very difficult to see why anyone would want to draw the line at that precise point. If we want to argue that being sufficiently careful in our mathematical reasoning requires us to reflect on the reliability of the formulae we employ in our calculations, why would we want to also maintain that being sufficiently careful does not preclude making simple miscalculations? Drawing the line at such a point seems *ad hoc.*

---

[6] Perhaps as a result of their teacher having made a typographical error when writing it on the board.

Furthermore, doxastic rules for mathematical reasoning that have antecedent conditions specifying the appropriate level of care one takes in performing calculations seem unnecessarily complicated. The appropriate level of care is going to vary across people and situations. People with a very good grasp of mathematics find calculations easier to perform than those with a weaker grasp. Plausibly, they don't—and are not required to—deliberate all that carefully when calculating. It would be an unacceptable view of mathematical justification that held that people who were excellent at performing calculations were less justified than those who are bad at it simply because they do it more quickly and less deliberately.

So working out how much care you ought to be taking with your calculations is then itself a complicated calculation: potentially as challenging as the calculation itself. This suggests that the doxastic rule is a bad guiding principle.

The following reformulation is both more straightforward and more reliable:

(8*)   If you conclude that *x* plus *y* equals *z after following a reliably veridical method of calculation*, then all things being equal you may believe that $x + y = z$ is true

Whether or not a method of mathematical calculation is reliably veridical is something that can be known through armchair reflection, so the rule is not obviously guilty of providing antecedent conditions that are unacceptably difficult to ascertain. This would also rule out beliefs resulting from misunderstood mathematical formulas or calculation methods.

But it would also rule out the sorts of miscalculation BonJour thinks are cases of warranted but false belief. If I have miscalculated then I have not *in fact* performed a veridical calculation. I have attempted to do so, but not succeeded.

Suppose a student employs a perfectly appropriate method of adding numbers but performs a standard miscalculation. For instance, suppose she attempted to work out the sum of 57 and 68 (while displaying the normal amount

of attentiveness) and arrived at 115, having forgotten to carry the '1' after adding the '7' and the '8'. Since she did not carry the '1', it follows that she did not, in fact, follow a veridical method in coming to her answer. The method in question prescribes something like the following:

(8.1)   When adding multiple digit numbers, arrange the numbers in rows such that the digits are properly aligned

(8.2)   If the sum of a column is more than one digit, leave the last digit at the bottom of the column and carry the first digit to the top of the first column to the left

Since she has not stuck to (8.2) she has not in fact followed this particular method: she has merely *attempted* to follow it. Her belief-formation process does not count as an instance of following rule (8*). And if we modify (8*) to accommodate these sorts of cases, we once again make the rule unacceptably broad. Suppose we change it to something like:

(8**) If you conclude that x plus y equals z *after attempting* to follow a reliably veridical addition method, then all things being equal you may believe that $x + y = z$ is true

The problem with this clause is that, generally, people using non-veridical methods of adding numbers still believe their method to be veridical. Plausibly, they are still *trying* to follow a veridical method; they just have false beliefs about how to accomplish that.

As a result, we have reasons to doubt that miscalculations can count as legitimate examples of properly justified but false a priori beliefs. Attempts to characterise the relevant epistemic norms in a manner that would allow beliefs based on mathematical miscalculations to be properly justified make justification

far too easy: these characterisations allow in beliefs that are exactly of the sort of belief epistemic norms are supposed to safeguard us against.

Miscalculations are far more plausibly understood as failed attempts to follow an epistemic norm. Depending on the circumstances under which it was performed, the miscalculation may perhaps qualify as an *epistemically blameless* failed attempt, but a failed attempt nonetheless.

2.4 Fallacious arguments

So let us turn to the second type of case BonJour mentions: cases of purely a priori philosophical arguments for conclusions that are false. The example I used to illustrate this case was the following:

(P1) I have an idea of a supremely perfect being (i.e. God)

(P2) Existence is a perfection

(C)   A supremely perfect being exists

I have chosen to focus on a straightforward (putatively) deductive argument, because these are the sorts of examples I take BonJour to have in mind: the most celebrated and discussed a priori philosophical arguments are set up as deductive proofs of their conclusions. While non-deductive a priori arguments are possible—I return to these arguments later— these are not as common or well-known as their deductive counter-parts.

In attending to this type of example, I will be making the following assumption about the philosopher putting forward the argument. I am assuming that the argument is not used *only* as a rhetorical device designed to convince her readers of the conclusion: the argument also serves to explain *why* the philosopher in question believes the conclusion. What we are considering is whether someone

would be epistemically permitted to believe the conclusion, assuming that she has no reasons for believing the premises other than the ones provided, and no reasons for believing the conclusion other than the premises. So in what follows, I will assume that these arguments and the evidential support offered for the premises are the only grounds the philosopher advancing the argument has for believing the conclusion.

Let us begin by rehearsing the reasons there might be to take the believer of an argument like this to be justified. The idea here is that since philosophers like Descartes have clearly thought carefully about this argument, and have attempted to defend it from critics, they must be justified in accepting its conclusion, even if the conclusion is false.

But is this obvious? If the conclusion of a (putatively) deductive argument really is false, then it follows necessarily that either at least one of the premises is false, or the argument is not really formally valid. In this section I will concentrate primarily on the second horn of this dilemma: I shall argue that the conclusion of any fallacious, putatively a priori argument is not *justified* a priori. Since the sort of formally invalid arguments under consideration here are fallacious arguments, it follows that these conclusions are not justified a priori.

However let us first briefly consider the first horn: the possibility that while the argument is formally valid, a premise is false. So, by hypothesis, the argument is such that it is logically impossible for it to be the case both that all the premises are true and the conclusion false, but yet one of the premises is indeed false. Under what circumstances would one be a priori justified in believing the conclusion?

The way a deductive argument justifies its conclusion, in my view, is by *transmitting* epistemic warrant from its premises to its conclusion. If I can establish that conclusion (C) follows logically from premises (P1) and (P2), for example, then what I have shown is that whatever reasons I have for thinking that both (P1) and (P2) are true, these also serve as reasons to think that (C) is true. On this view, (C)'s justification or warrant stands or falls with the justification or warrant for believing both the premises to be true. If it turns out that I have no good reason to think that both of the premises are true (perhaps (P2) is entirely unjustified, or perhaps both

premises enjoy some epistemic support, but there are stronger reasons for thinking that (P1) and (P2) are mutually inconsistent) then I have no good reason to believe the conclusion. Deductively valid arguments do not themselves generate new epistemic warrant. A valid argument is not a *source* of epistemic warrant, it is merely a means of passing it from one belief to another.

Another way to put this point would be to say that the epistemic norms governing deductive inferences *only* permit us to believe a conclusion that follows logically from its premises, if we are *already* epistemically permitted to believe that the premises are all true. Forming beliefs by performing valid inferences from unwarranted premises is a belief-formation method that is in violation of the norms of deductive inferences.

So I could only be warranted in believing the false conclusion of a valid, purely a priori, argument if I was warranted in believing all of the premises: including the false premise. This suggests that if it is indeed possible to be epistemically permitted to believe premises like (P1) and (P2) even when they are false, then it would be possible to be warranted in believing (C) even when it is false.

If an argument is purely a priori its premises must all be justified (if at all) a priori. Now, typically, the standard procedure in defending a premise of a deductively valid a priori argument is not to wheel out a separate a priori argument in favour of it: the most common strategy is for the philosopher defending the argument to make an *appeal to intuition* in motivating her premises. This is precisely how (P2) in the above argument is usually motivated: the defender of the ontological argument will invite us to compare an infinitely perfect being that actually exists, with a non-existent one. She will then urge us to see that the existent infinitely perfect being is even more perfect than the non-existent one, and hence that existence is a perfection.

For the sake of argument, let us assume (P2) is indeed the false premise in the above argument. Could this be justified but false? Answering this question requires some sort of account of how intuitions justify beliefs. More precisely it

requires us to adjudicate over whether a rule like the following could constitute an epistemic norm:

(9)   If it intuitively seems to you that *p*, then all things being equal you may believe that *p*

This is an important question, and one that will be considered in depth in sections six, seven and eight. In the meantime I turn to the other horn of the dilemma; the possibility that the argument in question has some sort of improper structure, rather than merely being unsound.

Now fallacies are widely explained as being examples of *incorrect* or *faulty* reasoning, so immediately there seem to be some intuitive grounds for thinking that fallacious reasoning is reasoning that is normatively improper. Calling a piece of reasoning 'incorrect' is a way of accusing the reasoner of reasoning in a way she ought not to have done. In what follows, I argue that accusations like this are entirely on the money.

Let us illustrate the point by considering an argument that is generally understood to be fallacious. Suppose we were to uncharitably interpret Descartes as replying upon an argument like this:

(P1*) I clearly and distinctly perceive that a benevolent God exists

(P2*) What I clearly and distinctly perceive is true

(C*)   It is true that a benevolent God exists

Let us further suppose that Descartes argues for (P2*) by appealing to the notion of a benevolent God, claiming that such a being would not have made the world such that his (Descartes') clear and distinct perceptions were non-veridical.

Such an argument is not invalid, but it is widely thought to have something in common with an invalid argument in that its premises, like those of an invalid

57

argument, fail to generate good reasons to believe the conclusion. The problem with an argument like this is that it would be of no use in persuading even the most rational, reason-responsive of interlocutors, insofar as that interlocutor doubted the conclusion. No one who rationally doubts (C*) should accept the reasons provided for (P2*). As a result, no rational doubter of (C*) should find the argument persuasive. In this sense the argument is thought to beg the question: it indirectly presupposes what it is supposed to prove.

But would it follow from this that reasoning in accordance with such an argument would be in violation of an epistemic norm? I take it that if a person who *generally accepts* that question-begging arguments are fallacious were to infer (C*) from those premises then she would not be following an epistemic norm in forming the belief that (C*). This would be the logical equivalent of a miscalculation: she accepts that these sorts of inferences are impermissible, but simply failed to recognise that her own inference is an inference of that sort. She has failed to follow the doxastic rule she intended to follow: so her belief cannot be plausibly understood as *actually following* an epistemic norm, no matter what the epistemic norms at play here are.

Suppose, however, that she is fully aware that her reasons for thinking (P2*) presuppose the truth of the conclusion, but that she does not believe that there is anything problematic about that presupposition. That is to say, she disagrees that this argument genuinely is fallacious, and instead takes inferring (C*) on the basis of (P1*) and (P2*) to be epistemically permissible, despite the fact that she is well aware that her warrant for believing that (P2*) is true presupposes the truth of the conclusion. Perhaps she has only come to think that this inference is permissible after endorsing a form of coherentism about justification that would render reasoning even in this tight a circle epistemically permissible.[7]

So construed, this example would be importantly different from the miscalculation case. This isn't a case of accidentally lapsing into an invalid argument, this is a case of someone coming to adopt a different doxastic rule as a

---

[7] It is difficult to think of how one could come up with plausible-looking coherentist grounds for accepting this argument, given the mutually supporting beliefs are so few but let's suppose it can be done.

result of her background beliefs. The difference between her and those who take this argument to be fallacious is a disagreement about whether the above argument is *cogent*.

Now over the last decade, an interesting debate has developed over the conditions under which a deductive argument is *cogent*.[8] There is widespread agreement however, that arguments like the one above are not cogent. Crispin Wright sets up a cogent argument as one where:

> [I]t is possible to *learn* of the truth of the conclusion by getting warrant for the premises and then reasoning to it by the steps involved in the argument in question. Thus a valid argument with warranted premises cannot be cogent if the route to warrant for its premises go – of necessity , or under the constraints of a current epistemic context – via a prior warrant for its conclusion (Wright, 2003: p. 57)

Now, while Wright's phrasing occasionally slips between descriptive or normative language when explaining cogency, Martin Davies is explicit that whether or not an argument is cogent is a purely normative matter: for him, saying an argument is not cogent is to say that inferring the conclusion, on the basis of its premises, is reasoning contrary to the way epistemic norms prescribe. That is to say, the primary issue for Davies is not that a non-cogent argument is not psychologically *persuasive*[9]: the issue is that inference from the premises of a non-cogent argument to its conclusion is impermissible. Davies is less interested in explaining what is wrong with arguments that beg the question; rather he is interested in explaining what is wrong with people who take the premises of question-begging arguments to provide good reasons for believing the conclusion. That is to say, he is interested

---

[8] The leading contributors to this debate are Crispin Wright and Martin Davies. See for example Wright 1985, 2000, 2003, Davies 1998, 2000, 2003, 2005.
[9] Sometimes Wright presents the question as if this is his primary concern. However, I take this presentation to be misleading; Wright is very much concerned with the question of how *epistemic warrant* is transmitted or not transmitted from premises to conclusion, which is a question about the evidential support such conclusions have.

in cases where epistemic warrant does not transmit from the premises of a valid argument to a conclusion, in the way it usually does.

That epistemic warrant may sometimes not transmit from the premises of a valid argument to its conclusion, is a surprising and interesting claim. In their series of papers on this topic, Wright and Davies have each developed and argued for a variety of different *limitation principles* which set out the conditions under which epistemic warrant fails to transmit, despite the argument being deductively valid. While they often disagree about what exactly the various limitation principles are, and employ their limitation principles as a means of resisting the conclusions of a variety of philosophical arguments, they take arguments like the one above as a prime example of a non-cogent argument: question-begging arguments are for them the clearest examples of non-cogent, yet valid, arguments. Wright, as the above illustrates, uses an argument that presupposes prior warrant for its conclusion as an illustration of a non-cogent argument. And Davies offers the following as one of his limitation principles:

> Epistemic warrant cannot be transmitted from the premises of a
> valid argument to its conclusion if, for one of its premisses, the
> warrant for that premiss counts as a warrant only against the
> background of certain assumptions and acceptance of those
> assumptions cannot rationally be combined with doubt about the
> truth of the conclusion. (Davies, 2000, p. 402)

It is important to see that any disagreement about whether an argument is cogent is essentially a disagreement *about* epistemic norms. If Davies and Wright are right that question-begging arguments are not cogent, for example, then it follows that coming to believe (C*) by inferring it from (P1*) and (P2*) is epistemically impermissible. Even if the person in question endorses any inferential rule that allows such inferences, she is nonetheless not following a genuine epistemic norm. Regardless of which participant in this debate have false beliefs about whether such arguments are cogent, that person also has false beliefs about how we ought to be

reasoning in certain circumstances. If those beliefs have directly affected how she actually reasons, then she will have come to follow a rule that is not a genuine epistemic norm. Whatever beliefs arise from her following that rule will not be beliefs that are genuinely justified. We might very well consider here to be *blameless* for this, depending on how plausible we take the false cogency-belief to be, but she will not be properly warranted.

So if Davies and Wright are correct about this, then our coherentist who thinks herself perfectly justified in inferring (C*) is not in keeping with the genuine norms governing deductive reasoning. Like the person who comes to change how she reasons after acquiring background beliefs about probabilities and the dangers of heuristics, her background beliefs (i.e. her acceptance of her version of coherentism) have altered the doxastic rule she follows. However, in this instance her background beliefs have led her *from* following an epistemic norm to following a mere doxastic rule, not the other way around.

In this way, then, it is difficult to see how we might plausibly maintain that someone inferring a false conclusion from the premises of an invalid a priori argument could be properly in keeping with an epistemic norm. Either she accepts that this type of invalid argument is fallacious or she does not. If she does, then her mistake is like a miscalculation and she cannot be construed as successfully following an epistemic norm in forming her beliefs. If she does not, and for some reason takes the argument to be deductively valid, then—as a result of her background beliefs about validity—the doxastic rule she follows is not an epistemic norm. Either way, the resulting belief is not properly justified.

2.5   Euclidean geometry

Let us turn to the most plausible of BonJour's examples: the case of Euclidean geometry. Unlike the products of miscalculations or the conclusions of philosophical a priori metaphysical arguments, for a substantial period of time there was an overwhelming consensus amongst mathematicians that Euclidean geometry was just *obviously* the geometric system that described physical space.

61

This gives us much stronger *prima facie* reasons for thinking this belief to be genuinely justified than either miscalculations, where even the calculator herself takes the reasoning to be incorrect, or the conclusions of philosophical arguments, which other philosophers will frequently take to be incorrect or insufficiently supported. If there was something epistemically incorrect in the reasoning behind the belief that Euclidean geometry was true, it is highly surprising that this went unnoticed for so long amongst those who are generally very rigorous and careful about their mathematical beliefs.

All the same, I will argue that such a belief was not justified, or at least not justified a priori.

The problem with Euclidean geometry, I claimed earlier, was that one of its axioms, the parallel postulate, turned out to be false as a description of space, and this invalidated the entire geometric system. However, axioms are not subject to mathematical proof. There was never anything that was widely taken to be mathematical *proof* that the parallel postulate was true.

The grounds for believing the parallel postulate, amongst mathematicians, was purely intuitive. Yet there is some historical evidence to think that even Euclid himself regarded this axiom as *less* intuitively obvious than his other four axioms or postulates, and that thought seems to have been shared even amongst those who followed him in accepting it as an axiom. The history of mathematics shows a number of (failed) attempts to prove the parallel postulate, suggesting that it was often regarded as being *in need of proof*, an unusual stance to take towards an axiom. Axioms are most commonly treated as self-evident, as propositions that need no further epistemic support. So while the parallel postulate might have been intuitively appealing to those who endorsed it, it seems to have been taken to be less obvious than the other four postulates.

So the question of whether the supporters of Euclidean geometry were *warranted* in believing it depends on whether they were warranted in believing the parallel postulate. And whether they were warranted in believing the parallel postulate depends on whether we are entitled to form beliefs on the basis of the sort of intuitive support that the parallel postulate enjoys. The question here is

about the conditions under which we are entitled to form beliefs *directly* on basis of an intuition like the intuition that the parallel postulate is true.

If we think that belief in the parallel postulate was genuinely warranted, then it looks like we are claiming that something like the following is an epistemic norm:

(9)   If it intuitively seems to you that *p*, then, all things being equal, you may believe that *p* is true

Now the question of whether or not something like (9) can be an epistemic norm has come up before. In section four I argued that this question was at the heart of whether or not those philosophers who employed *cogent* a priori arguments for false conclusions were properly warranted. There, once again, the possibility of warranted but false a priori belief depended on whether beliefs that were directly justified by a priori intuition could be justified. If we accept—as seems undeniable—that philosophical or mathematical intuitions are sometimes false, and we accept that we are epistemically entitled to follow a rule like (9), then we would have an easy recipe for generating false but warranted a priori beliefs. Any a priori belief that is either based directly on an intuition, or is the conclusion of a cogent argument or mathematical proof with premises that are based on an intuition, runs the risk of being false, despite the fact that it is formed entirely in keeping with the epistemic norms at play.

In the next few sections, however, I will argue that (9) is not an epistemic norm.

2.6   Intuitions as intellectual seemings

I take an a priori intuition to be what we might call an 'intellectual seeming'.[10] I take it to be clear that they are such 'seemings', independent of any corresponding

---

[10] See for example, Bealer 1996a, Sosa 2007, Weatherson 2003 for others who take this approach.

beliefs, for the standard reasons. Not all a priori beliefs intellectually 'seem' true to us: '741852 + 9369167 = 1678769' does not seem either true or false to me, even though I believe it is true.

And sometimes it intellectually seems as if something is true even if we do not believe it is. To borrow an example of Boghossian's, it *does* intellectually seem to me that there must be more natural numbers than there are even natural numbers, even though I do not believe that there are.[11]

Accepting that there are intellectual seemings in this sense is not to commit oneself to the existence of some sort of quasi-perceptual faculty of 'rational insight'[12], or to take such seemings to be some sort of '*sui generis*, irreducible, natural propositional attitude which occurs episodically' (Bealer, 1996b, p. 169). While these are some ways we *may* understand an intellectual seeming, they are not compulsory.

All it is to have an intellectual seeming that *p* is to feel intellectually attracted towards *p*; to feel what Sosa calls the 'pull' of certain considerations in favour of *p*. To talk of intellectual seemings in this loose sense is to remain neutral about what *type* of mental state it is, or where it comes from. It is neutral as to whether an intellectual seeming is dispositional or episodic, whether it is generated by a perception-like faculty or simply by the employment of other more commonplace cognitive faculties, like our ability to understand concepts and logical relations. I intend the notion of an intellectual seeming to pick out a familiar, everyday phenomenon, rather than to take a stance regarding what that phenomenon actually is.

It has become increasing clear that our a priori intuitions are not as reliable as we might once have supposed. While Bealer claims that:

[A]lthough different people do have conflicting intuitions from time
to time, there is an impressive corroboration by others of one's

---

[11] See Boghossian 2009.
[12] There is a long tradition of philosophers who endorse that sort of view. See BonJour 1998 for a prominent recent example.

elementary logical, mathematical, conceptual, and modal intuitions
(Bealer 1996a, p. 125)

And even more strongly:

The on-balance reliability of our elementary concrete-case intuitions
is without question one of the most impressive facts about human
cognition (Bealer 1996b, p. 163)

There is a significant body of empirical evidence that suggests the opposite. I have already offered some empirical grounds for thinking that people's logical and probabilistic intuitions are systematically unreliable in a number of different types of circumstances. There is also evidence suggesting this is true about some of our conceptual intuitions. Empirical findings indicate that people's intuitions about the conditions in which a person knows or is justified in believing that *p* vary significantly across culture, across socio-economic status, and across how much philosophy the subject has been exposed to (Weinberg, Nichols and Stich, 2003).

While this may have surprised Bealer, it would not have surprised William Lycan, who thinks 'philosophical intuition is and always will be *laughably unreliable*', and who also suggests that 'some [intuitions] are just mathematical or (worse) philosophical *opinions* that present themselves in the guise of seemings' (Lycan 1996, p. 144, his emphasis).

There is certainly some plausibility in Lycan's suggestion that one's philosophical opinions do shape how things intellectually seem to one. When I was first introduced to philosophy as an undergraduate I had, not atypically, very strong intuitions in favour of Cartesian scepticism and against any type of externalism about meaning or content. And yet both sets of intuitions, again not atypically, have faded over time: I no longer feel anything resembling the intellectual 'pulls' that I did then. The hypothesis that this change in intuition was the result of the various changes in my philosophical views since then is as good as any.

So what does this suggest about a norm like (9)? The rule was:

(9)   If it intuitively seems to you that *p*, then all things being equal you may believe that *p* is true

If this rule is to be a genuine norm, recall, then it must be reliable: following it must generally result in true belief. For that to be the case, our intuitions would need to be reliably veridical. Yet there is empirical evidence suggesting that a great many people have non-veridical mathematical and logical intuitions, and that people with different cultural backgrounds, socio-economic status and philosophical exposure have very *divergent* conceptual intuitions, at least with regard to epistemological concepts. Since these divergent conceptual intuitions cannot all be right, at least some people will have non-veridical conceptual intuitions as well. (9), this suggests, will not be a reliable rule as it stands.

Defenders of (9), I take it, might respond by offering a more precise account of what an intuition is, and then argue that these empirical cases are not largely cases of people with non-veridical *intuitions*; they are cases of people with non-veridical *pseudo-intuitions.* Bealer appears to take this route with regard to the evidence suggesting we have unreliable probabilistic or logical intuitions. He says:

> Intuitions are also distinct from judgements, guesses, and hunches…
> the work of cognitive psychologists such as Wason, Johnson-Laird,
> Eleanor Rosh, Richard Nisbett, D. Kahneman and A. Tversky tells us
> little about intuition in the restricted use of the term relevant here;
> they have simply not been concerned with intuitions in this sense.
> (Bealer, 1996a: 124)

In the following section, I argue that this approach is not tenable.

2.7 Intuition as pure understanding

There are two serious obstacles to this approach. First, arguing that the term 'intuition' is to be understood such that it does not apply in the problematic cases requires a more restricted, and less plausible, conception of what an intuition is. Second, even with a plausible restricted conception of an intuition, the restriction would still make the rule expressed by (9) unacceptably impractical.

What this approach amounts to is the claim that the way in which it *seem*s to some people that the answer to the Wason selection task is to turn over three of the cards, not two, and in which it *seems* to people that a coin is now more likely to land heads than tails given the previous tosses, is simply not the relevant sort of 'seeming'. This immediately raises the question: what is the right sort? Bealer's own view is that an intuition is a *sui generis* episodic mental state entirely distinct from beliefs, judgements, guess and hunches. So while if I have a hunch that *p* or take an educated guess that *p*, there is a sense in which it seems to me that *p*, and this seeming might be distinct from the belief that *p*, but it does not qualify as the intuition that *p*.

Yet this answer requires more motivation than Bealer gives it: as it stands it is not plausible. Why should we take intuitions to be ontologically distinct psychological episodes from hunches? There are no clear phenomenological differences between the two. As Williamson remarks (Williamson, 2007, p. 217), intuitions do not enjoy an especially rich phenomenology in the first place. Phenomenologically speaking, all it is to have an intuition that *p* is to feel attracted towards that proposition, to use Sosa's metaphor. But that feeling of attraction is present also in hunches.

Perhaps Bealer thinks that there is a difference in *degree* of intellectual attraction between hunches and intuitions if not in kind. But that thought is also not plausible: intuitions and hunches both come in varying degrees of strength. If we were to mark a distinction between the two in those terms, it is difficult to see what could motivate us to pick one cut-off point rather than another: there are no obvious phenomenological facts of the matter about where hunches end and

intuitions begin. So construed, the distinction between intuitions and hunches would not be a natural one.

If we are to mark a distinction here, I take it, our best option would be to distinguish intuitions from phenomenologically similar attitudes like hunches and educated guesses in terms of their causal history. Hunches and educated guess, perhaps, are to be understood as products of heuristic-based probabilistic reasoning and various background beliefs and opinions, whereas intuitions are seemings that are the result of some sort of rational insight or *pure* understanding, unaffected by folk-theories or background beliefs.

I will tackle these suggestions in turn. If we take intuition to be the product of a rational insight, the thought would be that we should understand a priori intuitions by means of an analogy with sense perception. Just as vision (for example) produces visual experiences, rational insight produces a priori intuitions. Cases of systematically mistaken intuitions then, would be the equivalent of either illusions or hallucinations.

The trouble with this suggestion, however, is well documented. Perceptual experiences, after all, do not just pop into our heads; they are the product of an easily identifiable perceptual faculty. But the idea of a special faculty of rational insight, that lets us somehow 'see' that certain modal truths obtain, is hard to make sense of. Endorsing such a view would run the risk of making intuition into the product of a mysterious, 'spooky' faculty.

Not all philosophers who endorse the rational insight model of intuition hold that there is a distinct faculty of rational insight: BonJour, for instance, doubts that intuitions are the products of a separate perception-like faculty.

> Clearly and trivially, a capacity or ability is involved, but that this
> must involve a distinct psychological faculty in any more interesting
> sense is anything but obvious. (BonJour 1998, p. 109)

However, if this is the case, it is difficult to see how seriously we should take perceptual metaphor. If all the account commits BonJour to is the 'trivial' fact that

there is some sort of capacity or ability involved in the production of intuitions, then we are no closer to understanding what intuitions are than when we started. As BonJour himself notes, that much was already clear. Calling an intuition a rational insight explains very little, on this sort of account.

So let us consider the more plausible of these two suggestions: the thought that a priori intuitions are simply the products of our ability to understand propositions, perhaps in combination with some other basic reasoning abilities.

On this sort of picture, an intellectual seeming is a genuine intuition only when it is derived purely from our grasp of the relevant concepts and, arguably, the logical relations between them. Pseudo-intuitions might arise when an intellectual seeming is derived not *only* from our grasp of the relevant concepts, but also from whatever folk-theories or background beliefs we might have about the things the concept refers to.

This is an improvement on the rational insight account, as at least we have a clearer, more plausible idea of what intuitions are and where they come from. We already know that we have the ability to understand concepts and logical relations. Intuitions would simply be produced by the employment of some combination of those sorts of abilities. I will discuss the merits of this sort of view in greater length in the next chapter, but for the moment it is worth noting that even if we *do* take this to be a sufficiently plausible account of intuition, the rule (9) expresses is still unacceptably impractical.

Earlier I suggested that the function of an epistemic norm is to serve as a guiding principle for the fundamental norm (or one of the fundamental norms) of beliefs. Given this function, an epistemic norm would need to be structured such that its antecedent conditions, if not transparent to us, had to pick out things that we were generally good at recognising. It would have to be easier for us to follow than a rule like:

(T) Believe that *p* only if *p* is true

If a doxastic rule is no easier to follow than (T) then the rule is of no practical use to us; it cannot play the role that an epistemic norm is supposed to.

On the present account of intuition, (9) is no easier to follow than (T). Hence it cannot play the role that an epistemic norm is required to play.

The issue here is that if our ontology of intellectual seemings is such that some of them are genuine intuitions and some of them are merely pseudo-intuitions, the resulting picture suggests that people frequently confuse intuitions with pseudo-intuitions. Consider the people to whom it seems that the answer to the Wason Selection Task is to turn over three cards, rather than two. No matter how *strongly* it seems to them that this is the right answer, this seeming cannot be called an intuition: rather it is a hunch, or some other sort of pseudo-intuition. This seeming is different in kind from the seeming involved when it seems to me that *modus ponens* is a valid argument form, for example. In the latter case, the story goes, this seeming is just the result of my basic understanding of logical connectives (or something to that effect), whereas in the former case the seeming in question is the result of something else in addition to my understanding of those connectives.

But it isn't obvious that we can reliably tell whether our own seemings really are the result of understanding alone or whether they are infected also by some opinion, folk-theory, or other result-skewing factor. The results of the Wason Selection task suggest that we are in fact generally poor at telling these seemings apart: less than 10% of participants got the answer to the task correct. In many cases, it often is not even obvious how we *could* go about working this out.

Consider the results of the Weinberg, Nichols and Stich study, for instance. If our supposed epistemic intuitions do vary systematically across cultural background, socioeconomic status and level of philosophical training, then, at the very least, the intellectual seemings of *some* of these different groups must be pseudo-intuitions, seemings that are infected by background beliefs. But how do we tell which group, if any, has the genuine intuitions? It is very difficult to see how to settle a question like that. Intellectual seemings are pre-reflective reactions to propositions; they do not wear their causal history on their sleeve.

The problem with this interpretation of intuitions then, is that it implies that (9) is formulated such that its antecedent conditions are *not* properly tailored to our recognitional abilities. This picture of intuitions gives us no reason to suppose that we would be good at telling whether any particular intellectual seeming is an intuition or something else. While the rule might come out as reliably true on this interpretation, it does so at the cost of making the rule impractical. This suggests that (9), as presently understood, is not the correct formulation of the epistemic norm that applies to a priori intuitions.

2.8   Intellectual and Perceptual Seemings

This, I argue, suggests that the epistemic norm governing a priori intuitions is structurally unlike the norms governing perception. Consider a norm like:

(4)   If it visually seems to you that *p*, then, all things being equal, you may believe that *p*

I take it to be uncontroversial that this is a reliable, easy-to-follow doxastic rule. There are no obvious reasons (4) could not serve as an epistemic norm.

And this is because, plausibly, visual experiences are not like intuitions in two key respects. First, we have a clear idea of where visual experiences come from. That is, they are the products of a remarkably veridical perceptual faculty: vision.

Second, they come with a rich, clearly identifiable phenomenology. Standardly we can easily distinguish a visual experience from something like an imagining. While it might well be *possible* for someone to mistake an imagining for a visual experience[13], in all but the most unusual cases, there is a clear phenomenological difference between the two. Perceptual experiences are

---

[13] This might be one way to understand hallucinations, and related phenomena. Kent Bach offers a plausible-looking characterisation of hallucinatory experiences along these lines of this in Bach (1985).

typically both far richer in content and far more vivid than imaginings. As a result, we are generally in a position to know very well when we are perceptually experiencing something and when we are just imagining it.

Intuitions, however, are not like this in either respect. Intuitions do not have a rich or clear phenomenology, and are harder to tell apart from related psychological episodes like hunches, educated guesses and even opinions. Furthermore, we cannot, with any great confidence, take any given intuition to be the product of a reliably veridical cognitive faculty. Even if we can give a plausible account the cognitive faculties that are responsible for producing *veridical* intuitions, empirical evidence suggests that intellectual seemings are subject to a great many cognitive biases. Our immediate, intuitive attraction to some proposition might very well have arisen as a result of social influence, or of our own personal background beliefs, or as a result of unknowingly employing a systematically unreliable heuristic in our thinking about that proposition.

So then how should we understand the relevant epistemic norm? My suggestion is that the norm in question should be something like this:

(10)   If it is self-evident that *p*, then you may believe that *p* without considering any further evidence for it

In the remainder of this section I will offer two reasons in favour of (10). First, (10) fits nicely with our epistemic practices. Second, (10) avoids the pitfalls of (9) by bypassing any mention of how it seems to us.

First, not all of our intuitions strike us as self-evident. But the ones we appeal to in philosophical arguments *do* strike us as self-evident. This is why philosophers appeal to those intuitions in the first place. Consider the sort of examples philosophers take to be typical examples of intellectual seemings:

(a)   1+1=2 (Sosa, 2007, p. 46)

(b)   Nothing can be red and green all over at the same time (BonJour, 1998, p. 100)

(c)   *Modus ponens* is a valid argument form (Boghossian, 2003a, p. 231)

(d)   Gettier's examples are cases where the believer lacks knowledge (Weatherson 2003, p. 1, Williamson 2007, p. 216)

These are all claims that intuitively 'seem' true to many of us. But they are also claims that many take to be *just obviously* true. Some philosophers have even gone so far as to suggest that anybody who declines to assent to (a), (b), and (c) does not genuinely understand what the propositions mean (Boghossian 2001, 2003a, Peacocke 2004). Frank Jackson has implied exactly this about (d) as well: he says:

> I have occasionally run across people who resolutely resist the Gettier cases. Sometimes it has seemed right to accuse them of confusion… but sometimes it is clear that they are not confused; what we then learn from the stand-off is simply that they use the word 'knowledge' to cover different cases from most of us. In these cases it is, it seems to me, misguided to accuse them of error (unless they go on to say that their concept of knowledge is ours). (Jackson, 1998, p. 32)

As far as Jackson is concerned, the Gettier cases are so overwhelmingly obvious that anybody who resists them is either horribly confused, or is using the word 'knowledge' differently from the rest of us.

Direct appeals to intuition, furthermore, are typically not tolerated when the intuition in question is *not* blatantly obvious or self-evident. Consider the following passage:

> I have heard a professional philosopher argue that persons are not their brains by saying that he had an intuition that he weighed more than three pounds. Surely there are better ways of weighing oneself

than by intuition. But such inapposite appeals to intuition should not

be dismissed as mere idiosyncratic misjudgements. They are clues to

the role of the term "intuition" in contemporary analytic philosophy.

(Williamson, 2007, p. 214)

While Williamson might be more disparaging about these sorts of appeals to intuition than most, it is not unusual for philosophers to find fault with arguments that appeal to intuitions they deem suspicious. Sosa, for instance, makes sure to accommodate this thought in his own theory of intuitions: he is careful to point out that an important difference between perceptual experiences and intuitions is that intuitions are *epistemically evaluable*. In his view, it makes sense to evaluate at least some intuitions as themselves warranted or unwarranted in a way that it does not make sense to do so for perceptual experiences. He says:

[A] consideration can be assigned the wrong weight, as it attracts

one too strongly or too weakly. Why should intuitive attractions be

any exception? The sheer considering of a proposition can attract too

much, if for example its attraction derives from enculturation into an

unfortunate bias or superstition. (Sosa 2007, p. 50)

Insofar as Williamson's reaction to that kind of appeal to intuition is not entirely inappropriate, and insofar as Sosa is right in thinking that it makes sense to consider people's intuitions as unwarranted, a doxastic rule like (10) is perfectly in keeping with our epistemic practices.

Moreover, (10) avoids the sorts of problems that (9) faces. If a judgement *really is* self-evident, then it is true. The rule is thus more than reliably veridical.

Furthermore, a rule's appealing to self-evidence in its antecedent conditions is no obstacle to its being easy enough to follow. While I will go into more detail about self-evidence in chapters four and five, for the moment I will say that whether or not a proposition is self-evident does not depend on extrinsic facts about it like its causal ancestry; self-evidence is a property that propositions wear

on their sleeve. While we might not always recognise the property when we see it, it is there to be recognised. And generally we are good at recognising self-evident propositions.

For example, we seem to have no trouble in recognising self-evident propositions whenever we make introspective reports about our occurrent beliefs, desires, sensations and other easily introspectable mental states. I will discuss introspection in more detail in chapter five, but for the moment I will simply note that we seem to have no trouble following a rule like (10) whenever we attribute a relatively wide range of mental states to ourselves. When we consciously and deliberately consider whether proposition *p* is true or false, for instance, it is self-evident to us *that we are considering it*. This is something we *can* know directly, without the need to appeal to any further evidence. And upon such reflection, it is almost always perfectly clear to us whether we do believe *p* or not. When we decide that we believe that *p* as a result of this sort of introspective reflection, it is self-evident to us *that we believe it*. It is *just obvious* to us that this is the case: that we believe *p* (if we do) is self-evident to us, and furthermore it is clear to us that it is self-evident.

2.9 Conclusion

An implication of this is that appeals to *false* a priori intuitions are better understood as failed attempts to comply with (10) rather than successful attempts to comply with (9). They are not, then, genuinely warranted. This is true even if it turns out that we cannot *hold people responsible* for their failure to comply with (10).

And if this is right, then the putative counter-examples to the Infallibility Thesis are not conclusive as they are commonly taken to be. The most plausible of BonJour's three examples are cases where the false conclusion depends on a plausible-seeming but false premise, a premise that it justified, if at all, by an appeal to intuition. This is the scenario we are being invited to imagine when we consider

the believer in Euclidean geometry, by far the most influential and persuasive of the counter-examples to the Infallibility Thesis.

The conclusion that the believer in Euclidean geometry was unwarranted even before General Relativity was established is certainly a surprising conclusion: after all, many of the greatest mathematical minds of their time endorsed that geometric system. How can we plausibly insist that they were all unwarranted?

However, the sorts of considerations we appeal to in motivating the thought that these excellent mathematicians were *justified* in their acceptance of Euclidean geometry are really only considerations relating to whether they are epistemically blameless. That the intuition seemed a plausible one, or that they had no reason to doubt it, or that they took great care when forming their mathematical beliefs, or that they displayed great intelligence in their mathematical reasoning are *not* decisive indicators that the resulting belief is warranted. All this indicates is that there is a strong case for saying that they are epistemically blameless. Given that epistemic blamelessness is not a sufficient condition for epistemic warrant, the intuition that these believers were warranted may be resisted.

What attention to these examples has uncovered then, is that for at least a central range of a priori beliefs, the long-discredited Infallibility Thesis seems to hold. It is an implication of the argument in section eight that the Infallibility Thesis holds for those a priori beliefs that are justified directly, without appeal to evidence: beliefs like 'nothing is red and green all over at the same time' or '1+1=2' are warranted *only if* they are also true. That propositions intuitively seem self-evident to us is not sufficient for their being warranted. In this sense, a priori warrant is not dependent on how things seem to us. The epistemic norm governing 'intuitive' a priori judgements is entirely unlike the epistemic norm governing non-inferred perceptual belief.

So what are the prospects of the Infallibility Thesis? The outcomes of this chapter suggest that it holds for an important range of a priori beliefs. But there are still good reasons to be suspicious of it as a general account of the a priori. The examples examined in this chapter have all been examples of *direct* or *demonstrative* a priori judgements: those beliefs that we either form directly,

without inferring them from any other beliefs, or those beliefs that we take to be logically entailed by other beliefs we hold. Yet it is far from clear that these options exhaust the possible ways of forming beliefs a priori.

It is incontestable that not all inferences that we perform are demonstrative. Sometimes we infer that some generalisation holds on the basis of a number of observations we have made. At other times, we infer that some conclusion is true on the basis that it is the best explanation of the phenomenon we are investigating.

Inferences like these are not infallibly truth-preserving in the same way that the demonstrative inferences examined in this chapter are. Some generalisations require only a single counter-example to be falsified. No matter how many observations we have made, it remains logically possible that there is an unobserved counter-example to the generalisation. Likewise, sometimes the most elegant and plausible explanation of some chain of events is not the correct one. Outrageous co-incidences are logically possible: so no matter how much more likely one explanation appears than another, it remains *possible* that the unlikely explanation is in fact the correct one.

In this way, if it is possible to perform an inductive or abductive inference *from premises that are warranted a priori*, then, arguably, we would have arrived at an a priori belief that has only fallible warrant.

Despite these concerns, the outcomes of this chapter are philosophically significant. As I will argue in the next chapter, the fact that direct a priori judgements are *not* governed by a norm like (9) suggests a novel and fruitful way of marking the divide between the a priori and the a posteriori.

3: A Priority As Seeming-Independence

In the previous chapter I argued that epistemically direct, intuitively obvious, a priori beliefs are governed by a different kind of epistemic norm to the one that governs analogously direct perceptual beliefs. The sort of norm governing perceptual belief may be structured such that its antecedent conditions appeal to what we might call 'perceptual seemings'. Yet I argued that we cannot plausibly construct a norm governing a priori belief-formation that appeals to 'intellectual seemings'. This is because these intellectual seemings differ from perceptual seemings in some key respects: the most important of which is that intellectual seemings are far less reliably veridical than perceptual seemings. Rather, I suggested, the norm that governs such belief formation would be more plausibly understood as something like this:

(10)   If it is self-evident that *p*, then you may believe that *p* without considering any further evidence for it

An implication of this was that it entailed that a central range of a priori beliefs are *infallibly* warranted, if warranted at all. Self-evidence is factive: all self-evident propositions are true. So anybody who correctly follows this norm when forming direct a priori beliefs will arrive only at true beliefs. Any *false,* psychologically direct, a priori belief would at best be a case where the believer has attempted to comply with a norm like (10) but failed. Depending on the attempt, we might judge the belief to be *epistemically blameless*. But it will not be epistemically rational.

And there are several notable examples of propositions that we have taken to be self-evident, but which were not: Euclid's parallel postulate for one. But the fact that we can sometimes be mistaken about whether a proposition is self-evident is no obstacle to (10) counting as an epistemic norm. Epistemic norms, I argued in chapter one, do not have to be transparent. As long as we are *generally*

good at telling self-evident propositions from ones that are not self-evident, there is no problem with (10).[1]

In this chapter, I intend to suggest that this substantial difference between the structure of the epistemic norms governing perceptual beliefs, and the structure those governing direct, 'intuitive' a priori beliefs, offers a theoretically fruitful characterisation of the well-known thought that a priori beliefs are those that are independent of *experience.*

3.1   Experience independence as seeming-independence

The difference between the a priori and the a posteriori is widely understood to be an epistemic difference. To say a belief is a priori is to say that it has a different type of justification or epistemic warrant from a belief that is a posteriori. Yet the fundamental distinction between the a priori and the a posteriori is surprisingly unclear. Traditionally, the a priori is generally distinguished from the a posteriori in terms of its relation to experience: a belief is a priori, the thought goes, if its justification or warrant is in some sense independent of experience. As Kant famously put the point:

> By the term "a priori knowledge", therefore, we shall … understand
>
> not such as is independent of this or that kind of experience, but
>
> such as is absolutely so of all experience. (Kant 1781, p. 28)

Yet it is not obvious what mental state counts as the relevant type of experience and what does not. More importantly, it is also unclear *why* the question of whether or not epistemic warrant depends on some type of mental state should make a difference to the type of epistemic warrant that it is. In this section, I argue

---

[1] Since self-evidence bears some resemblance to Descartes' notion of clear and distinct perceptions, it is worth pointing out that, unlike Descartes, I do not take appeals to self-evidence to do any substantial work in the fight against scepticism. It would be question-begging to claim that since some proposition *p* is self-evident, scepticism about *p* is refuted. A thorough-going sceptic at this point would simply contest the claim that we *can* tell that *p* is self-evident. The claim that we can easily distinguish self-evident claims from those that are not self-evident presupposes that scepticism is false: it does not *establish* that scepticism is false.

that a plausible way to characterise the sense in which a priori justification is independent of experience is to argue that whether or not a person is a priori warranted does not depend in any way on how it *seems* to her. This characterisation, I argue, has the key advantage of making it obvious why the distinction matters: on this reading, the archetypical a priori beliefs are *infallibly* justified, making them importantly different from a posteriori beliefs.

As Kant was well aware, very few beliefs, if any, could be plausibly understood to have arisen entirely independently of experience. Generally concepts are not innate. They have to be learned to be understood, and learning requires experience. Moreover, it is also now widely thought that semantic content can sometimes depend crucially on facts about our physical environment and linguistic community. Being able to think that *water is wet*, for example, is often thought to be possible only if your environment at some point contained samples of H_2O.

So Kant's claim that a priori knowledge is knowledge that is *absolutely* independent of experience must be understood carefully. The idea is often recast as the thought that a priori warrant is *epistemically* independent of experience. One does not need to have any *specific* experience to be warranted in believing that nothing can be both red and green all over at the same time. We need to have had certain kinds of experiences in order to grasp concepts like *red*, and *green*, but experience plays no further role once the concepts have been mastered. As Burge puts the point:

> I understand 'apriori' to apply to a person's knowledge when that knowledge is underwritten by an apriori justification or entitlement that needs no further justification or entitlement to make it knowledge. A justification or entitlement is apriori if its justificational force is in no way constituted or enhanced by reference to or reliance on the specifics of some range of sense experiences or perceptual beliefs. (Burge 1993, p. 458)

Now this is not true of beliefs warranted a posteriori. One of my current beliefs is that I am presently drinking coffee. Now once again, I needed to have certain types of experience to come to understand what coffee is, but I also need other, more specific, experiences for my belief to be justified. If I was not *currently perceptually experiencing* drinking coffee, then I would not be justified in believing that I was.

But it is not entirely obvious whether it is *only* perceptual experiences of that sort that counts as the relevant sense of 'experience'.

For instance, Michael McKinsey says: 'I will call knowledge obtained independently of empirical investigation a priori knowledge' (McKinsey 1991, p. 9). Since an empirical investigation is an investigation premised upon information obtained via the senses, McKinsey takes the a priori to be those propositions that we can come to know that do not depend on information acquired via the senses for their justification or warrant.

But this understanding of experience is controversial. While McKinsey's conception of the a priori was adopted by most participants in the debate about the compatibility of content externalism and self-knowledge that followed that particular McKinsey paper [2], many had clear reservations about doing so. Jessica Brown, for instance, describes this usage as 'not entirely happy' (Brown 1995, p. 149).

The reason for this unhappiness is that this formulation has the consequence of making introspective judgements, by definition, a priori. Now that would be an intriguing consequence, since it entails that there are a great many *contingent* propositions that we can know a priori to be true. That I am thinking about coffee, for instance, is contingent if anything is, and yet McKinsey understands 'experience' in such a way that I can know this a priori.

This strikes many philosophers as implausible. As a result, others understand 'experience' more liberally. Jim Pryor, for instance, says:

> I understand a priori justification to be justification that does not
> derive from occurrent experiences. By 'occurrent experience' I

---

[2] For early examples, see Brueckner (1992) and Warfield (1992).

include perceptual experience and I also include occurrent mental states and processes'. (Pryor 2006, p. 239)

It is important to see that a great deal does hang on which of these two has the right account of experience. If philosophers are free to simply stipulate what counts as experience and what does not then the distinction between the a priori and a posteriori begins to look dangerously ad hoc. If there is a significant epistemological difference between a priori warrant and a posteriori warrant, then the question of whether introspection counts as a priori cannot be settled simply by *stipulation.* Settling the question would require us to consider the epistemic facts of the matter. Is the sort of epistemic warrant enjoyed by introspective beliefs like perceptual warrant, or is it like the type of warrant that the belief that 1+1=2 has?

The point is that if we are to endorse an epistemological distinction like the a priori/a posteriori distinction in the first place, then it is important to ensure, as John Hawthorne puts it, that the distinction 'marks a joint in our epistemological lives' (Hawthorne 2007, p. 218). The more gerrymandered the distinction appears, the less obvious it is why we should be interested in it.

Yet to motivate a choice between whether to include or exclude introspective beliefs in the category of belief based on experience, we need to consider *why* dependence on experience matters.

Now experience, however construed, is a type of mental state. What needs to be made clear is why epistemic dependence on that type of mental state is significant.

I take it that epistemic dependence itself is not the issue here: that a belief depends *on another mental state* does not make the belief a posteriori. Many canonical examples of both a priori and a posteriori belief depend, epistemically, on other beliefs. Both scientific and a priori philosophical theories are generally arrived at via an inference from some other belief or set of beliefs.

So, plausibly, it is something about the nature of experience itself that makes the epistemic difference. But how can this be? Why does whether your

belief epistemically depends on one type of mental state rather than another make a difference to the *type* of warrant that belief has?

What is needed, then, is an account of experience that makes it clear first, why dependence on experience is epistemologically relevant, and, second, why a priori beliefs do not depend upon it.

Now if the account of the epistemic norms governing non-inferential a priori beliefs developed last chapter is correct, then we have a plausible-looking account that can do this. To see this, let us once again compare the following three possible norms:

(4)   If it perceptually seems to you that *p*, then, all things being equal, you may believe that *p*

(9)   If it intuitively seems to you that *p*, then all things being equal you may believe that *p* is true

(10)   If it is self-evident that *p*, then you may believe that *p* without considering any further evidence for it

Consider a perceptual belief, *p.* Now we might say that *p* depends on experience in the sense that whether I am warranted in believing *p* depends in part *on how it perceptually seems to me*. When, in normal conditions, it perceptually seems to me that *p* is true, I am entitled to believe that *p.* Norm (4) entitles us to (in normal conditions) rely or depend upon our perceptual experiences when forming beliefs.

Now if (9) were an epistemic norm, a priori beliefs would depend in exactly the same way on intellectual seemings. So the proponent of (9) would need to motivate the thought that intellectual seemings do not count as experiences. But to do that, she would have to argue that there was an important epistemological difference between intellectual and perceptual seemings.

But, if (10) is the relevant epistemic norm rather than (9), then a priori judgements do not depend epistemically on how things intellectually seem to us at all. There is no need to try to find an epistemological difference between

intellectual and perceptual seemings, given that the norm governing non-inferred a priori beliefs ignores intellectual seemings entirely. So this gives us an unproblematic account of how a priori beliefs are independent on experience. We would simply understand 'experience' in the relevant sense to refer to *how it seems to us*. (Henceforth, I shall refer to this characterisation of the a priori as Seeming-Independence.)

And we also have available an easy explanation of why it matters whether or not a belief depends on experience. This is because seemings are not factive. That it perceptually seems to you that *p* does not ensure that *p* is true. The specific perceptual experience that the belief that *p* depends upon might or might not be veridical. As a result, any belief that epistemically depends in part on a perceptual experience is *fallibly* justified: the epistemic norm at play cannot guarantee truth.

But, if the inferential rule I am following is (10) rather than (9), then I am not entitled to rely or depend upon how it seems to me when forming the belief that *p* in the way I am entitled to depend upon perceptual seemings. Whether or not I am a priori justified in believing that *p* depends, in this sense, only on whether *p* is in fact self-evident. That it seems self-evident to me would be insufficient for the belief being warranted, or epistemically rational. The fact that a priori beliefs don't depend on how it seems to us, then, is epistemologically important: it confers infallible justification on at least those obvious, intuitive a priori beliefs.

This gives us a nice and clear explanation of why dependence on experience should be epistemically important. It also makes it clear how we would go about finding out whether introspective judgements are dependent or independent of experience. The answer to this question can be determined by establishing whether introspective beliefs are governed by a rule like (10), or whether they are governed by something like:

(11)   If it introspectively seems to you that *p,* then you may believe that *p*

My own view is that they are governed by something like (10), but I will leave further discussion of introspection until chapter five.

It's worth emphasising that Seeming-Independence is not the view that the epistemologically important difference between the a priori and the a posteriori is that *all* a priori judgements are infallibly justified. There remains a concern over nondemonstrative inferences involving a priori premises.

But I take it that if we are to find *any* substantive epistemological differences between the a priori and the a posteriori, the sorts of a priori judgements that we should focus on are the direct, non-inferential variety: the ones where their relationship to experience is at its most clear. Focussing on beliefs arrived at via inference from other beliefs is just going to obscure any possible differences. This is because how exactly an inferred belief is supported by the beliefs it is inferred from will vary, depending upon the sort of inferential reasoning involved, and —crucially—it is not obvious that any particular sort of inferential reasoning belongs purely on one side of the distinction or the other. For instance, the paradigmatic examples of inferred a priori judgements are those that are reached via deductive reasoning. This is why it is examples of this sort that the previous chapter was primarily concerned with. But whether a deductive argument is a priori depends upon the warrant enjoyed by its premises. While there are deductive arguments containing purely a priori premises, there are also deductive arguments containing a posteriori premises. Consider this version of Kripke's (1980) argument for the claim that water is necessarily H_2O:

(I)   Water is identical to H_2O

(II)   If a true identity claim is flanked by two rigid designators, then it is necessarily true

(III) 'Water is identical to H_2O' is an identity claim flanked by two rigid designators

(IV)   'Water is identical to H_2O' is necessarily true

This argument is deductive, in that (IV) follows logically from (I), (II) and (III), but it is clearly not a priori. This is because premise (I) is justified a posteriori. This suffices to make (IV) warranted a posteriori. Since a person is only justified in believing (IV) on the basis of this argument if she is also justified in believing *all* of the premises, her justification for believing (IV) depends in part on her justification for believing (I). And since (I)'s warrant depends epistemically on information acquired through perception, (IV)'s warrant depends epistemically on information acquired through perception as well. Epistemic dependence on experience is transferred from the premises of an argument to its conclusion.

As such, it's not obvious how focussing on how the premises of a deductive argument support its conclusion could tell us anything interesting about a priori judgements in particular.

Likewise, while the paradigmatic examples of inferred a posteriori judgements are those reached via inductive or abductive reasoning, whether or not an abductive or an inductive argument is a posteriori once again depends upon its premises. A great deal of philosophy operates abductively; we frequently choose between different philosophical theories on the basis that one is a better explanation of the phenomena in question than the other. As Williamson points out:

> When [philosophers] cannot provide a deductive argument, they still
> offer supporting considerations. Often they cite phenomena which,
> they suggest, their theory best explains: they provide abductive
> arguments. (Williamson 2007, p. 208)

For instance, philosophers might run an abductive argument in arguing for one system of logic opposed to another. In this case all the premises of such an argument would be claims about logic: paradigmatically a priori propositions. On what basis could we claim that the conclusions of such arguments are themselves not a priori? They are not based on experience in any obvious sense.

Mathematics, the other paradigmatically a priori discipline, also admits of inductive or abductive reasoning. Burge points out the following:

> Some mathematical arguments are nondemonstrative, even broadly inductive, yet apriori … If a principle is accepted because its truth would explain or derive a variety of other accepted mathematical principles, the justification for accepting the principle is nondemonstrative; but it may not derive any of its force from perceptual beliefs. (Burge, 1993, p. 461)

As this example illustrates, abductive reasoning has a place in mathematics, just as it has in philosophy. The same is true for inductive reasoning. I might offer an inductive argument for the conclusion that Goldbach's Conjecture is true—that is, that every even integer over 2 can be expressed as the sum of two prime numbers. I might infer from the fact that every even integer greater than 2 *that I have counted so far* is the sum of two primes, that *all* even integers can be expressed as the sum of two primes. Each premise in this argument ('4 is the sum of two primes', '6 is the sum of two primes, '8 is the sum of two primes', '10 is the sum of two primes, etc.), will itself be justified a priori.

The point is that what *makes* the conclusion of an argument a priori rather than a posteriori will have nothing to do with the logical structure of the argument. We can have both demonstrative and nondemonstrative a priori arguments. What makes the conclusion a priori rather than a posteriori is the type of warrant enjoyed by the premises. If we want to find significant differences between the a priori and the a posteriori, the beliefs we should be examining are those that are not inferred from other beliefs. What we want to identify are those epistemic features that a belief has *in virtue of being a priori.* Not those features that are a result of it being the product of one type of inference rather than another.

For it is clear that the premises of a cogent deductive argument support their conclusion differently from how the premises of a good nondemonstrative argument do. The truth of the conclusion of a deductive argument is *logically*

entailed by the truth of the premises. As such if I have very strong reason to believe that all the premises of that argument are true, that reason also serves as very strong reason to think the conclusion is true. But the premises of a nondemonstrative argument only make the conclusion *probable*. The truth of the premises does not guarantee the truth of their conclusion. The reasons I have for thinking the premises are true serve as *less decisive* reasons to think the conclusion really is true. Some degree of epistemic warrant is lost in the transfer from premise to conclusion. No nondemonstrative argument, then, can guarantee the truth of its conclusion. Its logical structure makes the type of warrant produced necessarily *fallible*.

The point here is that the fact that nondemonstrative a priori arguments generate fallible epistemic warrant for their conclusions tells us nothing significant about the a priori in general. Nondemonstrative a posteriori arguments also generate less decisive reasons for their conclusions than deductive a posteriori arguments do. That is just a logical consequence of their inferential structure. But if the *only* examples of propositions that enjoy merely fallible a priori warrant are the conclusions of abductive or inductive arguments, then we have still discovered an epistemologically significant difference between the a priori and the a posteriori. We should not let the fact that no abductive or inductive argument can guarantee the truth of its conclusion blind us to the fact that the central, paradigmatic examples of a priori warranted beliefs *do* enjoy infallible justification. That is enough of a difference to make the a priori/a posteriori distinction epistemologically significant.

3.2 Traditional infalliblism and mathematical knowledge

Now any proposed account of the a priori faces two key challenges. First, it must demonstrate that the distinction is a theoretically significant one: that is, that a priori warrant really is interestingly different from a posteriori warrant. If it turned out to be either incoherent or coherent but ad hoc, then I take it that the account would fail.

On the other hand, however, it must also make its account of the a priori *realistic*: it must make it clear that a priori warrant is something that creatures like us are capable of achieving. If an account successfully demonstrated that a priori warrant was interestingly different from a posteriori warrant, but yet conceived of the a priori such that none or almost none of our beliefs actually *had* a priori warrant, then once again the account would fail. It would be hard to see why an empty or very nearly empty epistemic category should be worthy of philosophical interest.

In this section, I argue that Seeming-Independence is better equipped to meet these challenges than the traditional, Kantian, version.

Philip Kitcher has argued that the traditional conception of the a priori is such that 'a priori warrants are ultra-reliable; they never lead us astray' (Kitcher 1980, p. 9). Kitcher finds this implication in his reading of Kant:

> I believe that Kant's account implies that three conditions should be met. The same type of process [i.e. the process of pure intuition that produces a priori mathematical knowledge] must be available independently of experience. It must produce warranted belief independently of experience. And it must produce true belief independent of experience. (Kitcher, 1980: p. 8)

So on an account like this, there is clearly an important epistemological difference between a priori warrant and a posteriori warrant: a priori intuition is a cognitive faculty that infallibly produces true belief. Given that perception clearly does not do this it gives us a clearly epistemological reason to distinguish those beliefs based on a priori intuition from those based on perception.

In the last chapter I argued that explaining a priori warrant by appeal to this sort of intuition was problematic: there are excellent reasons to doubt that we do possess this sort of mysteriously infallible cognitive faculty. The account is arguably psychologically unrealistic in that sense. A separate problem is that, as Kitcher

argues, on this understanding there are only very few beliefs that could conceivably count as genuinely a priori.

In his view, it is a grave misunderstanding of the epistemology of mathematics to suggest that mathematical beliefs are typically produced straightforwardly from some sort of infallible intuitive faculty. In Kitcher (1983), he notes that mathematical beliefs, as they are in fact formed and maintained, depend strongly on the mathematical traditions in which the believer is located. That is to say, mathematical belief is largely dependent on information acquired through *testimony*. Kitcher takes this dependence to undermine the thought that mathematics is an a priori discipline: given that mathematical beliefs can be overturned by information acquired via testimony, in his view, it does not possess the sort of infallible justification it would need to possess to be properly a priori. So by taking the a priori to apply strictly to those beliefs that are produced purely by some sort of infallible intuitive faculty, we effectively exclude most mathematical beliefs from counting as genuinely a priori. Not only is the traditional account lumbered with the problematic notion of an infallible faculty of intuition, but it also seems ill-equipped to accommodate the actual epistemic practices of the mathematical community.

Now I take it that there is nothing objectionable to the thought that many non-experts are massively reliant on the testimony of experts in acquiring their mathematical beliefs. In fact, it might be that almost all of the more sophisticated mathematical propositions that they believe—say, 'Fermat's Last Theorem is true'—are known *entirely* through testimony if known at all. That is to say, we might expect the general public to have no idea what the actual proof for this theorem looks like; they have simply be told that it has been proven, and accept that at face value. I see no immediate problem with describing this sort of mathematical knowledge as a posteriori. That claim would not, in my view, be devastating to an account of the a priori.

What is far more problematic, however, is Kitcher's claim that *mathematical experts* do not have a priori knowledge of mathematics. Mathematics, as it is practiced by professional mathematicians, is widely considered a paradigmatically a

priori discipline: a discipline that is a priori if anything is. If it were to turn out not to be a priori after all, it would be a strong indication that the category of the a priori was either entirely empty or close enough to it to make it unworthy of so much philosophical attention.

So it seems that the traditional infalliblist approach does well at meeting the first challenge for an account of the a priori, in that it offers a clearly epistemic criterion for marking the distinction. However it does badly at meeting the second challenge. We have strong reasons to think that this account of the a priori is simply unrealistic in that it presupposes the existence of a dubious cognitive faculty *and* seems out of step with the realities of how mathematics—an a priori discipline if anything is—actually operates.

Seeming-Independence, I submit, does substantially better than the traditional model on this score.

Seeming-Independence has two key advantages over the traditional model. First, and most importantly, it does not presuppose any type of pure intuition that is capable of infallibly distinguishing true logical or mathematical propositions from false ones. Humans remain quite capable of arriving at *false* beliefs via pure, armchair reflection. Sometimes what seems to us to be mathematically obvious is not so. The point is simply that the falsity of such seemings undercuts their justification. The resulting beliefs will not count as being warranted a priori.

What makes direct a priori beliefs infallibly justified, on Seeming-Independence, is not that they are produced by an 'ultra-reliable' cognitive faculty. Rather, the infallibility derives from the structure of the relevant epistemic norms. Epistemic norms do not have to be transparent. It is quite possible that an epistemic norm is such that even those who put in a decent effort to comply with it might fail to do so. The norm governing direct, 'intuitive' a priori beliefs, on this account, is like that. It is structured such that its antecedent conditions pick out a non-transparent (but recognisable) property that is factive.

Second, and relatedly, Seeming-Independence allows that mathematical beliefs acquired purely via armchair reflection may be overturned by evidence acquired through testimony. Even if we grant Kitcher's claim that knowledge

acquired through testimony is knowledge that is dependent upon *experience*[3], the fact that our beliefs *can* be overturned in this way does not entail that mathematical beliefs in general lack infallible justification. It does not follow that if empirical evidence leads us to change our minds about mathematical claims, then we originally had fallible epistemic warrant for believing them. It is also possible that the claim was never warranted in the first place.

Given the distinction between subjective and objective epistemic norms, epistemic norms with transparent antecedent conditions and norms with non-transparent antecedent conditions, this is a viable option even in those cases where the overturned belief was held by a gifted mathematician.

Furthermore, as I have already stressed, Seeming-Independence is not committed to the claim that *all* a priori warrant is infallible. The infallibility is a result of the structure of the proposed epistemic norm governing propositions that strike us as a priori obvious, or self-evident. But a priori reflection is not restricted to self-evident propositions: for example, it might also involve inferring an unobvious proposition from a series of self-evident ones. Such beliefs will be answerable to a different set of norms: the norms governing the relevant type of inference.

In chapter two I suggested that generally the conclusions of deductive or putatively deductive a priori arguments have infallible warrant, if they have warrant at all. Yet, as argued above, some inferential mathematical reasoning may be *nondemonstrative*. And since nondemonstrative inferences are fallible inferences, the resulting belief will *not* be infallibly warranted.

Now the advantage of this thought is that it offers us further dialectical options with which to resist Kitcher's claim that mathematics is in fact an empirical discipline, rather than an a priori one. One explanation of how we revise our mathematical beliefs in light of findings elsewhere in the mathematical community, for example, is that we come to see that we were mistaken in taking some axiom to be self-evident, or in taking the conclusion of a supposed proof to follow logically from some axiom. These are cases in which our putatively a priori warrant is

---

[3] Burge (1993) famously rejects this view.

undercut: it is revealed not to be a priori warrant at all. But in other cases we might be in possession of genuinely a priori warrant that is *overruled*. This is a legitimate possibility for those cases where the mathematical belief in question was arrived at via a nondemonstrative inference.

For example, Euclid's parallel postulate was widely regarded as *less* obvious than his other axioms even before it was disproven. In the last chapter I suggested that one explanation of how this postulate came to be believed was that it was believed because it *seemed* true, if less obviously so than the other axioms. But another possible explanation might be that the belief was arrived at via an abductive or inductive inference. We could just as plausibly understand the sort of nondemonstrative mathematical reasoning Burge describes as applying to many of those mathematicians who accepted the parallel postulate. If someone took the best explanation of the nature of physical space to be the theorems of Euclidean Geometry, for example, and recognised the indispensability of the parallel postulate to those theorems, then she might well be inclined to abductively infer that the parallel postulate was true, regardless of whether or not it struck her as intuitively obvious.

Assuming that this abductive reasoning was good, she would have had fallible a priori justification for believing the parallel postulate. This fallible justification would be open to be overturned by any stronger evidence to the contrary, irrespective of whether that evidence was a priori or a posteriori.

So not every false mathematical belief, then, would need to be explained as an instance of someone failing to follow some epistemic norm. This gives us the flexibility to consider the epistemic status of false mathematical beliefs on a case by case basis.

This flexibility is a crucial dialectical advantage: it allows Seeming-Independence to be adaptable enough to cope with the sorts of features of mathematical belief formation that Kitcher mentions. The ruling that mathematics, as it is actually practiced, is *not* genuinely a priori would then require a great deal more motivation than has been provided.

3.3 Seeming-Independence and ultra-reliability

There is also a case to be made for the claim that Seeming-Independence does *even better* than the traditional Infalliblist view of the a priori at making the a priori/a posteriori distinction an overtly epistemological distinction. One worry for understanding the distinction in terms of infallibility— or 'ultra-reliability' as Kitcher prefers to put it—is that being ultra-reliable is not a clear hallmark of the a priori. There are clear cases of ultra-reliable a posteriori beliefs as well. Furthermore, and relatedly, the question of whether a belief is ultra-reliable is not *obviously* an epistemological question. This is because the ultra-reliability of a belief is a property that need not be explained by any epistemological features of the belief.

Consider a demonstrative judgement like 'that pen exists'. Now demonstratives take their semantic content directly from the objects that they pick out: so a thought *with that content* can only be considered in the first place if I successfully make reference to a pen. In situations where there is no pen in front of me, my judgement is not demonstrative, even though it might seem to me that it is. It has an altogether different content.

And we individuate beliefs not in terms of the content they *seem* to have, but in terms of the content they *do* have. So in possible worlds where there is no pen in front of me, I do not believe the same thing as I do in possible worlds where there is a pen in front of me. [4]

But a clear implication of this is that when I hold the demonstrative belief 'that pen exists', it must be true. The belief is ultra-reliable: there is no possible way to hold this kind of belief without there being a pen to pick out. What makes it possible for me to think that thought also makes the belief true. Yet it is also clearly a posteriori: I know that there is a pen there only because I can *see* that there is. So, what we have here is an ultra-reliable a posteriori warranted belief.

But on the traditional infalliblist account of the a priori, ultra-reliability is supposed to be the defining characteristic of the a priori. Infallibility was supposed to serve as the clearly epistemological criterion they could point to in support of the

---

[4] See Pyor 2006 for a discussion of these cases.

view that the a priori/a posteriori distinction is an epistemic one rather than a metaphysical or modal distinction. The presence of ultra-reliable a posteriori beliefs, then, suggests that this view does not cleanly capture the distinction between the a priori and the a posteriori.

A closely related point is that ultra-reliability is not an essentially epistemological property. That is to say, the fact that a particular belief cannot possibly be false does not, *in itself*, add to the epistemological status of the belief. Whether this infallibility makes a difference to the epistemological status of the belief depends on *why* the belief is ultra-reliable.

Suppose that I am aware that I have been slipped a drug that makes me hallucinate extremely realistic-looking pens. This is a context, I take it, in which norm (4)[5] would not permit me to rely purely on how it seems to me in forming beliefs about pens: this would be a situation where the *ceteris paribus* clause of norm (4) comes into play. Coming to believe that there is a pen in front of me just because it perceptually seems to me that there is *in those circumstances where I have reason to mistrust my perceptual experiences* is clearly and even paradigmatically irrational: this cannot be construed as a belief that follows a legitimate epistemic norm.[6]

But suppose that I ignore this and maintain my (misplaced) faith in my perceptual faculties. Suppose further that I do manage to make reference to the single genuine pen in my environment and form the demonstrative belief 'that pen exists'. This belief is ultra-reliable. But that has no relevance to the question of whether the belief is *rational*. Our epistemic assessment of the belief is not, in this context, altered by considerations of its ultra-reliability. My belief is entirely unwarranted, even though it is ultra-reliable. In this case, then, questions of whether I am warranted in holding my demonstrative belief come apart from questions of its status as an ultra-reliable belief. Ultra-reliability in and of itself is not an indicator of epistemic standing.

---

[5] (4) If it perceptually seems to you that *p* then all things being equal you may believe that *p*.
[6] I will explain what this particular sort of epistemic irrationality consists in, in section 6.

Now Seeming-Independence has an advantage in that it holds that direct a priori beliefs are ultra-reliable in virtue of their having been formed in a manner that is in compliance with a special kind of epistemic norm. So we might say that such beliefs are ultra-reliable *because* they are warranted.

While Seeming-Independence does not hold that *every* a priori belief is ultra-reliable, it does make ultra-reliability a product of the belief's epistemic status.

To rephrase the point, there is something importantly different about *why* this kind of demonstrative thought is guaranteed to be true, and why a non-inferential a priori warranted judgement is guaranteed to be true. With direct a priori beliefs, this ultra-reliability applies only to those beliefs that are in full compliance with the relevant epistemic norms; that is, to those beliefs that are fully and genuinely rational. The infallibility, then, comes from the *way* in which the belief was arrived at.

Consider some logical belief I hold, *l*. Let us suppose *l* is a true logical principle that strikes professional logicians as plainly self-evident, but which is beyond the intuitions of some less reflective non-philosophers. Now in this case, professional logicians will believe *l* because it is self-evident. An unsophisticated layman, however, might come to believe *l* only on the basis of the testimony of an expert logician.

Now it is important to see that, on Seeming-Independence, whether the belief is infallibly warranted depends on whether one is following norm (10) in believing *l*. There is nothing wrong with relying on expert testimony when forming logical beliefs. This is most commonly how laymen acquire information about technical subjects. For instance, all or almost all of my scientific beliefs are formed like this. There is no reason to think it inappropriate for logic.

There might even be good reason to prefer that method to the normal, direct method. If I have reasons to doubt that I am good at telling self-evident logical propositions from plausible-looking false ones, then relying on an expert in forming logical beliefs would be *better* than not doing so.

But doing this can bring different epistemic norms into play. By employing different belief-forming methodologies, we can actually make our beliefs answerable to different epistemic norms. If norm (10) has nothing to do with either why I formed the belief, nor why I continue to believe it, then the warrant for that belief must come from another source. The self-evidence of a proposition can only justify my belief in those cases where I *follow* norm (10). There is nothing preventing that warrant coming from my trusting the word of an expert. Seeming-Independence does not claim that *all* logically self-evident propositions are warranted a priori. It is possible that someone might hold that belief for reasons entirely unrelated to its self-evidence. Depending on what those reasons are, this might make the logical belief, unusually, a posteriori. On Seeming-Independence, whether a belief is a priori depends exclusively on the norms the believer is complying with. Norm (10) offers us an infallible route by which to come to believe *l*, but there might also be other acceptable routes that are not infallible.

The point is that on Seeming-Independence, the infallibility of direct a priori beliefs is to be explained not by the *content* of the beliefs themselves, but by the way we go about acquiring and maintaining that belief. What this type of a priori reasoning does, this suggests, is provide us with a special route to knowledge—a route that, if followed correctly, guarantees truth. The fact remains that a posteriori *warrant* cannot itself provide this sort of guarantee, even if sometimes the content of certain a posteriori beliefs can provide it.

This, I think, suffices to retain a significant epistemic difference between the a priori and the a posteriori. According to Seeming-Independence, the type of warrant that we typically have for those logical, conceptual or mathematical beliefs that strike us as intuitively obvious, is different in kind from the warrant we have for those similarly immediate perceptual beliefs. Propositions like '1+1=2', or 'nothing is both red and green all over at the same time' are typically governed by norm (10): that is, we believe these propositions *because* they are self-evident. [7]

---

[7] Though, as we have seen, they are not *always* governed by this norm. Conceptual truths don't *have* to be a priori: there are other possible ways of acquiring beliefs about them.

This supplies them with a kind of infallible epistemic warrant. Since self-evidence is factive, anybody who follows norm (10) correctly will arrive at true belief. Epistemic warrant and truth are inseparable in this case: the former ensures the presence of the latter. Typically, beliefs like 'there is a pen on my desk' are governed by norm (4): it is partly *because* it perceptually seems to me that there is a pen on my desk that I believe there is. Since how it seems to me is not factive, (4) supplies merely fallible warrant: it cannot absolutely guarantee that there is indeed a pen on my desk.

3.4   Problems of length and width

Seeming-Independence, I have argued, enjoys a crucial dialectical advantage over the traditional infallibilist view. It is better able to explain how a priori warrant and knowledge is achievable for creatures with our cognitive limitations and epistemic practices. Yet it also retains a commitment to a highly restricted version of the Infallibility Thesis: that is, it maintains that non-inferential a priori beliefs are *infallibly* warranted if warranted at all. Arguably, then, it retains and even improves on the key strength of the traditional view—that it makes the a priori/a posteriori distinction an overtly epistemological one—while avoiding its costs.

However, just because this account uses some sort of epistemological criteria in dividing a priori and the a posteriori, it does not follow that the distinction is a clear, natural one. Just because there is an epistemological difference of sorts, it does not entail that the distinction is not ad hoc.

Hawthorne (2007) has recently argued that the supposed distinction between the a priori and the a posteriori is simply not a natural one: it does not divide up our 'epistemological lives' in a natural way. In pressing this point, he targets two ways of construing Kant's claim that a priori knowledge is independent of experience. The first construal understands a priori knowledge to be knowledge that is independent of *the believer's environment*. The second construal understands a priori knowledge to be knowledge that is 'sustained by a method that is not experience-involving' (Hawthorne 2007, p. 7). Hawthorne argues that

neither construal is a natural one. On both ways of marking the divide, the distinction does cleanly separate the a priori from the a posteriori. On natural readings of both construals, paradigmatically a priori beliefs end up on the wrong side of the divide. There is no way to preserve the distinction, Hawthorne concludes, without making it unacceptably ad hoc.

While I have thus far focussed primarily on a priori warrant, rather than on a priori knowledge, Seeming-Independence does indeed appear to be committed to a version of this second construal of a priori knowledge. However, in the next two sections, I will argue that it can avoid Hawthorne's objections.

While I will focus primarily on Hawthorne's criticisms of the second construal ('Experience-Independence') I will first briefly sketch Hawthorne's problems with the 'Environment-Independence' construal of a priori knowledge.

Hawthorne takes the idea behind Environment-Dependence to be this. For a belief to count as knowledge, in his view, it must be sufficiently safe from error. Generally, in order for a belief to be sufficiently safe the believer's environment must be configured in such a way that it provides a 'safe haven' (Hawthorne 2007, p.3). My perceptual belief that I have just driven past a barn, for instance, is only safe if I am not in an area filled with realistic barn facades. In fake barn country, that belief, even if true, is not safe enough to count as knowledge. The claim that a priori knowledge is Environment-Dependent, in his view, is the claim that a priori belief 'stands in no special need of a safe haven from the environment since the environment can't but provide a safe haven' (Hawthorne 2007, p. 3).

Now, as Hawthorne points, out, this claim is simply not true. Just as perceptual knowledge requires the thinker's environment to be free of 'Bad experiences' so too does a priori knowledge. For instance, in the same way that it is possible for one's environment to contain an abundance of fake barns, it is also possible for one's environment to contain something like an a priori gas; a gas that 'induces the phenomenology of blatant obviousness' (Hawthorne 2007, p. 4) whenever the believer considers manifestly false propositions. The presence of such a gas in one's environment would render any 'obvious' a priori belief unsafe in the same way that the presence of fake barns renders barn-beliefs unsafe. Just as

99

perceptual knowledge depends in part on facts about our environment, so too, in Hawthorne's view, does a priori knowledge.

I take Hawthorne to be correct in that *the safety* of an a priori belief does depend on the believer's environment in a variety of respects. Just as environmental conditions can affect the veridicality of our perceptual beliefs, so too can they affect the veridicality of armchair reflection.[8] Armchair reflection is only reliably veridical when we are thinking clearly: and there is no good reason to suppose that environmental factors cannot affect how clearly a person reasons. If our environment was such that the possibility of reason-affecting conditions obtaining was a real one, then our judgements arrived at via non-empirical armchair reflection would lack safety. So if we think that safety *really is* a necessary condition for a priori knowledge, then we must concede that Environment-Independence rules out far too many beliefs to be credible.

Before turning to Experience-Independence, it is worth pointing out that the safety requirement is at its most plausible as a requirement for knowledge, and not as a requirement for warrant or justification. Consider once again the fake barn case. Plausibly, we have seen, I would not *know* that I had just driven past a barn if I happened to be driving through an area that for some reason had a high prevalence of fake barns. But would I be justified[9] in believing that I had just driven past a barn? On this question, intuitions diverge. Some reliablists about justification might think that since our perceptual faculties are not reliably veridical barn-detectors *in those surroundings,* the relevant belief is not the product of a reliable process and is thus unwarranted. Philosophers with stronger internalist intuitions, on the other hand, might think that the presence of the fake barns *by themselves* does not undermine my justification. Provided I had no reason to suspect that I was in fake barn country, the belief in question remains justified. Anyone attracted to this way of understanding the picture might then be inclined to understand fake barn

---

[8] I take the term 'armchair reflection' to apply to those beliefs acquired non-emipirically, through pure reflection. Note that beliefs arrived at via armchair reflection are *not* necessarily warranted a priori. Armchair reflection might generate miscalculations, or other rational errors. The resulting beliefs would not count as warranted a priori.

[9] I use justification and warrant interchangeably here. The literature on these cases typically refers to justification rather than warrant, so I am introducing talk of justification only to make the connections between the literature and what I say here clearer.

scenarios as being like the Gettier examples: cases of justified true belief that are not knowledge.

The same holds for the a priori gas scenario. It is not obvious that the mere presence of an a priori gas in the surrounding environment undercuts the epistemic warrant of an obvious a priori judgement made by someone *unaffected* by the gas. Whether philosophers take these cases to be warranted or unwarranted depends on the particular brand of a priori warrant endorsed. And on my account, this sort of case would not be a case of unwarranted a priori belief.

Let me explain. Whether a belief is warranted depends, in my view, only on whether the believer has followed the relevant epistemic norm or norms. The presence of an a priori gas in her surroundings does not entail that she has failed to comply with those norms. The way the a priori affects belief, I take it, is that it disrupts our reasoning in such a way as to render us *incapable* of following certain epistemic norms. So suppose while in this environment a person comes to consider whether it is possible for something to be both red and green all over at the same time. If she is lucky enough not to be affected by the gas, she will decide, rightly, that this is not possible. There is nothing problematic about how she went about forming this belief: she recognised that it is self-evident that an object can be entirely one colour *and* entirely another colour simultaneously, and formed her belief accordingly.

If she *was* affected by the gas, however, she would be temporarily incapable of telling self-evident propositions from non-self-evident ones: false propositions rather than true ones strike her as 'blatantly obvious'. So if she had been under the influence of this gas while considering this question, she would have believed it blatantly obvious that something could indeed be both red and green all over at the same time. What the gas has done is make her temporarily unable to follow norm (10) by making her unable to identify its antecedent conditions.

The presence of a priori gas in the believer's vicinity makes it *lucky* that she is able to comply with norm (10) in those cases where she is not affected by it. If we hold that the involvement of this sort of luck is inimical to the belief counting as knowledge then we will agree with Hawthorne that these are not cases of

101

knowledge. Yet there is no reason to suppose this sort of luck is inimical to her belief counting as *warranted*.

Bear in mind that whether or not one is properly warranted in believing that *p* depends on whether one has successfully *followed* the relevant epistemic norms in arriving at *p*. While I do not offer an account of rule-following here, any plausible account of rule following must hold that following an epistemic norm—however it characterises this—is not the same as simply believing in a way that is consistent with a norm. To be following an epistemic norm, it must be the case that the norm in some specific way *explains* the belief.

So there must be a non-accidental, luck-free connection of some sort between the norm and the belief, otherwise the belief is not properly warranted. So some types of epistemic luck certainly *are* inimical to a belief counting as warranted. But the type of epistemic luck that a priori gas introduces does not appear to be so. The presence of a priori gas does not entail that an unaffected person cannot properly follow epistemic norms. As such, we have no reason to doubt her belief is properly warranted.

So, if we agree with Hawthorne that the sort of luck that a priori gas brings into play *really is* inimical to a belief counting as knowledge, then our only recourse is to take this case to be a Gettier-like case: a situation where a justified true belief is not knowledge.[10]

I take this conclusion to be not unacceptable: it is well established that it is possible to have a justified true belief that is not knowledge. Moreover, the a priori gas case closely resembles Gettier's famous examples of such cases. I take it that the reason we do not take Gettier's examples to be cases of knowledge is precisely because the believer is *lucky* that his or her belief turned out to be true. Smith, the

---

[10] I will not pursue this thought here, but it does not strike me as *undeniable* that this sort of luck is knowledge-defeating, given the level at which the epistemic luck takes place. This is not, as it is in most cases of epistemic luck, a situation where she is lucky *that her belief-forming method arrives at truth*. It isn't a situation in which the believer follows a line of reasoning that could have easily have led her astray. I agree that those cases are clearly cases where the believer lacks knowledge. But in this scenario, there is no luck whatsoever about the fact that her reasoning method was truth-conducive. What is lucky is just that she was epistemically competent enough to follow that method. But it isn't obvious to me that competency-luck is epistemically relevant. And since cases where one is *luckily competent* are still cases where the belief is not safe, I do not consider Hawthorne's safety requirement to be undeniable either. I accept it here simply for the sake of argument.

job applicant who believes that the person who will get the job has ten coins in his pocket, is lucky that this belief turned out true, given that it was based on the reasonable but false assumption that his competitor for the post will get the job. It is this luck that defeats his knowledge. In the a priori gas case, as I have stressed, it is also the presence of luck that prevents the belief from counting as knowledge (assuming it does). The case mirrors Gettier's in this respect.

With that in mind, let us now consider Experience-Independence. So on this conception, a priori knowledge is understood to be knowledge sustained by a non-experience-involving method. Now Hawthorne's concern with this construal is that whether a cognitive process counts as experience-involving or not depends on how we *choose* to individuate it. And, in many cases, there are no obvious facts of the matter that make one choice better than any other. This makes the distinction unnatural. In this picture, whether an item of knowledge falls into one epistemological category or another depends on nothing more than how we individuate the process, and there are no obviously 'correct' ways to individuate it. As such it is hard to see why it should *matter* which category it falls into.

Meeting Hawthorne's challenge here requires addressing two distinct problems: problems of *width*, and problems of *length.* Problems of width stem from concerns similar to those we have already come across: that the presence of empirical evidence can sometimes prevent a particular a priori belief counting as *knowledge.* As Hawthorne says:

> Even if I have carefully worked through a mathematical proof that *p*, I will not know that *p* if I get empirical evidence that I am mad, or that human and mechanized experts agree that not-*p*, or that there is a priori gas in the area, or that I have made lots of mistakes using a very similar proof technique in the past or that lots of smart people are inclined to laugh when they hear my proof. (Hawthorne 2007, p. 8)

Given the presence of these Bad experiences in or around the process that sustains the belief that *p*, that belief does not count as knowledge. Now one possible way to interpret this thought is to say that the process that produced *p* is knowledge-generating if it includes only *Good experiences*. Hawthorne states his case as follows:

> That my proof counts as knowledge appears to depend crucially on it being accompanied by Good experiences. But if the process of arriving at putatively a priori knowledge is individuated so as to include Good experiences, then it will count as a posteriori by the experience dependent criterion … There is a general problem of width here. All sorts of facts at one time have some causal bearing on belief at a slightly later time. Which of them are to count as part of the belief forming process? (Hawthorne, 2007, p. 9)

In addition to the problem of width, there is also a problem relating to where we take the relevant belief-forming process to have begun.

Consider a scenario where a person learns a set of natural laws as a child. Much later in life, she recalls these laws to mind, and from them infers that some more specific conditional obtains: say, 'if I drop a penny, it will fall downwards'. If we take the cognitive process to begin with her learning the laws, then, according to Experience-Independence, her knowledge of the conditional is a posteriori. However, if we take the relevant cognitive process to begin from the point where she brings these laws to mind, then it shall count as a priori. Hawthorne sets up the problem here as follows:

> It is important to realize here that when one considers a stream of events, various processes can be distinguished, some longer in temporal extent than others … Suppose someone extracts a conditional prediction about the course of events. There is a process that begins with the teacher telling him the laws and ends with applying some laws to derive a conditional prediction. But there is a

shorter process that begins with retrieving the laws from the

relevant internal information bank and ends with producing the

conditional prediction … Which shall we use to test whether the

belief is a priori? … Is there any deep mistake in taking [the shorter

process] to be the relevant safe method? (Hawthorne 2007, p. 10)

Once again, what is required is a non-ad hoc way of determining which way of individuating a cognitive process is the appropriate one.

The problem here is exacerbated by the fact that to prevent the belief inferred from natural laws from counting as a priori, we need to take the relevant process to be the longer process. Yet to avoid an unacceptable result in other cases, we would need to take the relevant process to be a shorter process. For instance, Hawthorne notes that all our initial mathematical beliefs are epistemically based on the testimony of our teachers. If we were to take the cognitive process involved in any mathematical calculation to be a long process, then it would start to look like all our mathematical beliefs are a posteriori. This would be an unacceptable conclusion. To avoid it, we must take the relevant cognitive process in this case to be the shorter one. But is there any principled reason that the process in the natural laws example is long, and the process in the mathematical calculation case is short? The decision to take one to be short and the other long looks to be ad hoc.

In the remainder of this chapter I will argue that these concerns are only pressing for those that endorse the sort of account of knowledge that Hawthorne does. On the account of a priori warrant and knowledge I favour, it can indeed be a 'deep mistake' to take the relevant process to be too short or too wide.

3.5 Cognitive processes and epistemic norms

On a normative approach to knowledge and epistemic warrant, a belief is warranted when the believer follows the relevant epistemic norms in forming and maintaining her belief. A belief is knowledge when the belief that is formed and maintained in this way is also (lucklessly) true. So when deliberating about whether

some belief counts as knowledge, we must consider whether the believer has indeed been following the relevant epistemic norms.

Now every belief is the result of some cognitive process or another. And any cognitive process can be individuated in a variety of ways. But when we are considering whether the belief is knowledge or not, what we are considering is whether the cognitive process that produced the belief could be *legitimately* individuated as an instance of the believer following an epistemic norm. If it can be so individuated, then it is warranted. If it is also lucklessly true, then it will count as knowledge. If it cannot accurately be so individuated, then the belief is not warranted and is not known.

Now whether a belief is known a priori on the suggested construal depends primarily on whether, as Burge put it, it is 'underwritten by an a priori justification or entitlement that needs no further justification or entitlement to make it knowledge' (Burge 1993, p. 461). And whether it is underwritten in this way will depend on *which* epistemic norms were followed. If those norms are structured such that their antecedent conditions appeal to non-factive 'seemings' then the resulting belief is a posteriori. If the relevant norms are not so structured then the belief is warranted a priori. If it is also known, then it is known a priori.

In settling whether a belief is a priori or not then, the *right* way to individuate the cognitive process that resulted in the belief is clear. It must be individuated as an instance (or a series of instances) of epistemic norm-following. No other way of individuating the cognitive process is relevant. And this allows us to settle conclusively how long or how wide the relevant cognitive process should be taken to be: the relevant norms *determine* what the proper length and width of the relevant cognitive process is.

Consider the following norms:

(2)   Do not believe that *p* if some alternative proposition incompatible with *p* has a higher degree of support

(6) If you are justified in believing that *p,* and are justified in believing that if *p* then *q*, then believe *q* or give up one of the other beliefs

(10) If it is self-evident that *p* then you may believe that *p* without considering any further evidence for it

Now the cognitive process individuated as instances of following each of these norms will have very different lengths and widths from each other. Suppose a believer comes to believe that one plus two equals three in the usual sort of way. Once again we can individuate the process as leading up to this belief in a variety of ways. We could consider it as part of a process that began when she was first introduced to numerical concepts. Or we could think of it as part of a holistic process that stretches all the way out to her background beliefs about her own sanity and what mathematical experts think. But what makes this belief warranted is that the believer arrives at this belief as a result of following norm (10). This means that the *relevant* way of individuating the process is to take it to be both very short and very narrow: it begins with her considering the proposition, taking it to be blatantly obvious, and concluding that it is true.

However following norm (2) involves a very different methodology. Consider a concrete case of a person's following norm (2). Suppose a person who believes that *p,* but then suddenly discovers that another belief she holds, *q*, is inconsistent with *p.* After careful deliberation, she decides that overall the evidence in favour of *q* is substantially stronger than the evidence in favour of *p*, so she stops believing that *p* and maintains her belief that *q*. Now what we have described here is what Hawthorne would call a wide cognitive process: the believer has brought to mind and weighed up a range of different considerations in determining whether to believe that *p*.

Following norm (6) is different from both of these. Suppose a person comes to believe that *p* after being told that it is true by a highly reliable source. Some time later she discovers that the conditional 'If *p* then *q*' obtains, and immediately forms the belief that *q*. Now once again, we could take this process to have begun when she discovered that the conditional obtains, and leapt to the conclusion that

q. But that cognitive act in isolation is *not* an instance of following norm (6). The question of whether the believer is justified in believing that *p* is relevant to the question of whether or not she has followed (6). If a third party were ask to consider, for example, whether this person *knows* that *q*, he would need to determine whether or not the belief that *p* was justified. In other words, he would need to view her belief-forming process as a long one; as having begun when she acquired the belief that *p* and ended much later after she arrived at *q.*

The point is that different epistemic norms have different demands. Meeting these demands sometimes requires us to consider a wide range of propositions, and other times it requires us to rely upon evidence we received in the past. How long or wide we should take a cognitive process to be can be definitely determined by considering what epistemic norms are at play in any given example.

So it is not the case that we have no principled means available of determining the appropriate length and width of the cognitive process that produces some belief. If we can determine exactly which epistemic norms would apply to the belief in question, then we can determine how long or wide the cognitive process should be taken to be.

This highlights a crucial advantage that a normative approach enjoys over Hawthorne's preferred safety account of knowledge. The issue with Hawthorne's view is that the *only* restriction on a cognitive process counting as knowledge-producing is that it must be a cognitive process that produces safe beliefs. So all that is required is that it be a process that is reliably veridical in the context in which the believer formed her belief. But, since this is the only requirement, it is difficult to motivate one characterisation of a cognitive process over another. But since a normative approach is explicitly only focussed on whether the belief in question was formed as a result of following an epistemic norm, the question is significantly easier to cope with.

3.6 Following norms and the transmission of epistemic warrant

However, even if Hawthorne's examples fail to establish that there is *no* way to motivate construing a cognitive process to be of some specific length and width rather than some other length or width, the cases he draws attention to might yet serve as counter-examples to Seeming-Independence. It might be that one could argue that these examples are cases where one's epistemic warrant is *not* properly independent of how it seems to you, given that 'Good experiences' do possibly play some sort of significant epistemic role. The suggestion here is that what these examples might also call attention to is the fact that Seeming-Independence has not yet been fleshed out enough to allow us to determine whether it really does capture a plausible conception of the a priori. Does this conception of the a priori/a posteriori distinction get the right results in the more complicated cases that Hawthorne brings up?

I take it that Hawthorne's suggestion that inferences from long-remembered natural laws might count as a priori is easily settled. The above discussion gives us a quick and clear rebuttal to this thought. Given that all inferences function by passing on warrant from premises to conclusion, the norm involved in this sort of case will be like norm (6), in the sense that the norm will have the general form of 'If you are *justified* in believing that *x*, then φ'. Since the manner in which the belief was acquired is relevant to whether it is justified, considering whether any belief is an instance of following a norm like this would require us to construe the cognitive process involved as an extremely long process, beginning when the natural laws were initially learned. On Seeming-Independence, then, we can definitely rule that this case is not a case of a priori knowledge.

A more problematic case for Seeming-Independence is case where one has empirical evidence that indicates that a putatively a priori item of knowledge is not known: that is, where one has evidence that indicates that there is a priori gas in the area, or that one is mad, or that the belief in question is generally derided by mathematical experts, or something to that effect.

The first point to note is that a case like this is importantly different to the case where a priori gas was *merely present* in one's environment. Above I argued

that the presence of a priori gas would not undermine the belief's epistemic warrant. However, *evidence* that there was a priori gas in one's environment would indeed be warrant-defeating. I take it to be untenable to deny that anybody accepting at face value what *seemed* self-evident to them, in cases where they knew they were vulnerable to something like a priori gas, would not be warranted. But how does Seeming-Independence account for this?

The worry here is that the person blithely accepting what seems to her to be self-evident in cases where all the evidence suggests she is vulnerable to a priori gas is clearly in violation of an epistemic norm. It is undeniable that she *ought* not to ignore this evidence. But the evidence in question is empirical: it is thus ultimately seeming-*dependent.*

So—arguably—the natural explanation for this is to suppose that self-evident a priori beliefs are answerable to an epistemic norm that ultimately appeals to 'seemings' in its antecedent conditions. In addition to norm (10), perhaps there is a set of seeming-dependent norms at play with regard to these beliefs.

Yet Seeming-Independence is not *committed* to positing the existence of a further epistemic norm to explain why such cases are not warranted. Indeed, we would have excellent reasons to suppose that the person who accepts what strikes them as obvious in cases where she has reasons to suppose that she is vulnerable to a priori gas, is *not following norm (10).* As a result there is no need to posit a further, seeming-dependent norm to explain why such cases are not warranted.

Recall that in order for a believer to be following a norm, the belief has to be in some way explained by the norm itself. And for a belief to be explained by norm (10), I take it, it must be the case that the belief is held *because* it is self-evident. That the belief has, or appears to have, the property of self-evidence must be salient in the causal explanation of how the believer arrived at the belief in question. If this were not to be the case, then it would be difficult to make sense of how the person in question is genuinely following norm (10).

But if a person believes that *p* in the face of evidence that indicates that *p* might very easily not be self-evident after all, it does suggest that it is not the self-evidence of the proposition in question that explains her believing it. If self-

evidence was causally salient we would expect evidence that casts into doubt whether a proposition is self-evident to have some sort of effect on her believing it. That it does not, suggests the believer is not genuinely following norm (10) after all.

The point of this is that, on Seeming-Independence, there are two sets of considerations that need to be taken into account when assessing whether a belief is properly warranted. Set One pertains to whether the belief in question is *consistent with* the relevant epistemic norms. If we are considering whether some belief, *p*, is consistent with the norm 'If *x* then believe that *p*', what we would be concerned with here is whether *x* obtains. In other words, Set One contains just those considerations relevant to *whether or not the antecedent conditions of the norm obtain*. If so, then the belief that *p* is consistent with the norm 'If *x* then believe that *p*'.

Set Two pertains to whether the believer has actually *followed* that norm in forming and maintaining the belief. Now I take it that the belief being consistent with a norm is necessary for a believer to have followed it. But mere consistency, as I have stressed, is not sufficient for rule-following. In determining whether a believer has followed some norm, we would also need to establish that the norm explains the belief in the right sort of way. That is to say, we would need to see whether it was *because x* obtained that the subject believes that *p.* Considerations relevant to this more complicated explanatory question are the considerations that I take to belong in Set Two.

Now the considerations that belong in these two sets play distinct epistemological roles. The sorts of considerations that belong in Set One play what we might call a *warrant-generating* role. The considerations that belong in Set Two play what we might call a *warrant-transmitting* role.

Let me explain this distinction by means of an example. Suppose we are considering whether I have followed norm (10) in believing that nothing can be both red and green all over at the same time. Now, what would make my belief warranted would just be the fact that the proposition is self-evident. It is this feature that makes it rational for me to believe that particular proposition. In other words, the epistemic warrant for that belief *comes from* its self-evidence.

However, in order for the self-evidence of that proposition to justify my belief that it is true, there must be the right sort of connection between it being self-evident, and my believing it. If I didn't think it was blatantly obvious that one object cannot be entirely one colour and entirely another at the same time, but rather held this belief on the basis of some deviant theory about the nature of redness and greenness in particular, then the fact that the proposition *is* self-evident would not be able to justify my believing it.

So in order for the proposition's self-evidence to justify my believing it, there must be some sort of causal chain between its self-evidence and my belief. To speak metaphorically, the links in this causal chain would serve as warrant transmitters: they allow the epistemic warrant to pass from the self-evidence of the proposition, to my belief that the proposition is true.

Now the point here is that Seeming-Independence, as I set it up in section one, is just the view that none of the epistemic norms that apply to an a priori belief have seemings *in their antecedent conditions.* So essentially, all this view is committed to is the claim that perceptual and intellectual seemings are not what *makes* the belief justified, in the way that perceptual seemings are part of what *makes* perceptual beliefs justified. Seemings are not warrant-generators for a priori beliefs. However, the theory remains neutral on whether or not seemings might sometimes be warrant-transmitters.

And it seems plausible to think that sometimes seemings— both intellectual and perceptual—might be warrant-transmitters. For instance, if it did not *seem* blatantly obvious to me that the proposition 'nothing can be both red and green all over at the same time' was true, it would be difficult to imagine how the self-evidence of that proposition could explain my believing it. That a proposition strikes me as self-evident looks like it would be relevant to the question of whether it is the self-evidence of the proposition that explains my belief.

An intellectual seeming, recall, is just a kind of intellectual attraction. If I *don't* experience that attraction towards a proposition, if it does not strike me as blatantly obvious, then it is doubtful whether I have *recognised* that the proposition genuinely is self-evident. And that I have not recognised its self-evidence, casts

doubt on the hypothesis that the self-evidence of the proposition explains my belief.

Let me illustrate this point by means of an analogy. Burge (1993) argues, against Chisholm, that the fact that memory is involved in any long mathematical or logical demonstration does not prevent the conclusion of the demonstration from counting as a priori. He says this:

> Chisholm's conception of the role of memory in demonstrative reasoning seems to me to be off the mark. If memory supplied, as part of the demonstration, "contingent propositions about what we happen to remember", the demonstration would not be purely logical or mathematical. But the normal role of memory in demonstrative reasoning is, I think, different. Memory does not supply for the demonstration propositions about memory, the reasoner, or past events. *It supplies the propositions that serve as links in the demonstration itself*. Or, rather, it preserves them, together with their judgemental force, and makes them available for use at later times. (Burge 1993, p. 462, my emphasis)

Burge thinks that what actually justifies the conclusion of a long demonstrative argument is its premises alone. Memory is involved, but it serves merely to *make it possible* for those premises to justify the conclusion.

Seeming-Independence is open to the possibility that intellectual or even perceptual seemings play *that* sort of role in a priori belief-formation. Just as Burge thinks you need memory in order for the premises of a long demonstrative argument to be able to justify its conclusion, you might sometimes need an intellectual seeming in order for the self-evidence of a proposition to justify your belief in it. Seemings, like, memory, might sometimes play a warrant-transmitting role. All Seeming-Independence claims is that, for a priori beliefs, seemings *do not*

play the sort of role that Burge takes the premises of the demonstration to play in that example. [11]

So the fact that Hawthorne's so-called 'Good experiences' —all those experiences that are *not* experiences indicating that there is a priori gas around, or that you are mad, or that experts think your belief is laughable, etc.— *do* seem to be necessary for an a priori belief to be warranted does not undermine Seeming-Independence. Good experiences can very plausibly be understood to be playing a warrant-transmitting role. That is, we might plausibly take the absence of Good experiences (or the presence of Bad experiences) to indicate that an a priori belief is not held *because* its antecedent conditions obtained.

3.7 Conclusion

Seeming-independence, then, has some important advantages. A cursory glance suggests that it divides the epistemological terrain such that there is a natural, epistemologically interesting difference between the a priori and the a posteriori, without raising the suspicion that beliefs with a priori warrant are entirely beyond creatures with our cognitive limitations and epistemic practices.

There remains, however, some work to be done in fleshing Seeming-Independence out. A fundamental assumption on which Seeming-Independence depends is the assumption that we can understand the epistemic norm governing

---

[11] And a very clear example of seemings playing the same role that memory plays in Burge's example would be those familiar cases where you perform a long mathematical calculation by writing the steps down on paper. Intuitively, this is clearly a case of a priori reasoning: I take it that it would be disastrous if this everyday type of mathematical belief formation was deemed a posteriori. However, as Hawthorne points out (Hawthorne 2007, p. 7) the a priori status of such beliefs is tricky to explain, given that there does seem to be a sort of dependence on perceptual experiences involved in retrieving the information from the paper. Now I take it that in this case, the role that writing on and perceiving the paper plays is precisely the same as the role that memory plays in Burge's example. We are, after all, only writing down the steps in the calculation as we go because we cannot easily *remember* all those steps. We are using the paper as a memory substitute. So, if Burge is right about the role that memory plays in long calculations, there is no reason that our perceptual experience of the paper we have written on could not play the precise same role. Plausibly, we write down those steps as a way to preserve them for later use: seeing what we wrote on that paper does not provide us with *new* information. It just allows us to continue to access what we already worked out. Our ability to write and read what we have written, arguably, merely *allows* the previous steps in our calculation to justify the conclusion, in just the way that memory sometimes does.

direct, 'obvious' a priori judgements to be governed by norm (10): 'if p is self-evident, then you may believe that p without considering any further evidence for it'. The plausibility of this suggestion depends, in turn, on the whether we can make sense of the notion of *self-evidence* that it appeals to.

In the remaining chapters I undertake to unpack and defend this notion.

4: Self-Evidence and Analyticity

I have thus far provided some reasons to think that Seeming-Independence is a plausible and theoretically fruitful account of the a priori. Seeming-Independence has the advantage of using a purely epistemological criterion in dividing the a priori from the a posteriori: whether or not a belief is based on a non-factive *seeming* or not has clear epistemological implications. Seeming-Independence also has the advantage of not wedding itself to an implausible philosophy of mind. There is no need to posit a separate faculty of a priori intuition, nor to account for how intellectual seemings are as reliable as—if not more reliable than—perceptual seemings. Seeming-Independence is perfectly consistent with the thought that humans are limited and fallible reasoners. Seeming-Independence also is well equipped to cope with two common charges against the a priori: the charge that a priori warrant is a property that either no or very few beliefs possess, and the charge that the distinction is either incoherent or ad hoc.

However, in arguing for Seeming-Independence, I have appealed to a property that is problematic in its own right: the property of *being self-evident.* This account presupposes that self-evidence is indeed a property that some propositions have. And this presupposition, I take it, would be illegitimate without any explanation of what self-evidence is. Boghossian nicely sums up the challenge that is to be addressed:

> Here the problem is that no one seems to me to have shown how this notion is to be spelled out. In particular, no one has supplied a criterion for distinguishing those propositions that *are* self-evident from those that … merely *seemed* self-evident to many people for a very long time. (Boghossian 2001, p. 9)

This chapter will go some of the way towards meeting Boghossian's challenge. I take there to be two types of self-evident beliefs.[1] There are beliefs about analytic propositions, and there *introspective* beliefs: beliefs about our occurrent propositional attitudes and sensations. In this chapter I will concern myself only with analyticity, leaving discussions of self-knowledge for the next chapter.

My aim here is not to offer a thorough defence of analyticity: such a task would take me too far off course. My intention is the more modest aim of offering some reasons to think that there *is* a plausible account of analyticity to be had. What I will suggest is that the account of normativity that I have put forward offers us a way of defending an epistemic conception of analyticity from what I take to be its most important challenge: the counter-examples Williamson offers in *The Philosophy of Philosophy* (2007).

4.1 Metaphysical and epistemic analyticity

Analyticity has been enjoying something of a resurgence. While it was once widely considered to have been decisively undermined by Quine (Quine, 1951), Boghossian, in a series of influential papers[2], has introduced an understanding of analyticity that he thinks avoids Quine's objections. The crucial move is that Boghossian distinguishes metaphysical analyticity from epistemic analyticity. The thought here is that the problems Quine pointed out only apply to analyticity understood as a *metaphysical* thesis. Reimagined as a purely epistemological thesis, analyticity is free from the problems that have traditionally plagued it.

Now according to metaphysical analyticity, an analytic proposition like 'all bachelors are unmarried men' is *made true* by the meaning of the terms involved alone. This claim is difficult to reconcile with a plausible semantics. The issue here is this account seems to be committed to the implausible thesis that analytic propositions are made true by an entirely different class of considerations from the considerations that determine that a synthetic proposition is true. And it is difficult

---

[1] Though I am not in principle opposed to the idea that there may be more.
[2] See Boghossian 1996, 2001, 2003a, 2003c.

to see how this could be so. Specifically, it is difficult to see how something like the *meaning* of a proposition could by itself make a proposition true.

The trouble here is that the meaning of a proposition seems to establish only what the proposition represents to be case. What makes a proposition true or false, however, are not facts about what it represents to be the case. Roughly, what makes a proposition true or false are the facts about *whether* what it represents to be the case, actually is the case. What makes propositions true are the facts about whether the states of affairs that the sentence represents as being the case, actually do obtain.

According to metaphysical analyticity, however, the truth of analytic statements is fixed not by the facts of the matter but rather the facts about what the proposition means. And it is hard to see exactly how this could be the case, given that it differs so radically from the how truth is usually determined. Boghossian says:

> In general, I have no idea what would constitute a better answer to the question: What is responsible for generating the truth of a given class of statements? than something bland like 'the world' or 'the facts'; and … I cannot see how a good answer might be framed in terms of meaning, or convention, in particular. (Boghossian, 1996, p. 36)

The problem with metaphysical analyticity, then, is that it does not seem to cohere naturally with a general account of truth and meaning.

Epistemic analyticity, however, is not a theory about what *makes* statements true and as such it avoids that problem. According to epistemic analyticity, for analytic statements, a person's *understanding* of a proposition suffices for him being warranted in believing it. Holding a proposition to be epistemically analytic, then, does not entail holding there to be any kind of unusual connection between the meaning of that proposition and its truth. We may quite legitimately continue to assume that the proposition is made true in the usual way.

118

All it entails is that understanding a proposition itself generates warrant for believing it.

This does, however, require some careful articulation. How is it possible that understanding a proposition allows you to be justified in believing it? Is this genuinely a less implausible hypothesis than the hypothesis that facts about meaning can generate truth?

The standard response is to explain this by appealing to the fact that there are constitutive links between understanding a certain range of propositions and assenting to them. Anybody who understands the proposition 'All bachelors are unmarried men', the thought goes, considers it to be true, or would do so if the sentence is one they have not yet considered. And this connection between understanding and assent entails that anybody assenting to the sentence is warranted in doing so. Here is Boghossian's rationale for this thought:

> Suppose it is true that my taking A to be a warrant for believing B is constitutive of my being able to have B-thoughts … in the first place. Then doesn't it follow that I could not have been epistemically blameworthy in taking A to be a reason for believing B, even in the absence of any reason for taking A to be a reason for believing B? For how could I have had antecedent information to the effect that A is a good reason to believe B, if I could not so much have had a B-thought without taking A to be a reason for B in the first place? If inferring from A to B is required if I am to have the ingredient propositions, then it looks as though so inferring cannot be held against me, even if the inference is blind. (Boghossian, 2001 p. 240)

Here Boghossian is defending the claim that if making certain *inferences* is constitutive of understanding 'if… then…' statements, then those inferences are warranted. But the reasoning here would apply equally well to the claim that if assenting to certain propositions was constitutive of understanding them, then one would be epistemically warranted in believing those propositions.

And while I have misgivings about Boghossian's appeal to epistemic blamelessness, I take the point to largely correct. If believing that all bachelors were unmarried men *really is* constitutive of my understanding the proposition, it is difficult to see how we could rationally expect a person to have in place antecedent reasons to think that all bachelors are unmarried men before arriving at that belief. This would require a person to have gathered evidence about the gender and marital status of bachelors *before* coming to understand what a bachelor was. That expectation is altogether unreasonable. For how could you expect people who do not understand what a bachelor is to gather evidence about bachelors? There is no way to even explain to them what this task involves, without first providing them with some detailed information about what a bachelor is. And that cannot be done without first explaining what 'bachelor' means.

So the task would have failed before it could begin. By hypothesis, understanding what a bachelor is constitutively entails assenting to 'all bachelors are unmarried men'. In virtue of coming to understand what she is supposed to be gathering evidence *about*, the subject would have failed to have gathered evidence for believing the proposition, before actually coming to believe it.

And as a result, we may reasonably conclude that it is not the case that a subject *ought* to have prior reasons to believe the proposition before coming to believe it. Assuming this understanding-assent link does obtain, the believer cannot be subject to any normative demand that would require her to obtain evidence for the proposition 'all bachelors are unmarried men'. Her belief, then, is warranted independently of any empirical evidence or background information: *ex hypothesi*, it is Seeming-Independent.

So establishing that there are understanding-assent links of this kind would suffice to establish that the proposition in question is self-evident. A self-evident proposition is just a proposition that is what we might call *groundlessly* warranted: it needs no further evidence, rationalising explanation, background beliefs or experiences of any sort to be warranted. A self-evident proposition, fundamentally, is a proposition that in some sense generates its own epistemic warrant.

So if understanding-assent links do obtain for some range of propositions, then it follows from this that the propositions will be self-evident. So if we can establish that there are understanding-assent links, we will have demonstrated one possible way that a proposition can be self-evident.

4.2 Williamson on understanding-assent links

The supposition that there are understanding assent-links is a highly plausible one. Consider Grice and Strawson's contrasting examples of, on the one hand, a person (X) saying 'My neighbour's three-year-old child understands Russell's Theory of Types' and, on the other, another person (Y) saying 'My neighbour's three-year-old child is an adult'. Consider how we would respond to X and Y once we were satisfied that neither of them was lying, or joking, or speaking metaphorically. We can safely assume that the natural reaction would be to take X to be *wrong* in thinking a three-year-old really could understand Russell's Theory of Types. Three-year olds do not yet have anything like the intelligence required to understand complicated philosophical theories.

But that is importantly different to how we would typically respond to Y. On the contrary:

> [W]e shall be inclined to say that we just don't understand what Y is
> saying, and to suspect that he just does not know the meaning of
> some of the words he is using. For unless he is prepared to admit
> that he is using words in a figurative or unusual sense, we shall say,
> not that we don't believe him, but that his words have no sense. And
> whatever kind of creature is ultimately produced for our inspection,
> it will not lead us to say that what Y said was literally true, but at
> most to say that we now see what he meant. (Grice and Strawson,
> 1956, p. 151)

In the first case, the natural assumption is that the speaker is gullible or naive. But in the second case, the assumption is not only that the speaker is gullible or naive, but also that he is *semantically incompetent*: that he simply does not properly understand what an adult is. A natural presumption, then, is that there is a link between his belief that a three-year-old child could be an adult, and his failing to understand what an adult really is.

Despite the intuitive force behind this idea, Williamson has recently argued that there is *no* necessary connection between understanding and assent. Williamson argues that Grice and Strawson fail to recognise that a person with some very unusual combination background beliefs may assent or fail to assent to just about anything. In such a case, the source of the difficulty will lie in the believer having adopted set of irrational beliefs, not in his inability to understand the proposition in question. Consider the following:

> Someone may believe that normal human beings attain physical and psychological maturity at the age of three, explaining away all the evidence to the contrary by ad hoc hypotheses or conspiracy theories … However foolish these beliefs, they do not constitute linguistic incompetence. (Williamson, 2007, p. 85)

So the thought here is that it is not *necessarily* true that anybody who assents to 'My neighbour's three-year-old child is an adult' fails to understand what an adult is (or what a three-year old is). It is quite easy to imagine someone who knows full well what an adult is, but comes to believe that three-year-olds generally mature at a much faster rate than is commonly presumed. He thinks it takes children three years to achieve the sort maturity that the rest of us think it takes eighteen years to reach. While it is hard to imagine how anyone might *rationally* believe such a story, nothing prevents people from irrationally believing it. But the irrationality of this belief is just an epistemological irrationality. The problem is that he is not complying with epistemic norms, rather than simply failing to understand what an

adult is. The understanding-assent links, Williamson argues, are illusory in this instance.

Yet, as Williamson acknowledges, this example is not the clearest example of an analytic truth. So this point might establish only that this particular case is not a case where understanding-assent links obtain, rather than the stronger point that there are no understanding-assent links whatsoever.

However, Williamson argues that we can employ this strategy even for paradigmatically analytic sentences like 'every vixen is a female fox'.

Consider a person with the following strange combination of beliefs. Suppose that Peter has, as a result of a complex argument he devised, come to think that 'every *x* is a *y*' entails that 'there is at least one *x*'. Suppose further that Peter has also come to believe that all foxes became extinct some years ago. He thinks that all recent fox sightings are the result of some sort of elaborate government cover-up.

Now, given this combination of beliefs, Peter would be disinclined to assent to 'every vixen is a female fox'. But, Williamson stresses, Peter does not fail to *understand* this proposition. While there is a strong intuitive impression that there is something irrational about Peter's denial of the proposition in question, the best explanation of this irrationality is that it is an epistemic irrationality. There is nothing wrong with his linguistic or conceptual abilities: the problem comes from his irrational beliefs about foxes and his non-standard logical beliefs. And these problems do not affect his understanding.

I take it that if we *were* to insist that Peter fails to understand the proposition 'every vixen is a female fox' the most likely source of the problem would be his understanding of the term 'every'.

Nothing about this example suggests that Peter does not understand what a fox or a vixen is. Nothing gives us any reason to suppose he has trouble understanding 'female', 'is', or 'a'. However Peter *does* have non-standard beliefs about the truth conditions of 'every'. If there is to be a problem with his understanding of any of the constituent concepts of 'every vixen is a female fox' it must surely lie with his understanding of 'every'.

Williamson offers the following considerations in favour of the thought that Peter does not misunderstand 'every'. First, in practice Peter uses the term 'every' in contexts that are very much like the contexts in which everybody else uses them. His deviation from the norm appears only in unusual cases like this one.

Second, Peter is capable of offering a careful and articulate defence of his view on the logical implications of 'every' when challenged to do so. We can coherently imagine his papers on this topic to be very well-written and cleverly argued.

Third, Peter takes his theory to apply to the common English sense of 'every', not to some idiosyncratic usage. Given all of this, Williamson argues that the only plausibly conclusion is to take it that Peter means the same thing by 'every' as the rest of us do. He is not semantically incompetent.

As I reckon it, if Williamson *has* successfully shown that there are no links between understanding 'every vixen is a female fox' and assenting to it, he will have successfully scuppered the analytic conception of analyticity. Unlike Grice and Strawson's example of a three-year-old adult, this is a paradigmatic example of an analytic truth. If this *were* to turn out to not be an analytic proposition according to epistemic analyticity's criteria, then so much the worse for this conception of analyticity.

So it is crucial that the defender of an epistemic conception of analyticity has some response to this kind of example. What might such responses look like?

I take there to be two possible avenues of response. One option would be to argue that Peter is not fully competent, semantically speaking: that is, that he does not *fully understand* the sentence 'every vixen is a female fox'.

Or alternatively we could argue that there *is* an understanding-assent link of some sort at play in this example: that Peter does have some sort of disposition to assent to the sentence, but he is prevented from acting on it in this case.

Both responses face serious challenges. Arguing that Peter is semantically incompetent, as Williamson claims, initially appears to set the bar for semantic competence implausibly high. Suppose Peter is an eminent scholar whose work on

logic in widely admired. There is nothing implausible about suggesting he might be: there are many examples of brilliant academics having decidedly odd background beliefs, and of excellent logicians endorsing unconventional theories.

Now Williamson suggests that it is implausible to think that someone like that might misunderstand 'every'. If even an eminent scholar who publishes excellent papers on the implications of 'every' misunderstands the word, then what hope do laymen have? The implication of that suggestion seems to be that not only are you required to think carefully about what 'every' entails in order to be fully competent in using it, but you are required to think even *more* carefully about it than Peter, a renowned logician, does. That is altogether implausible. Only supernaturally intelligent people would qualify as being fully semantically competent. Virtually everyone would count as only partially epistemically competent. And this makes analyticity uninteresting. For:

> Understanding-assent links that do not apply to most humans would be of limited epistemological interest. The picture was that those who appear to reject analytic sentences can be excluded from the discussion because they lack the linguistic competence to engage in it; but we cannot exclude humans who reject such sentences on those grounds if the connection between rejecting them and lacking competence holds only for super-humans, not for humans.
> (Williamson 2007, p. 92)

But on the other hand, taking Peter to have some sort of disposition to assent to 'every vixen is a female fox' also seems *prima facie* implausible. On this picture, Peter is fully semantically competent, but his competence is somehow blocked: something prevents his disposition to assent to the sentence from being realised in how he goes about forming beliefs. As Asa Wikforss points out:

> [T]o fall back on the notion of 'blocked competence' seems perfectly
> ad hoc, since there is no evidence that speaker has the relevant
> disposition in the first place. (Wikforss, 2010, p. 8)[3]

In the remainder of this chapter I will argue that the account of epistemic norms developed in chapters one and two allows us to generate a non-ad hoc account of how Peter may indeed be understood to have a 'blocked competence' in cases like this. This, I shall argue, provides us with a recipe for resisting Williamson-like counter-examples to the claim that there are understanding-assent links.

4.3 Cognitive and non-cognitive accounts of understanding

Now I take it that the central idea behind the hypothesis that there are understanding-assent links is that our *semantic competency*, our ability to understand specific propositions, is in some special cases all that is required to generate assent. So let us begin by considering how we might conceive of our ability to understand propositions.

One way of categorising the various available positions in this regard would be to divide the field up into two broad categories: *cognitive* and *non-cognitive* accounts of understanding.

Cognitivists about understanding take the possession of semantic competence to involve the possession a range of *beliefs* about meaning, whereas non-cognitivists about understanding will see it merely as a kind of know-how: an ability that need not involve beliefs about meaning at all.

---

[3] Wifforss does argue against Williamson that there *are* understanding-assent links. However, her own view would be of no use to anyone hoping to base an account of analyticity on understanding-assent links, which is why I do not include it as a third option. Wikforss thinks that there are understanding-assent links that apply holistically: not to individual sentences but to entire sets of beliefs. Her position is summed up as follows:

> [T]here is after all a constitutive connection with belief, only a holistic one; it suggests that although we cannot single out single sentences as a litmus test for understanding, unless a plausible background story emerges such that the 'unorthodoxy is compensated for by orthodoxies at other points', [Williamson's example] does entail failure of understanding (Wikforss, 2010, p. 12)

Now I take it that Williamson's insistence that Peter is fully semantically competent is far more plausible on a non-cognitivist understanding of semantic competence. There are some obvious dialectical options available to a cognitivist that would allow her to argue that anybody in a situation like Peter's would clearly fail to properly understand 'every vixen is a female fox'.

Let me explain. If semantic competence *really is* a matter of possessing a set of beliefs about meaning, then there are grounds available to argue that anybody who has certain *wrong* beliefs about what words mean would simply not understand a sentence that involves those words. And this, I take it, would indicate that she cannot understand the proposition that sentence expresses.

But Peter, by hypothesis, has the wrong beliefs about what 'every' means. As such, there are grounds available from which to argue that he does not, on this view, understand the proposition 'every vixen is a female fox'. The example is explicitly set up as an example in which the subject has false beliefs *about* the logical implications of 'every': he thinks, incorrectly, that it carries logical implications that it does not carry. These sorts of logical beliefs have a clear knock-on effect with regard to semantic beliefs: anybody who takes terms to have logical implications that they do not have, and is consistent, will also develop false beliefs about how to correctly use sentences involving the term 'every'. This is precisely what seems to have happened in Peter's case: given his beliefs about the implications of 'every' he has altered his *use* of the term. And for a cognitivist about semantic competence, this change in use is to be explained by a change in belief about what the word means.

Now I take it that a cognitivist need not—and probably should not—insist that in order to understand a sentence, we must have entirely correct beliefs about what *all* of the words in that sentence mean. The cognitivist might insist that in at least some cases, having *mostly* right beliefs about what the words in a sentence mean suffices for understanding the sentence.

Suppose that someone believed that 'river' and 'stream' meant exactly the same thing. Let us assume this is because she has false beliefs about what 'river' means. I take it that a cognitivist would want to insist that that even though such a

person has a few wrong beliefs about what a river is, she can still fully understand sentences like 'the river is running very high today'.

But I assume that there will still be some grounds available to her to insist that some semantic beliefs are more important than others. That is, she might argue that Peter's mistakenly believing that 'every *x* is a y' entails 'there is at least one x' is *too important a mistake* for Peter to be properly said to be fully competent with regard to sentences like 'every vixen is a female fox'.

So the cognitivist has available the resources to resist Williamson's conclusion.

But could we *plausibly* resist Williamson in this way? Given that there is such powerful intuitive support for the thought that Peter *really is* semantically competent, we might be tempted to take Williamson's example to be a counter-example not only to the thesis that there are understanding-assent links, but also to the version of cognitivism about semantic competence sketched here.

However, the case for that is not as strong as it initially appears. In chapter one I argued that, given the non-transparency of epistemic norms, it is possible to blamelessly fail to comply with an epistemic norm. I suggested that one way in which this might occur would be if the person mistook a mere doxastic rule for an epistemic norm.

Now in chapter two I argued that those who have acquired *false epistemological beliefs* will be susceptible to this sort of error. The example I used there was someone who endorsed a version of coherentism according to which the following argument is perfectly cogent:

(P1*) I clearly and distinctly perceive that a benevolent God exists

(P2*) What I clearly and distinctly perceive is true

(C*) It is true that a benevolent God exists

Now Davies and Wright have argued—rightly in my view—that inferring (C*) on the basis of (P1*) and (P2*) would not be an acceptable inference: that type of reasoning is in violation of an epistemic norm. If they indeed are right about this, then it remains true *even if* the person in question defends an epistemological theory according which such beliefs really are warranted.

I bring this up because a cognitivist about semantic competence would be well within her rights to attempt a similar approach with regard to Peter. That is to say, she could understand possessing semantic competence to be a matter of having the correct beliefs. False semantic beliefs, irrespective of the amount of intellectual effort and epistemic care that has gone into them, could be treated the same way that I treat following inferential rules that are not reliably truth conducive. On my view, if the inferential rule one follows is not a genuine epistemic norm, then beliefs acquired via that pattern of inference are *not* warranted. This is true even if one has some sort of argument in favour of that unreliable inferential rule.

So we might take the considerations relating to the intellectual effort and epistemic care Peter has taken with regard to his semantic beliefs similarly to be considerations that relate primarily to whether or not he can be blamed for his semantic incompetence in this regard. They do not, it might be pressed, provided decisive reasons to deny his semantic incompetence.

Essentially this response would be centred around the *denial* of Williamson's accusation that claiming Peter to be semantically incompetent commits us to the view that that a speaker would need to be superhumanly intelligent in order to be semantically competent.

The point is that there is no need for a defender of this view of semantic competence to hold that Peter has been led into semantic incompetency by *not being sufficiently careful* about his semantic beliefs. If that was the case then Williamson would indeed be correct in thinking that this account of semantic competency is implausibly strict, given that Peter is far more careful than most.

But there is another option available: a cognitivist about semantic competence could simply hold that Peter has taken more than enough care in how

he formed his semantic beliefs: nevertheless, he remains semantically incompetent with regards to sentences like this one. Just as it is possible to fail to comply with an epistemic norm despite being very careful about one's beliefs, one might argue that it is possible to fail to be semantically competent despite being very careful about one's semantic beliefs.

4.4 Blocked competencies and epistemically ideal conditions

Let us turn to how a non-cognitivist about understanding would respond to Williamson.

In my view, non-cognitivism is the more plausible of the two accounts of semantic competence. So as far as I am concerned, if Williamson has succeeded only in demonstrating that there are no links between understanding and assent *on a non-cognitivist account of understanding*, that would suffice to establish that there is no good reason to think that there are understanding-assent links of any form. Let us consider, then, whether he has succeeded in demonstrating this.

According to non-cognitivism about semantic competence, beliefs about the truth conditions of any given proposition are not necessary for understanding it. Understanding a proposition, on this view, consists in how one uses the proposition, not in what one's beliefs about it are. Being semantically competent is not a matter of having a set of beliefs of any sort; rather it is a matter of possessing an ability or a skill. One understands a given word or concept if one tends to use it in roughly the right way.

Understanding the word 'river', for example, consists only in *generally* employing the term correctly. So not having a clear idea of the differences between a river and a stream does not indicate that one does not understand what a river is. Being confused over where exactly the difference lies does not preclude *generally* using the term 'river' in the right way: insofar as she uses the term in the

right way[4] most of the time, we may legitimately conclude that she understands the term.

Now As Williamson points out, Peter does use 'every' in *roughly* the same way everyone else does. For the non-cognitivist, this is a very strong indication that he really does understand 'every vixen is a female fox'.

So a non-cognitivist, unlike a cognitivist, would not be able to straightforwardly deny that Peter understands the proposition in question. However, I will argue that the non-cognitivist has available other means to resist Williamson's conclusion.

Let us begin by considering what a plausible version of the sort of understanding-assent links necessary to underwrite epistemic analyticity *would have to look like* on a non-cognitivist account.

Let us begin by considering the notion of assent. To assent to a proposition is a sort of performance: it is to *express* approval of, or agreement with, that proposition. That expression might take various forms: it could take the form of *asserting* the proposition or, more commonly, it might take the form of simply judging it to be true.

And to judge a proposition to be true is to consider the proposition and *occurrently believe* that it is true. Judging is the *act* of forming or reaffirming a belief. So while there might be close constitutive connections between assent and belief, assent differs from belief in being a performance, not a disposition.

Now according to non-cognitivism, being *semantically competent* is a matter of possessing an ability or skill: it is a kind of know-how. But, crucially, like standard dispositions, abilities persist even when we are not acting on them, and even sometimes when we are prevented from acting on them.

---

[4] What factors determine what the right way to use the term 'river' is, will I take it, vary according to our theory of meaning. It is important to see that non-cognitivism here is not an account of meaning: it does not entail that our use *determines* the correctness conditions of the words we use. Nothing prevents a non-cognitivist from assuming, with the orthodoxy, that a term's correctness conditions are fixed in part by environmental features and facts about one's linguistic community. The point is only that to understand the word our use of it must be *approximately* correct.

For example, a person with her writing-hand in a cast will not be able to write for a period of time. This does not entail that she has lost her *ability* to write. That would be a wholly unacceptable understanding of how skills or abilities work.

Rather, let us say that she knows perfectly well how to write, but that her present circumstances *prevent* her from acting on this know-how.

And many types of circumstances might prevent us from being able to do what we know how to do. She would be prevented from writing if she was stranded on a desert island, far away from pen and paper. Once again, those circumstances conspire to prevent her from writing without affecting her ability.

So if understanding is a matter of possessing an ability, then we should expect that this will be true for understanding as well. So, a non-cognitivist must allow that even if we *do* understand some proposition, we may sometimes be prevented from articulating or expressing this understanding.

And this has implications for how a non-cognitivist ought to interpret the proposal that understanding propositions like 'every vixen is a female fox' is connected in some immediate way to assenting to them. The proposal would be that assenting to 'every vixen is a female fox' just is an expression of my ability to understand it. But if understanding is an ability, then it follows that it is logically possible for us to be *prevented* from expressing or instantiating it. Circumstances can sometimes prevent us from assenting to propositions without undercutting our understanding.

The point here is that on any plausible non-cognitivist version of epistemic analyticity, what connects understanding analytic propositions to assenting to them, must be an underlying *disposition:* when *p* is analytic, to understand that *p* entails being disposed to assent to it *when* in circumstances that do not prevent us from expressing our semantic competence.

The clearest examples of such a circumstance would be cases where the subject is temporarily cognitively impaired. Consider Hawthorne's a priori gas example.

As I read the example, a priori gas affects our ability to assent to analytic propositions in the same way that alcohol affects our ability to drive. The fact that a person *would* drive poorly while under the influence of alcohol does not entail that she loses her ability to drive well every time she drinks. She still knows how to drive: her intoxication has simply placed her is in a non-ideal condition for driving. Alcohol has rendered her unable to *exercise* her competency, without actually undercutting it.

In exactly this way, falling victim to a priori gas would not entail that a person does not understand what 'every vixen is a female fox' means, even though she would not assent to it. The a priori gas simply renders her unable to bring her semantic competency into play when forming a priori beliefs. In cases like this it should be no surprise that understanding-assent links break down: this is not, I take it, a counter-example to any *plausible* version of the claim that there are links between understanding propositions and assenting to them.

So on a plausible version of the thesis that there are understanding-assent links of the sort Boghossian proposes, the connection between understanding and assent is only present in what we might call *epistemically ideal conditions*: those circumstances where our coming to assent to certain propositions is unaffected by extraneous factors that in some way have prevented us from bringing our semantic competency to bear on our beliefs.

The thought, then, is that if proposition *p* is analytic, then, anybody who understands it will assent to it when in epistemically ideal circumstances. And we have good reason, I will argue, to think that the conditions that Peter is in are not epistemically ideal. Claiming that cases like this are cases where the subject has a blocked semantic competency, then, is not ad hoc.

My argument has two stages. First, following Wikforss, I argue that understanding-assent links can only plausibly be understood to break down in those cases where the subject has some non-conventional background beliefs.

Second, I argue that the role these background beliefs play can be plausibly understood as rendering one's situation epistemically non-ideal. As a result, these need not be understood as counter-examples to the thesis that if someone

133

understands the proposition 'every vixen is a female fox' then she will assent to it *if in epistemically ideal circumstances*. So there remains an avenue by which epistemic analyticity can escape Williamson's objections, even on this conception of semantic competence.

Williamson takes examples like the example of Peter to serve as particularly good illustrations that it is possible for understanding and assent to come apart for paradigmatically analytic propositions. They are not, he thinks, the only ways in which this is possible. A community of non-logicians reasoning just like Peter might be another example of people for whom understanding-assent links break down. Once we have accepted that deviant logicians fully understand the words the propositions they resist assenting to, 'we can hardly refuse that same classification to other speakers merely on the grounds of their unacquaintance with formal semantics' (Williamson, 2007, p. 99).

But this is not convincing. Wikforss puts the problem with this suggestion as follows:

> Imagine a speaker who walks around questioning sentences like
> 'Every vixen is a female fox', and related inferences, without
> providing any reasons whatsoever for her deviance ... Is there any
> inclination at all to say that such a speaker is semantically competent?
> I should think not, and the reason is precisely that there is not a set
> of background beliefs that allows us to make sense of the deviance.
> (Wikforss, 2010, p. 13)

I take it that Wikforss is entirely correct on this point. Williamson must surely allow that there are speakers who are semantically incompetent: who *do not* properly understand some of the sentences they assent to. And the sort of speaker Wikforss describes is as clear an example of such speakers as any. If people like these do not count as failing to understand the proposition 'every vixen is female fox' it is difficult to see how anybody could fail to. It is one thing to conclude that Peter fully understands the proposition in question without assenting to it. But it is quite

another to conclude that speakers *without* Peter's background beliefs who resist assenting to the proposition, still understand it.

Now Wikforss takes this point to establish only that there are very general 'holistic' understanding-assent links that apply to a set of beliefs as a whole.

But another option is to understand these background beliefs as rendering the subject's situation epistemically non-ideal, and hence that these are not the sorts of cases where we should expect there to be any connections between understanding and assent.

In motivating this thought, I appeal to the account of epistemic rationality developed in the first three chapters. Now I have argued that the presence of certain types of false background beliefs can lead to one performing *irrational* inferences: inferences that are in violation of the relevant epistemic norms. We have excellent reasons to suppose that our epistemological beliefs affect how we actually reason. Coming to believe that Gambler's Fallacy is a fallacy is in most cases sufficient to change our probabilistic reasoning: to shift from following one doxastic rule to another doxastic rule. And it is perfectly possible to develop the wrong beliefs about how one *ought* to be reasoning.

Consider Clifford's famous declaration: 'it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence' (Clifford 1877, p. 76). Now this dictum has clear implications for what we ought to believe and what we ought not to. For one thing, if this is true, then it entails that we ought not to adopt any beliefs through something like religious faith. This is in effect a very stringent interpretation of what epistemic norms demand from us.

And it is not an interpretation that is universally shared. For disagreement we need look no further than William James' famous response to Clifford in *The Will to Believe* (1896). There James argues that there are indeed occasions where one is entitled to hold a belief based purely on faith.

Suppose for the sake of argument that James is wrong about this: that Clifford was right in thinking that it is 'wrong' to believe propositions based on faith alone. It follows from this that James has false beliefs about how we *ought* to go about forming and maintaining our beliefs.

135

It is perfectly possible, then, that James might resultantly adopt incorrect doxastic rules. All that we need suppose in order to imagine such a possibility is that James' epistemological beliefs affect how he reasons in roughly the way that our beliefs about probabilities lead to us abandoning the representativeness heuristic when calculating the probabilities of the next coin toss landing 'heads' or 'tails'. The difference is that we are supposing that James is coming to reason in ways he ought not to. His philosophical arguments have actually resulted in him *believing irrationally*. Given the doxastic rule we are supposing he adopts is *ex hypothesi* incorrect, beliefs based on that rule are not beliefs arrived at by following an epistemic norm. They are unwarranted, and hence irrational.

Now—depending on how plausible we take James' argument for his epistemological view to be—I take it that we may legitimately understand James' coming to reason incorrectly to be something that we cannot hold him responsible for; that is, we might consider him epistemically *blameless* for the irrational beliefs arrived at by following his own doxastic rule.

If that is the case, then in effect we are taking James to be *epistemically unlucky* in a distinctive sense. He is unfortunate in that his best efforts to be an epistemically responsible agent have back-fired. His philosophical theorising has resulted in him adopting irrational belief-forming methods: it has led him down a blind alley.

Now it is important to see that the fact that he is unfortunate in this sense *only* indicates that he is blameless for his beliefs, and not that those beliefs are genuinely rational.

With that in mind, let us compare the situation James is in with Hawthorne's a priori gas thought experiment. What I want to point out here is that holding false epistemological beliefs of this sort is in some respects very much like having fallen victim to a priori gas.

One way to understand such cases is to understand them as situations where the subject has, through misfortune, been placed in a position where a specific range of his beliefs will be *irrational*. Once we accept that our false epistemological views may lead us to reason irrationally, there is no significant

epistemological difference between cases where one has fallen victim to a something like an a priori gas, and cases where one has adopted incorrect reasoning methods as a result of false epistemological beliefs. In both cases, there are hindrances interfering with the subject's ability to reason correctly.

But if that is the case, then it seems that *nothing prevents us* from viewing Peter's background beliefs as rendering his situation as epistemically non-ideal. Here we have a case of a person's false theoretical background beliefs being directly responsible for his denying a true proposition. Once we allow that having false theoretical beliefs, no matter how well reasoned, can actually have devastating, large-scale effects on the epistemic standing of one's beliefs, then we must conclude that false background beliefs of this sort are epistemologically relevant in largely the same way that a priori gas is epistemologically relevant.

Given, first, that false logical beliefs are *required* to provide a decisive break between understanding and assent in this case, and, second, that false logical beliefs affect the *epistemic rationality* of a believer, there is room to maintain that these examples are not decisive.

We already know that understanding-assent links only plausibly apply to those in the right sort of epistemic context: when we are thinking clearly or when we are not under the influence of anything that might impair our ability to follow epistemic norms, for instance.

My suggestion is that the examples Williamson has provided can plausibly be understood as being just like these cases. Just as a priori gas, or extreme tiredness, or alcohol, may block the connection between our general ability to understand a sentence and our assenting to it, so too may the sorts of background beliefs Peter has.

The root of the problem is to be found in Williamson's reflection on whether Peter and Stephen (a similarly deviant logician who also does not assent to 'every vixen is a female fox') would have assented to it had they lacked those background beliefs. Rhetorically, he asks:

> Would Peter and Stephen assent to ['every vixen is a female fox'] if
> they lacked their conscious theoretical commitments? Perhaps not,
> but that counterfactual would show little. (Williamson, 2007, p. 102)

My suggestion is that Williamson underestimates the importance of that counterfactual. Theoretical commitments of this sort lead to systematically incorrect reasoning, just as falling victim to a priori gas leads to systematically incorrect reasoning. Neither are situations in which we should expect the link between understand and assent to be preserved.

The point is simply that the question of whether a person has the sorts of theoretical commitments that Peter has, is epistemologically relevant. It affects the subject's epistemic standing. As a result, we have some basis from which to claim that Peter may be understood to be in epistemically non-ideal conditions: conditions in which his general semantic competency is blocked. The account of epistemic normativity generated in the previous chapters strongly suggests that being in epistemically ideal conditions is not *just* a matter of being cool, calm and collected. It is also a matter of having avoided any deviant theoretical commitments.

4.5 Analyticity and semantic intuitions

Now there are naturally a number of possible ways in which one might respond to Williamson's counter-examples. What I want to point out in this section is that the response I have sketched here has some welcome theoretical consequences.

Foremost of these is that if the argument presented above is right, then it is possible that we could legitimately extend analyticity to apply not only to what seem to be very trivial, uninteresting truths like 'every vixen is a female fox' and 'all bachelors are unmarried', but also to more substantive, theoretically interesting philosophical conclusions.

I tentatively offer the following as a possible example of a theoretically interesting conclusion that can be construed as epistemically analytic: 'If Smith is

justified in believing that Jones has ten coins in his pocket, and Smith is justified in believing that Jones will get the job they have both been interviewed for, and Smith infers from these beliefs that the person who will get the job has ten coins in his pocket, then, if it turns out that in fact Smith gets the job, and that Smith unknowingly also has ten coins in his pocket, Smith does not *know* that the person who will get the job has ten coins in his pocket even though he is justified in believing it, and it is true' (henceforth, the Gettier Claim).

Before offering some reasons for thinking that we *can* legitimately include claims like that as epistemically analytic, let me first offer some reasons for thinking that being able to do so is a welcome consequence.

Most philosophers, I assume, take themselves to *know a priori* that in that scenario, Smith does not know that the person who will get the job has ten coins in his pocket. That we can have a priori knowledge of that sort is a central epistemic intuition: that intuition underpins the widely held view that philosophy is an a priori discipline.

Now I take it to be an explanatory advantage to be able to appeal to analyticity in explaining *how* we have such knowledge. Not only does this mean that we do not need to provide separate explanations in order to explain how we know that all bachelors are unmarried *and* how we know that Smith does not have knowledge, but there is also strong intuitive support for the idea that those propositions that are knowable via *conceptual analysis* are analytic.

So it would be a welcome consequence of epistemic analyticity if it could provide an account of analyticity that could apply to more than trivial, uninteresting statements.

And it is not immediately obvious that it can do so. If the Gettier Claim is analytic, then there are a great deal *real life* cases like the imaginary case of Peter. If the Weinberg, Stich and Nichols study mentioned in chapter two is anything to go by, then there are a vast number people who simply do not have the intuition that Smith lacks knowledge. And so they would not assent to the Gettier Claim.

Moreover, Brian Weatherson, to name one example, would be a real life equivalent of Williamson's Peter. Weatherson, like Peter, does not assent to the

supposedly analytic proposition in question, and backs up his denial with plausible-looking philosophical arguments.

The point here is that given that there are a number of real life philosophers who deny Gettier's intuition, and that there are apparently a large number of laymen who lack the intuition altogether, it is not immediately obvious that there are understanding assent-links at play with regard to this particular type of proposition.

This strongly indicates that the conclusion that propositions known through conceptual analysis are analytic does not fit naturally with epistemic analyticity. Epistemic analyticity only holds for those propositions for which there are understanding-assent links: since there do not seem to be such links in these cases, there is a significant *prima facie* problem here.

However the response I offer to Williamson would apply equally effectively in this case. Rather than insisting—implausibly, in my view—that none of these laymen or deviant philosophers *understand* what 'knowledge' means, we might simply say that, like Peter, they are in epistemically non-ideal conditions.

I take it that I do not need to defend the idea that the considerations I brought to bear in favour of thinking Peter is in epistemically non-ideal conditions apply also to Weatherson. However, the idea that they might apply also to non-philosophers who lack Gettier's intuition requires some explication.

In chapter two I suggested that a plausible way to explain how people can come to have divergent philosophical intuitions, and also how people can *change* their intuitions over time, is to take intuitions to be open to infection from our background beliefs, opinions, philosophical inclinations, folk theories, idiosyncratic presumptions, and so forth.

The fact that intuitions are so diverse and liable to change played a significant role in my argument that a priori warrant is *independent* of intellectual seemings: intellectual seemings, I argued, are the sort of things that provide a viable epistemic basis.

But here we can also use this claim to argue a related point: that the deviant intuitions resulting from these background beliefs, opinions, presumptions, and so forth, block the links between understanding and assent in these cases.

The fundamental claim in the last section was that having certain epistemological and logical background beliefs can actually affect whether or not one forms beliefs *rationally*. Our background beliefs can have a direct effect on the epistemic status of what we might call our foreground beliefs. So if that can be true of deviant logical and epistemological beliefs, why not think that—for instance— one's deviant semantic intuitions can likewise have an adverse effect on the rationality of our foreground beliefs?

Consider the following scenario. Suppose Mary is a normal English-speaker who, through no fault of her own, comes to incorrectly employ a particular concept —*sofa*. Let us suppose that Mary acquired the concept initially from her parents, who are in full mastery of the concept. However, during the initial learning-stage, Mary was exposed to no single-seated, upholstered armchairs, and so never picked up that her parents would not employ the concept *sofa* to refer to such chairs. Mary encounters her first upholstered armchair in the company of her schoolteacher, who incorrectly calls such chairs 'sofas'. Co-incidentally, in the years that follow, Mary only ever describes upholstered armchairs as 'sofas' in the company of those who also misuse the word 'sofa'. As such, Mary is never corrected when she uses the word in this way.

So Mary, through no fault of her own, has simply acquired deviant semantic intuitions about 'sofas' and these intuitions lead her to systematically misapply the term.

Now suppose that Mary performs an a priori conceptual analysis of the concept *sofa*. She would inevitably arrive at false beliefs about the necessary and sufficient conditions of sofas.

Now would her resulting sofa-beliefs be warranted? On my account, such beliefs would not properly count as warranted. While we may not *blame her* for the erroneous sofa-beliefs she acquires via conceptual analysis, it is nonetheless not the case that these beliefs are properly rational.

141

I take it that this scenario very closely resembles Peter's. Just as Peter's theoretical beliefs had a direct effect on the rationality of a range of his beliefs, so too do Mary's semantic intuitions about sofa's undermine the rationality of her sofa-beliefs. These intuitions are epistemically relevant in just the way that Peter's beliefs are epistemically relevant. Like Peter, then, we can legitimately take Mary to be in epistemically non-ideal circumstances.

Once we have allowed that this is possible in cases like Mary's, why not suppose that the same thing is going on with those who have deviant semantic intuitions about philosophically interesting concepts like 'knowledge'? There seems no obvious reason that a similar approach would not apply just as well.

What this indicates is that this response is particularly well equipped to insist that even epistemically analytic propositions do not need to be *transparently true* for everybody, all the time. By narrowing the scope of the links between understanding and assent to apply only in epistemically ideal conditions, and by allowing for the possibility that our background philosophical beliefs, folk theories and cultural opinions can place us in non-ideal conditions, we are well insulated from counter-examples.

This has the further advantage of potentially allowing us to extend the scope of epistemic analyticity, so that it can apply to a broader range of beliefs than we might initially have thought. The notion of epistemically non-ideal conditions developed here allows —or initially appears to allow—the notion of epistemic analyticity to do some substantial explanatory work. It becomes a notion we might plausibly appeal to in order to explain our knowledge of a great many intuitively a priori propositions.

4.6 Conclusion

The central argument in this chapter has been that Williamson's counter-examples are not decisive. There is still conceptual space to maintain that, *when a subject is in epistemically ideal conditions*, it is true that if she understands an analytic proposition like 'every vixen is a female fox', then she will assent to it. This way of

narrowing the scope of the connection between understanding and assent is not, I think, ad hoc. It is clear that such a connection could *only* conceivably be reliably present when one is in the right sort of epistemic context. And I have also provided good reasons to think that false theoretical commitments, and perhaps even false semantic intuitions, can indeed affect one's epistemic context. We have good theoretical grounds, then, for taking cases like Peter's to fail to establish that there are no links between understanding and assent.

Given that we can maintain the thesis that there are constitutive links of a sort between understanding and assent, epistemic analyticity remains a viable option.

And this gives us reason to think that the category of *self-evident propositions* is not an empty category. Epistemically analytic propositions meet the criteria for being self-evident.

5: Self-Evidence and Self-Knowledge

In the previous chapter I suggested that one way in which a proposition may be self-evident is if it is an *epistemically analytic* proposition: a proposition that is such that understanding it warrants the belief that it is true. And this happens, I argued, when it is the case that there are constitutive connections between *understanding* the proposition and assenting to it (when in epistemically ideal conditions). This type of self-evident proposition, then, is made self-evident in virtue of possessing an unusual type of *content*: a content that generates or entails this sort of constitutive connection between understanding and assent.

In this chapter, I will argue that there is also a range of propositions that are self-evident not in virtue of possessing a special type of content, but rather in virtue of us having a special kind of access to them. As I will show, epistemically analytic propositions are not the only obvious examples of self-evidence. There are also a range of *self-ascriptions* that bear all the hallmarks of being self-evident.

This chapter has two closely interrelated aims. One aim is to offer some reasons for thinking that a certain range of self-ascriptions are self-evident. The other aim is to show that taking self-knowledge to be a priori is plausible both as an account of self-knowledge, and as an account of the a priori. The central contention to this effect will be that appealing to the a priori offers an elegant, if surprising, solution to the Problem of Self-Knowledge.

5.1   Self-evidence and the Problem of Self-Knowledge

One good reason to suspect that at least some of the claims we make about our own minds are self-evident is that our conversational and epistemic practices strongly indicate that this is so. Consider, for example, those thoughts that Burge (1996) has famously called 'cogito-like judgements': judgements of the form 'I am thinking that *p*'. Now these are paradigmatic examples of propositions that strike us as obviously true when we think them. When I deliberate about what to eat for

lunch, for instance, it strikes me as self-evident that I am *thinking about* what to eat for lunch. When I am so deliberating, I take the second-order claim 'I am thinking about what to eat for lunch' to be just as obvious as the claim 'nothing can be both red and green all over at the same time'. In neither case is the claim in question obviously based on any sort of evidence, and in both cases, I treat the proposition as beyond any epistemic reproach. Both statements are treated as if they neither depend upon nor require evidence.

This apparent self-evidence is part of what we might call the *Problem of Self-Knowledge*. That is to say, one of the philosophically puzzling features of self-knowledge is that we do treat a range of self-ascriptions—including cogito-like thoughts—as if they are self-evident to those who make them. For it is not obvious why we *should* treat them as self-evident.

As Elizabeth Fricker puts it, the phenomenon that is to be explained by any account of self-knowledge is that self-ascriptions like these are *Language-Game* (henceforth, LG) authoritative, LG-basic, and apparently psychologically non-inferred (Fricker, 1998, p. 157).

What Fricker finds to be the central puzzle about this range of self-ascriptions is the following. First, we treat people as being special authorities on what they are thinking. Unless we have reason to doubt the sanity of the person involved, we treat people as being experts about what their beliefs, desires, intentions and sensations are. If person *a* and person *b* give conflicting reports about what *a* is thinking, then we invariably take *a*'s word for it over *b*'s. The burden of proof is taken to lie with *b* in this sort of case. Person *b* would need to provide very strong reasons indeed to persuade us that *a* is mistaken about what she is thinking.

One familiar scenario in which we *would* be inclined to side with *b* over *a* would be if we had reasons to suspect that *a* is lying. In a totalitarian regime, for example, members of the secret police might take the word of an informer (*b*) over the word of a suspect (*a*) about what the suspect thinks about the government. But this is not a counter-example to the claim that *a* is an expert on what she believes. The fact that *a* would lie about her anti-government beliefs if she did have them

does not conflict with the thought that *a* is in a better position than *b* to judge the content of her own thoughts. That we assume that *a* is a special authority on her own thoughts entails that some range of self-ascriptions are LG-authoritative.

A related, but distinct, point is that we do not require people to defend or explain self-ascriptions like cogito-like thoughts. If *b* claims that *a* harbours anti-government sentiments, we would be well within our rights to ask *b* how she knows this. Is this based on something *b* saw *a* doing, or heard *a* say? Is this based on some form of statistical profiling? However, asking people how they know what they themselves believe is not a live conversational option. Asking people to defend or explain their cogito-like thoughts, for instance, is a grossly unreasonable request by the standards of our epistemic and conversational practices. This is the sense in which self-ascriptions like cogito-like thoughts are LG-basic. People are not just treated as experts on their own minds; they are treated as experts who are not expected to explain or defend their assertions.

Finally, the third feature to be explained is that cogito-like thoughts strike us as being *psychologically direct*. They do not appear to us as having been inferred from anything else we believe. While we might sometimes arrive very quickly at beliefs about other people's minds, we could, upon reflection, trace the inferential pattern we followed. Not so with cogito-like judgements: these appear to us to be psychologically non-inferred.

A quick point on what these features add up to. The most epistemologically interesting of these three features is LG-basicness. LG-authority is not by itself an uncommon feature. There are many different ways in which one person may be in an epistemically privileged position regarding one type of enquiry. Many fields of enquiry require specific skills, experience, or background beliefs.

For instance, suppose that while sitting in the garden in the evening, I spot a bird that I take to be a nightingale. Suppose further that this is disputed by my companion, an ornithologist, who insists that the bird is a robin. Now for any third party who has not seen the bird in question, I take it that the reasonable thing to conclude here would be to side with the ornithologist rather than me. I have little interest in birds and am poor at telling them apart. In effect, an ornithologist is LG-

authoritative when it comes to identifying birds: she is taken to be *in an epistemically privileged position* on such topics as a result of her expertise.

Likewise, there are many other beliefs that are not—or not taken to be—psychologically inferred from anything else. Arguably, perceptual beliefs are like this. While there might be some sort of *cognitive transition* from sense experience to perceptual belief, this transition is not an *inference.* An inference occurs only when the transition in question is a move from one belief to another belief. Cognitive transitions like the move from perceptual experience to belief do not count as inferences. So the fact that cogito-like thoughts are psychologically non-inferred is so common as to make it hardly worth commenting on: that characteristic is shared by a substantial portion of our beliefs.

However, LG-basicness is far rarer than it may first appear. The assertions of experts, for instance, are not normally LG-basic. An expert witness in a criminal trial may legitimately be asked to explain and defend whatever claims she makes. Questioning experts is also perfectly appropriate in other contexts. We may question experts not only to ensure that their assertions are correct but also out of epistemological interest. It would be perfectly appropriate for me to ask the ornithologist how she knew the bird was a robin rather than a nightingale, even if I do not doubt that she is right about this. I might simply be curious about how she knew.

This latter example also illustrates a related point: that while we standardly take perceptual experiences at face value, perceptual judgements are not LG-basic either. It is unusual to question people's perceptual beliefs, but it is not unheard of, nor is it always inappropriate. Some people are better at perceptually identifying certain things than others. Just as some people might be better than us at telling one type of bird from another, some people are better than us at identifying wines by taste or by smell, or at identifying a composer by the sound of their music. In all of these cases it is perfectly acceptable conversational practice to ask such people *how* they knew it was one type of bird or wine rather than another type, or one composer rather than a different composer.

We might ask for an explanation out of simple epistemological curiosity, or we might do so because we doubt her conclusion. In the court room, we consider it perfectly appropriate to thoroughly examine not only the testimony of experts but also the testimony of eye-witnesses. The fact that it strikes an eye-witness as having been *perceptually obvious* that the defendant was the person she saw fleeing the scene of the crime does not render her belief to this effect beyond reproach. We might, perfectly legitimately, ask the witness to produce some sort of reason for thinking that the person that she saw fleeing the crime *really was* the defendant rather than someone else. If she fails to provide compelling reasons for this, we might well suspect that she might have made an error of misidentification.

In even the most rigorous and careful of court-rooms, however, once it was satisfactorily established that the witness was not lying, it would remain unchallenged whether the witness *believes* it was the defender who she saw fleeing the crime. Her sincerely asserting that to be the case is taken to settle the question.

The only clear examples of LG-basic beliefs other than self-ascriptions, interestingly, are those very simple conceptual, mathematical or logical beliefs, like 'nothing is red and green all over at the same time', '1+1=2', or 'all bachelors are unmarried men'. Beliefs like this are LG-basic precisely because, for most of us, there is *nothing to say* in response to questions about how we know them. We *just know* that propositions like this are true: we know them simply because they are self-evident.[1]

So the combination of these three features, I submit, suggests that cogito-like thoughts are taken to be self-evident for those who think them. That they are LG-basic and (believed to be) non-inferred indicates that we treat cogito-like thoughts as if they are self-evident: as if they are propositions that are just immediately obvious to those that think them.

---

[1] That is not to say that a logician, epistemologist or a mathematician would not be able to offer a substantial explanation of how we know such things; clearly they could. But an epistemologist might also offer a substantial explanation of how we know our cogito-like thoughts. These beliefs are LG-basic because these theory-based explanations are *the only* types of explanation available. These sorts of explanations are not the sort that non-experts can be expected to offer. Our 'language game' does not insist on people having explanations like these available.

But that they are LG-authoritative suggests that if cogito-like thoughts *really are* self-evident, their self-evidence is to be explained by the type of access the subject has to them, rather than, say, appealing to them having the type of content that produces understanding-assent links. That a person is thinking that *p* might be self-evident to her but it is not self-evident to an observer, even though the observer understands the proposition 'Person *a* thinks that *p*' perfectly well.

5.2 Vindicating and non-vindicating explanations

Of course, just because our conversational practices seem to indicate that something is the case, it does not follow that it *is* the case. On the contrary, I take it that there are no less than three possible solutions to the Problem of Self-Knowledge. The first solution is to endorse what Fricker calls a non-vindicating explanation of our practices regarding self-knowledge. This account can acknowledge that we tend to treat self-ascriptions as if they are LG-authoritative, LG-basic and psychologically non-inferred, but will hold that this practice is, as it turns out, based upon some misconception on our part. Commonly, such a view will be supported by empirical evidence that, it is claimed, shows that our access to our own thoughts is substantially more fallible than we assume. We have, it is claimed, no special epistemic access to the contents of our own minds, and our practice of acting as if this is the case is based upon an illusion.

The second solution is to endorse a vindicating but deflationary explanation of our practices. On this account, it *is* perfectly appropriate for us to treat certain self-ascriptions in this way. But this is not because people are capable of knowing their own cogito-like thoughts in a way that no others can. That we treat them in this way has no significant epistemological implications at all: it is merely an 'artefact of grammar' that explains the appropriateness of this practice.

The standard way of defending such a view is to argue that cogito-like thoughts and other related self-attributions (avowals), function very differently in our language-game from the way reports function. Essentially, the idea is that there is no plausible way of explaining how avowals can be reports of independent states of affairs while accounting for the features unique to them—that is, as Crispin

Wright would put it, the fact that they are 'groundless' and 'authoritative' (Wright 1998). Reports, it is claimed, follow a different set of grammatical rules from the ones followed by avowals, and, consequently, avowals cannot be genuine reports. This is a kind of vindicating explanation of our common practices. We are right to treat avowals as LG-authoritative, -basic and psychologically non-inferred, according to this account, but this is simply because of the way the language-game is played. These features of avowals are not to be explained in terms of some underlying cognitive ability, but can be understood as mere artefacts of grammar. This, it is claimed, is the most basic explanation we can give here—the explanatory bedrock is language.

Third, and finally, we might endorse a vindicating, non-deflationary account of our practices. On this account, it is perfectly appropriate for us to treat self-ascriptions in the way we do treat them, and it is appropriate *because* people are able to know, in a way that no others can, exactly what they are thinking. This account, then, takes the special status of cogito-like thoughts not to be a status that is *granted* by our linguistic mores, but a status that is *earned* by the believer. That we treat self-ascribers as beyond epistemic reproach is a result of a genuine cognitive achievement on their part.

Now, of these three possible solutions, I take it that a deflationary vindicating explanation is only worth considering as a last resort. Its plausibility hinges crucially on a non-deflationary account being *untenable*. That is to say, if we had at hand a satisfactory explanation of how avowals *are* the result of the sort of 'genuine cognitive achievement' that is required for knowledge, it would be difficult to see how a deflationary account would be in any way appealing. For what possible reason would such an account be preferable? As epistemologists, I take it that our aim should be to offer a proper epistemological solution to the Problem of Self-Knowledge. We should only offer a non-epistemological solution if we have reason to think we can do no better.

I also see no reason to be pessimistic about the prospects of the 'groundlessness' and 'authoritativeness' of avowals being fully accommodated by an epistemological theory. Authoritativeness is not a difficult characteristic to

incorporate into an epistemological theory. There are many fields of enquiry that admit of experts, people who we consider to be in a better epistemic position than the rest of us.

And to say that self-knowledge is groundless is simply to say that it is *basic* in the terminology I have been using here. Here is Wright's explanation of what groundlessness is:

> The demand that someone produce reasons or corroborating
> evidence for such a claim about themselves—'How can you tell?'—is
> always inappropriate. There is nothing they might reasonably be
> expected to say. In that sense, there is nothing on which such claims
> are based. (Wright 1998, p. 14)

However, note that on the epistemological account I have put forward, the groundlessness of a statement does not indicate that it is not *known.* Non-inferential a priori knowledge is groundless in just this way and yet is still a form of *knowledge*. That is to say, I have argued that non-inferential a priori knowledge is the knowledge of self-evident propositions, and self-evident propositions are groundless propositions. A priori knowledge, on the account I have put forward, just is knowledge of, and knowledge epistemically based upon, self-evident judgements. As such we already have an epistemology in place that can accommodate groundless beliefs. So if it turns out that cogito-like beliefs really are groundless (rather than just *assumed to be* groundless) that would not prevent them from counting as knowledge.

So I shall set this possible solution to the Problem of Self-Knowledge to one side.

5.3   Scepticism about self-knowledge

In this section I discuss the prospects of a non-vindicating solution. I shall argue, first, that empirical research does not support the sort of scepticism about self-

knowledge required for a non-vindicating solution to the Problem of Self-Knowledge. And second, that reflection on the implications of such a scepticism strongly indicates that it is false.

A cornerstone in the psychological literature about self-knowledge is Nisbett and Wilson's seminal paper 'Telling More than We Can Know' (1977). In that paper, Nisbett and Wilson report that studies they conducted indicate that people are surprisingly poor at correctly identifying *why* they make certain choices or why they hold certain beliefs. Their general methodology was to put people into situations where their choices were being manipulated, and then to ask them to explain their choices.

One scenario was where subjects were asked to pick out the best item of clothing from an array of similar-looking items, and then explain why they chose the item they did. Now, by happy coincidence, Nisbett and Wilson found that the overwhelming majority of the subjects fell prey to a cognitive bias that Nisbett and Wilson call 'The Position Effect' (Nisbett and Wilson, 1997, p. 243) whereby the right-most item is strongly preferred.

Yet, when asked to explain their choices, the subjects denied that the *placement* of the item had anything to do with their choice. Nisbett and Wilson found that subjects would insist that the item was chosen because it appeared to be of a higher quality than the others.

On the basis of this evidence, Nisbett and Wilson argue that we have little or no introspective access to our own higher order 'cognitive processes'.

While these findings are sometimes taken to indicate that we ought to adopt a more sceptical approach to people's self-ascriptions than we currently do, it is important to see that these findings do not supply us with any reason at all to doubt that self-ascriptions about *what* we thinking are unreliable, or even fallible at all. Nothing here suggests that we are bad at identifying what we prefer. It merely suggests that we are bad at identifying *why* we prefer it. This study lends itself to scepticism about our knowledge of the processes that produce our thoughts, not to scepticism about self-knowledge generally.

Nisbett and Wilson are well aware of this. Consider their explanation of *why* people are prone to think of themselves as being experts about their own cognitive processes, despite this clear evidence to the contrary:

> The individual knows … the focus of his attention at any given point in time; he knows what his current sensations are and has what almost all psychologists and philosophers would assert to be "knowledge" at least quantitatively superior to that of observers concerning his emotions, evaluations, and plans. Given that the individual does possess a great deal of accurate knowledge and much additional "knowledge" that is at least superior to that of any observer, it becomes less surprising that people would persist in believing that they have, in addition, direct access to their own cognitive processes. (Nisbett & Wilson 1977, p. 255)

So, illuminatingly, Nisbett and Wilson account for our mistakenly taking ourselves to be authoritative about our own cognitive processes by hypothesising that we *over-generalize.* In their view, people are indeed experts about what their own 'emotions, evaluations, and plans' .The problem is that they think that their expertise extends further than it does. They go on to say:

> A related point is that we are often capable of describing intermediate results of a series of mental operations in such a way as to promote the feeling that we are describing the operations themselves. (Nisbett & Wilson 1977, p 255)

So it is important to see that the findings of Nisbett and Wilson do not support a wholesale scepticism about self-knowledge. On the contrary, the idea that we do enjoy a special, distinctive kind of self-knowledge is posited as an explanation for the fact that people are so easily misled into thinking they possess an epistemic authority that they do not have.

Alison Gopnik, on the other hand, uses empirical evidence to draw a stronger conclusion than Nisbett and Wilson. Gopnik argues, controversially, that studies indicate that we do not know our own psychological states differently to how we know about the psychological states of others.

In the studies in question, psychologists would present a young child with some object that appeared to be another sort of object, and then reveal the deception. The child would then be asked about what she initially believed the object to be. It was found that the children would generally get this wrong. They would maintain that they had *always* believed the object to be the way they currently believed it to be.

Children improve at such tasks, Gopnik claims, once they started to pick up the common folk psychology of their parents and the people around them. This, Gopnik argues, supports the view that how we come to know about our own minds in the same sort of way to how we come to know the minds of others—by interpreting behaviour in light of some theory of how the mind operates, and how it is connected to behaviour. She says:

> Suppose the commonsense and philosophical account of privileged first-person beliefs about the mind were correct. Then we should predict that, however erroneous children's views of the psychological states of others might be, they would not make similar errors in their understanding of their own psychological states. (Gopnik 1993, p. 6)

One concern with this, however, is that Gopnik seems to be taking these results to be more conclusive than they really are. For instance, should we really expect, on the 'common-sense' view, that very young children should not make these sorts of errors?

It is not obvious we should expect this at all. One way of interpreting this data would be that what happens when children reach the point where they suddenly become much better at both reporting on their own immediately past beliefs and on the states of mind of others, is that it is only at this point that they

properly *understand* what a belief is. That is, the reason that they become so much better at telling both what they themselves previously believed, and what others believe, is that before that point they did not have a clear understanding of what type of thing a belief was. So, the hypothesis goes, the young children used in these studies were bad at reporting on their immediately past beliefs because, despite Gopnik's protestations to the contrary, they did not properly understand what they were being asked to report on.

On that interpretation, the fact that children are bad at these tasks does not entail that the 'common-sense' view is wrong. Any account of self-knowledge, I take it, will restrict itself to applying only to those who understand what beliefs, desires, intentions and so on actually are. If we suppose that a person does not understand what a belief is, it is hard to see why we should expect her to be able to report reliably on *anyone's* beliefs, including her own beliefs.

Furthermore, even if we lay the possibility that the children do not properly grasp what a belief is to one side, these cases still only indicate that children are bad at judging what they *previously* believed, not that they are bad at judging what they *currently* believe. At best, these cases might support a scepticism about diachronic self-knowledge. They do not at all support scepticism about synchronic self-knowledge: the knowledge of what we are thinking *now*.

This point is crucial. If it is only past beliefs that children are poor at reporting on, their failure might be a result of *misremembering,* rather than what we might call an introspective failure. It might be—and indeed Gopnik does not dispute this—that the children knew very well that they believed the object in question was object *o* at the time when they actually believed that. So none of this empirical evidence here suggests that normal adult humans do not know what it is that they believe *at present*. This evidence is far less conclusive than Gopnik makes out.

That is not to say that the results of the studies cited by Nisbett and Wilson or Gopnik are not important in their own right: for clearly they are. What they do, however, is help us get a clearer idea of where the *borders* of that range of self-ascriptions we are authoritative about lie. In other words, research like this can

serve as a useful guide as to how far the authority of self-knowledge extends. It gives us reason to suspect it does not apply to the people's reports of the *reasons* their beliefs are based on, and that it *might* not apply to even our immediately past beliefs. This in itself is a significant outcome. But it does not suggest that we should be sceptical about synchronic self-knowledge: the knowledge of what our current beliefs, desires, intentions, sensations and so forth are.

And there are powerful reasons to think that any large-scale scepticism about synchronic self-knowledge is false. The concern here is that if a sweeping scepticism about self-knowledge were true, then would be difficult to make sense of how we could be as accurate as we are in our judgements about our own states of mind, and in our judgements about the states of mind of other people.

It is worth stressing that a great deal of what we do requires some level of cooperation from the people around us. We frequently have to take into account the beliefs, intentions, desires and preferences of those we live or work with, in our day-to-day routines. When we make plans to eat at a restaurant, we rely on other's people reports on what they want to eat in deciding which restaurant to go to. Doctors rely on the patient's report of whether she is in pain in making their diagnosis. We rely on people's reports on their religious beliefs in determining whether certain conversations would be appropriate. We make long-term plans to accommodate the goals and intentions of other people, and we rely on their reports of what these intentions are. Knowing what it is that other people think plays an indispensible role in our epistemic lives. Without this, we would not be able to co-ordinate our actions with those around us. And, crucially, the primary method of discovering what it is that other people believe, desire, or intend, is to *ask* them.

The point here is that scepticism about self-knowledge does not just threaten our knowledge of our own minds. It also threatens our knowledge of *other people's* minds. And this sort of scepticism, I take it, is untenable. If scepticism about self-knowledge were true, we should expect our reliance on other people's reports on their own states of mind to be exposed as a poor way to find out what they think. Social coordination would surely break down if our primary method of

finding out about other people's states of mind was as unreliable as scepticism about self-knowledge would indicate.

Another way of framing the point as follows: if scepticism about self-knowledge were true then it follows we are not *significantly* more reliable guides to our own minds than anyone else is. But if that were the case, then it would be very difficult to understand how it is that we are able to coordinate with others as effectively as we can. What would cry out for explanation is why it is that this type of scepticism would not have a clear, observable knock-on effect on our day-to-day lives. For every indication is that the assumption that people are reliable guides to their own minds is a cornerstone of social coordination.

The point is that unlike with, say, a global Cartesian scepticism, we should expect self-knowledge scepticism to impact on our daily lives. Reflection strongly indicates that it should have noticeable effects. The fact that such effects are not present, then, is a strong reason to suppose that such a scepticism is not true.

5.4 Immunity from brute error and self-knowledge

This does not, of course, establish that there is a range of self-ascriptions that are self-evident to those who think them. There is conceptual space for a thoroughly epistemic solution to the Problem of Self-Knowledge that denies the self-evidence of self-ascriptions. For example, one such option would be an 'inner sense' view of self-knowledge: a view that accounts for our knowledge of our own thoughts by positing an introspective faculty that operates in the same way that perceptual faculties do.

This would account for the fact that self-ascriptions appear psychologically non-inferred since perception is likewise non-inferred. A follower of this view might account for the LG-authority and -basicness of self-ascriptions by insisting that introspection is a particularly reliable faculty. Beliefs generated via introspection are far more likely to be true than those arrived at via an inference from the person's behaviour.

But yet, on this view, self-ascriptions would be no more self-evident to us than our perceptual beliefs. That is, on this account, self-ascriptions are not *genuinely* groundless in the way 'nothing is red and green all over at the same time' is groundless. Introspective judgements, on this view, are grounded in my introspective experiences: how it introspectively seems to me. It is just that the connection between how it introspectively seems to me and how it in fact is, is so reliably veridical that we do not bother to question it. 'How do you know?' questions are out of place not because self-ascriptions are groundless, but because the answer is obvious: the self-ascriber knows because she has a highly reliable introspective faculty.

To put this in context, this view of self-knowledge holds that self-ascriptions like 'I believe that *p*', 'I am thinking about what to eat for lunch', 'I intend to φ', etc. are not governed by the norm:

(10)   If it is self-evident that *p*, then you may believe that *p* without considering any further evidence for it

But are instead governed by:

(11)   If it introspectively seems to you that *p,* then you may believe that *p*

This inner sense view is an example of what we might call a *partially* vindicating (epistemic) solution to the Problem of Self-Knowledge. That is, it is a genuinely epistemic solution, in that it does think that we *know* our own minds in an epistemically special way, but it does not claim that self-ascriptions possess all three qualities our practices suggest they have. While self-ascriptions really are authoritative and non-inferred, they only appear to be basic. Taking them to be genuinely groundless is a mistake, even if it is an excusable one.

While the most influential arguments against this inner sense view are those offered by Sydney Shoemaker[2], the argument I wish to focus on here is Tyler Burge's.

The rationale for this is as follows: Shoemaker's central argument against the inner sense view is that there are two conditions (the Causal Condition and the Independence Condition) that would have to obtain with regard to our judgements about our own thoughts in order for perception to be a useful epistemological model for 'introspection'.

The idea is that in order for introspection to be operating in anything like the way that perception operates, there would have to be a *causal* connection between our first-order mental states and our beliefs about them. If that is not the case, then Shoemaker thinks that introspection and perception are not similar enough for us to be able to model introspection on perception. And since *causal* connections and *logical* connections are mutually exclusive (that is, if the connection between *x* and *y* is genuinely causal, then it cannot be the case that *x* logically entails *y*) it follows that it must be logically possible for there to be such mental states without us knowing about them, or without there being 'the mechanisms that make such knowledge possible' (Shoemaker 1994b, p. 271). However Shoemaker argues that it is not possible for there to be certain mental states without there also being the mechanisms by which we know about this state.

Even if this argument is successful, however, it is not clear that it conclusively establishes that (11) is not the norm that applies to self-ascriptions. There is conceptual space, I take it, for a *non-perceptual* account to insist that despite self-knowledge being unlike perceptual in those crucial ways that Shoemaker points out, a norm like (11) is still the epistemic norm at play with introspection.

The problem here is that it is not clear what the *normative* implications are of the Causal and Independence Conditions are. Is perception governed by a norm like:

---

[2] See Shoemaker 1994a, 1994b, 1994c.

(4)   If it perceptually seems to you that *p*, then, all things being equal, you

may believe that *p*

because the Independence and Causal Conditions obtain? Would the fact that some set of beliefs are governed by a norm that is structurally similar to (4) indicate that the Causal and Independence Conditions hold for that set of beliefs? This is an interesting and suggestive thought, but I know of no argument in favour of it.

Burge, on the other hand, offers more explicit reasons to think that self-ascriptions are governed by a norm like (10), rather than something like (11).

Essentially, the argument is that introspection differs from perception in that beliefs acquired via introspection are immune from what Burge calls 'brute error' (Burge 1988, p. 120). Burge's point is that it is an integral part of the epistemology of perception that it is possible to have brute perceptual errors, errors that are not due to 'any sort of carelessness, malfunction, or irrationality on our part' (Burge 1988, p. 120). He claims:

The possibility of such errors follows from the fact that no matter

what one's cognitive state is like (so, no matter how rational or well-

functioning one is) one's perceptual states could in individual

instances fail to be veridical—if physical circumstances were

sufficiently unfortunate. (Burge 1998, p. 120)

For instance, even when my eyesight is working properly, and I am looking carefully at my surroundings, there remains a possibility that my judgements about my surroundings are false.

To employ a familiar example, suppose I am visiting zoo and see what I take to be a zebra in one of the cages. Suppose there is nothing wrong with my eyes, I am not under the influence of any narcotics or anything that affects my rationality, and I look closely at the animal in question before concluding that it is a zebra. In

other words, I have done everything that would normally suffice for my knowing that the creature is a zebra. Now if, in this instance, the animal in question is a cleverly disguised mule, then what has happened is that I have been led into making a *brute* perceptual error: an error that is not due to 'any sort of carelessness, malfunction, or irrationality' on my part. My error is to be explained by my having been unlucky enough to have attended a zoo that resorts to this sort of trickery, not in virtue of any epistemic carelessness or cognitive malfunction on my part.

The possibility of this sort of brute, unlucky error follows straightforward from the fact that how it perceptually seems to me is not factive. While perception is remarkably reliable, it is by no means infallible. Even when my perceptual faculties are free of any malfunction, it is still entirely possible that how it perceptually seems to me is not how it is.

Burge goes on to argue that self-knowledge is not like this. For some range of self-ascriptions, these types of errors are simply not possible.

This claim, if true, has some interesting implications. A brute error is to be contrasted with *an error of rationality*. Brute errors are those that may occur no matter how well the relevant cognitive faculty is operating, or how epistemically careful and reasonable one is. If it is not possible for a belief of a certain sort to be brute error, then the only remaining possible error is one that is the result of a *rational failing* on the believer's part. And the relevant notion of rationality here is clearly epistemic rationality. As such, claiming that a belief is immune from brute error is just to say that one cannot have a mistaken belief of that sort without being epistemically irrational.

This has straightforward implications for the relevant epistemic norms at play. If epistemic irrationality —failure to comply with the applicable epistemic norm—is the only way to induce an error in a given type of belief, then it follows that the epistemic norms at play are truth-guaranteeing: that is that the norm is structured such that its antecedent conditions are factive.

In effect, then, what Burge's claim entails is that self-knowledge is governed by an epistemic norm like norm (10). So, if he can offer good reasons to think that introspective reports are immune from brute error, then I take it that this would

suffice to establish that self-ascriptions like *cogito*-like thoughts *are* self-evident to those that think them.

Now consider how this relates to the Problem of Self-Knowledge as I set it up above. I argued that the apparent self-evidence of self-knowledge was a substantial part of problem. To say a belief is basic or groundless as well as non-inferred, I take it, just is to say that the belief in question is self-evident.

Essentially, then, Burge is arguing for what we might call a *fully* vindicating solution to the Problem of Self-Knowledge. I take the view he defends to be the view that self-knowledge really is authoritative, basic and psychologically non-inferred.

I shall make one final point about the implications of the claim that self-knowledge is immune from brute error before moving on to discuss Burge's argument for the claim. Now while it is clear that the claim that self-knowledge is immune to brute error entails that direct self-ascriptions are governed by norm (10), it might also be thought to be saying more than that. One way of reading Burge's claim about brute error is that the suggestion is that anybody who makes a false, direct introspective report is not only irrational but also *blameworthy* for this irrationality.

In the first two chapters I indicated that there are various possible ways in which a person can fail to comply with a norm like (10) without being epistemically blameworthy for that failure. So the suggestion here might be that Burge is arguing that self-knowledge admits of no parallel cases: that people are always to be *blameworthy* for their false self-ascriptions when those self-ascriptions are 'avowed' in the usual non-evidential way. This then would make the claim that self-knowledge is immune from brute error to be significantly stronger than the claim that self-knowledge is governed by norm (10).

While this is an interesting and suggestive claim, it is, I think, unmotivated. As I will illustrate, Burge's argument for this immunity from brute error establishes only that one cannot be *rational* when one erroneously avows some attitude or intention. This is because the considerations Burge appeals in his argument are considerations related to *the rational status* of our beliefs, and of the inferential

162

relations between them. I interpret Burge as endorsing only the weaker thesis that self-knowledge is typically governed by norm (10).

Now the argument Burge gives in favour of the claim that self-knowledge is immune from brute error, interestingly, appeals to the claim that self-knowledge is necessary for a type of reasoning he calls critical reasoning.

Critical reasoning is the reasoning we engage with when we consciously reflect on our reasons for believing some proposition. Non-critical reasoning is the reasoning that occurs when we make an inference from belief *a* to belief *b* without actually thinking about whether the inference is a good inference, or whether *a* really is a good enough reason to warrant belief in *b*. Most of our reasoning, I take it, is like this. Critical reasoning occurs when we deliberately construct —or evaluate— something like a philosophical argument or mathematical proof. The reasoning involved in debating with another person on any subject is critical reasoning.

Now Burge argues that this type of reflective reasoning requires not only that we have self-knowledge, but furthermore that self-knowledge be immune from brute error. Reflective reasoning would not be possible at all unless self-knowledge enjoys this immunity. The idea here is that any account of self-knowledge that denies that certain self-ascriptions are immune from brute error entails:

> [A] dissociation between cognitive review and the thoughts reviewed that is incompatible with the norms of epistemic reasonability that are basic to all critical enquiry, including empirical, mathematical, philosophical, and practical enquiry. (Burge 1996, p. 256)

The central thought here is that the way critical reasoning works is that the results of any enquiry into my own reasons for believing has an immediate effect on my first-order beliefs.

Consider the following example: suppose, after walking past my neighbour's house while deep in thought, I find myself believing that my neighbour is not home. When a friend looking for my neighbour asks how I know this, I recall that I saw that

her car was missing from her driveway when I walked past her house, and offer this as my reason for thinking that she is not home. In the very act of providing this explanation, however, I suddenly remember that her car was taken to the garage this morning and hence that the fact that it is not outside her house is not a good reason to believe that she is not home.

Note the impact that this simple piece of critical reasoning has on my first-order beliefs. We would all expect, I take it, that I would immediately *stop believing* that my neighbour is not home in this instance. And this expectation is a rational expectation. This point is not merely a point about the causal connections between first-order beliefs, and second-order beliefs about them. This is a point about the *normative* connections between first- and second-order beliefs. It would, I take it, be deeply irrational for me to continue to hold the first-order belief that my neighbour is not home, when I hold the following second-order beliefs: first, that this belief is epistemically dependent on my belief that her car is not in her driveway, and, second, that the absence of her car is not a good reason to believe she is not home. Having those second-beliefs entails I rationally ought to immediately stop believing she is not home. Failure to do so, interestingly, is treated like the contravening of an epistemic norm.

So the higher-order critical enquiry into *my own* reasons for believing that my neighbour is home generates rational commitments. It entails that I *ought* to modify my first-order belief.

Why does this entail that self-ascriptions of the sort involved in critical reasoning are immune from brute error?

To see this, it is worthwhile comparing the way we treat the outcomes of our own epistemic reflection on our first-order beliefs, with the way we treat the outcomes of other people's epistemic reflection on our beliefs. The crucial point is that the fact that *someone else* judges that I do not have sufficiently good reasons to believe that my neighbour is home does not immediately entail that I ought to take steps to modify or find additional reasons for my first-order beliefs. The reason for this is straightforward: that somebody thinks my reasons for holding the beliefs I

hold are bad does not entail that they really are. It is quite possible that they are wrong about what I believe, or about what my reasons for that belief are.

Yet, interestingly, we treat the fact that I myself have come to think that my own reasons are bad reasons *as if* it actually entails that my reasons really are bad reasons. That is to say, the fact that *I* now think my reasons are bad is treated as decisive in a way that someone else's opinion is not. Treating my own higher-order reasons this way is an essential part of what it is to reason critically. As a result, for the practice of critical reasoning to be an epistemically *legitimate* way of acquiring and modifying beliefs, it must be the case that we are *right* to treat my higher-order beliefs as decisive in this way.

Essentially, then, when we reason critically, we presuppose that our judgements about the mental states reflected on, and the reasons we have for them, are *self-evident* to us.

So the thought here is that any theory that denies that higher-order judgements of the sort that are involved in reflective, critical reasoning are self-evident from the first-person perspective would be forced to abandon critical reasoning as a legitimate type of reasoning. And this, the reasoning goes, is an unacceptably high a theoretical cost.

Note that Burge has not yet said anything about *what makes* self-ascriptions immune from brute error. All he has done, at this point, is offer reasons to think that they *must be* so immune: on pain of some unpalatable theoretical consequences.

This is an interesting dialectical strategy. The problem we started with here was that it was powerfully intuitive that self-knowledge has a number of puzzling features. Reflection on our pre-theoretical intuitions and practices suggests a strong, widely-held inclination to think of self-knowledge as authoritative, basic and non-inferred.

What Burge has done at this point is effectively to raise the stakes. Not only is it powerfully intuitive that self-knowledge has these properties, but the *assumption* that they do underpins some of our epistemic practices. If we are to deny that self-knowledge has these features, not only do we have to face the

unwelcome prospect of cutting very much against the grain of intuition, but we are also forced to accept the even more unwelcome consequence that a substantive portion of our epistemic practices are entirely misguided.

This argument is structurally similar to the argument I gave in section three. There I argued that a non-vindicating explanation of self-knowledge, an explanation that takes us to be *misguided* in treating self-ascriptions as authoritative, basic and non-inferred, would undermine our knowledge of *other minds* as well, and that this was an implication that would be wholly unacceptable.

Burge's argument is similar in that he is arguing that the denial that self-ascriptions are immune from brute error likewise has unacceptable consequences. An implication of such a view, he argues, is that it undercuts the rationality of a great many of our first-order beliefs. And this he takes to be wholly untenable.

While this does not itself solve the Problem of Self-Knowledge, it does give a clearer idea of where a proper solution must lie. If he is right about this, he has established that a fully vindicating explanation is the only acceptable solution on offer.

5.5 Epistemological explanations

Unfortunately, Burge's actual solution to the Problem of Self-Knowledge is not a good one.

Burge explains this immediate rational link between the outcomes of a 'critical enquiry' and the belief being enquired about as follows: in critical reasoning, he claims, the belief under review and the reviewing process itself operate from 'the same point of view' (Burge 1996, p. 257).

> [M]y checking my belief and finding it wanting normally itself
> provides immediate prima facie reason to change it from within the
> perspective of the review. This is *because* the first – and second-
> order perspectives are the same point of view. (Burge 1996, p. 258,
> my emphasis)

However, as Tom Stoneham (2004) points out, this explanation is more like a restatement of the Problem of Self-Knowledge, than a solution to it. The claim that second-order beliefs are from the same point of view as the corresponding first-order is no less puzzling than the original puzzle an account of self-knowledge is supposed to solve. This is because it is difficult to understand what *makes* the point of view of a second-order judgement the same as the corresponding first-order belief. Stoneham puts point as follows:

> This is a puzzle, because, on the one hand, their form is not sufficient since there can be judgements of the same form which are not so epistemically privileged, and on the other their directness is not sufficient either because there can be direct judgements of other matters … Either they have some other feature, or it is the combination of these two features, form and directness, which explains how come the judgement is necessarily from the same point of view as its subject matter. Until we have such an explanation, the account is incomplete. (Stoneham 2004, p. 658)

Stoneham takes the driving question of an account of self-knowledge to be the question of 'how judgments about our own minds made directly and without recourse to the evidence could constitute knowledge' (Stoneham 2004, p. 658).

Consider the following statements:

(i) When judgements about mental states are made directly and without recourse to the evidence, they constitute a special kind of knowledge

(ii) Judgements about our own minds are made from the same point of view as their subject matter

Burge is offering (ii) to explain how (i) is possible. The problem here is that (ii) no less cries out for explanation than (i) does. More to the point, if we were to articulate what exactly was philosophically puzzling about (ii), we would point to the many of the very same features that we would appeal to in explaining what makes (i) puzzling.

On Burge's view it is possible for people to be 'dissociated' from themselves in the sense that they make second-order judgements that are not from the same perspective as the first-order judgements. A person might, for instance, discover that she harbours racist beliefs by realising that she tends to behave differently towards people who of different ethnic backgrounds to her.

So, the thought is that it is *only* when second-order judgements are made in the usual, direct introspective way that they are from the same point of view as their subject matter. The claim is this:

(iii) When judgements about mental states are made directly and without recourse to the evidence, they have the same point of view as their subject matter

But (iii) is not a clear improvement on (i).

To claim that second-order judgements have a 'point of view' *at all* is to employ a metaphor, and not an especially clear one in this context. Burge vehemently denies that perception is a useful model by which to understand self-knowledge. But yet he appeals to a *visual* metaphor in his explanation of what makes self-knowledge epistemically special. So immediately, it is unclear in what way second-order judgements have a 'point of view' in the first place.

So for Burge's answer to be a satisfactory solution to the Problem of Self-Knowledge he must make it clearer what exactly is involved in two judgements having the same point of view. Without such an explanation, this is not an acceptable solution to the problem.

One way of reading this point would be to see it as making a criticism analogous to Quine's famous attack on analyticity. Quine (1951) argues that the problem with attempts to explicate analyticity in terms of synonymy, definition, interchangeability *salva veritate*, or semantic rules, is that all those terms are no less philosophically puzzling than the term we sought to explain: analyticity. As he says at one point:

> Still there really is no progress. Instead of appealing to an unexplained word 'analytic' we are now appealing to an unexplained phrase 'semantical rule' … Semantical rules are distinguishable, apparently, only by the fact of appearing on a page under the heading 'Semantical Rules'; and this heading is itself then meaningless. (Quine 1951, p. 34)

On this interpretation, Stoneham is pressing the point that Burge's claim that second-order judgements are from the same point of view as their subject matter is likewise 'no progress' towards a solution to the Problem of Self-Knowledge.

However, one important difference is that Stoneham does not take these points to support any kind of scepticism about self-knowledge analogous to Quine's scepticism about the analytic/synthetic distinction.

As I understand the objection, it is not even primarily an objection to the claim that self-ascriptions *are* authoritative, groundless and non-inferred. The claim is only that we have no proper explanation of these features. The problem being raised is a meta-philosophical problem about the kind of explanation a philosopher, or an epistemologist in particular, is supposed to give.

So on the reading I favour, the issue is that it is not enough for Burge to establish that there is a range of second-order beliefs that *are* immune from brute error. In addition to this, he owes us a proper explanation of what it is about these beliefs that gives them this special immunity.

Now I take it that it is not uncommon for philosophers to establish to our satisfaction that something is the case without providing a full explanation of how it

is the case. Suppose I am defending some philosophical thesis—say, realism about *x*. Now one way I might go about defending this view is by offering a *reductio ad absurdum* argument against anti-realism about *x*. If my argument convincingly establishes that any anti-realist position about *x* is committed to some deeply implausibly consequence, then this argument would successfully show *that* realism about *x* is true. I would have established that realism is the only tenable position available, which is of course an excellent reason for thinking that it is true.

But this sort of argument does not provide the last word in the philosophical discussion of *x*. Even if we suppose that I have conclusively established that anti-realism about *x* logically entails a view that is manifestly false, there would still more philosophical work to be done on the topic. This is because this type of argument does not provide us with a substantive realist account of *x*. All it establishes is that we have excellent reasons for thinking that there must be such an account to be had. Just because we are convinced that realism about *x* is true, it does not follow that there is nothing further to be gained by providing a proper explanation of what this realism consists in.

Another way to put this point is that we do not always ask 'How do you know?' because we doubt that our interlocutor *does* know some proposition. It is perfectly legitimate for me to ask an expert ornithologist how she knows that a bird we have both seen is a robin rather than a nightingale. In this scenario, I have no doubt whatsoever that she *does* know this. I am not asking her how she knows because I am looking for reasons to think that the bird *really* is a robin. I take her testimony to be more than adequate reason to believe that it is. I am asking her how she knows because I am epistemologically curious. What I am interested in is what it is about the bird that tipped her off that it was a robin. I want to know what the basis of her knowledge is.

So the suggestion here is that the epistemologist is, or should be, asking the general 'How do we know that *p*?' question in a similar sense.[3] It is not a question

---

[3] Though, to be clear, the questions are not asked in exactly the same sense. One important difference is that the epistemologist is asking for the general conditions for knowledge of some type. When I ask an ornithologist how she knows a bird is a robin, I am not asking for general conditions that make knowledge possible, I am asking for her specific reasons for that judgement. (And, as I

that can be fully answered by pointing out that we *do know* that *p.* We also want to know what the basis of this knowledge is, or what this knowledge consists in. We do not simply want to know *that* we do indeed have knowledge of a certain sort. We are also curious about what makes this knowledge possible: what it is about beliefs of this type that gives them their positive epistemic status.

And it is this sort of question, arguably, that Stoneham is claiming that Burge does not properly answer. The claim that self-ascriptions and their corresponding first-order beliefs are from the same perceptive is simply not a good enough answer to this sort of question.

5.6   A priority as a solution to the Problem of Self-Knowledge

What I want to consider now is whether appealing to the a priori would serve as a better explanation of the special epistemic status of self-knowledge.

What we are considering is the following solution to the Problem of Self-Knowledge: self-ascriptions have the special features that they do in virtue possessing genuinely a priori warrant.

On this account, we simply take it as a brute fact that self-ascriptions are *psychologically non-inferred* and *groundless*. What we can stress, however, is these features are not obstacles to them constituting knowledge. There are psychologically non-inferred a priori judgements. And, on the account of the a priori argued for here, psychologically non-inferred a priori judgements are necessarily *groundless* judgements. Psychologically direct a priori judgements are warranted simply because they are self-evident: the believer requires no further grounds in order to be warranted in believing them.

Finally, the solution goes, self-ascriptions are *authoritative* because our judgements about our own minds are warranted a priori, whereas other people's judgements about our minds are merely a posteriori. Since the epistemic norms

---

argued earlier, it does not look like there are those sorts of specific reasons we can point to with regard to self-knowledge, so this dissimilarity is an important one.) The similarity is just that both questions are asked out of pure intellectual curiosity: they are not interrogative. There is a spirit that is common to both questions, if not a sense.

governing non-inferred a priori judgements are *truth-guaranteeing*, the type of epistemic warrant we have regarding our own thoughts is stronger than the warrant other people have. Treating people as experts on their own minds, then, is eminently sensible. By appealing to the a priori, we can fully account for the puzzling epistemic features that generate the Problem of Self-Knowledge.

Now one advantage that this explanation has over an explanation like Burge's is that it is neither metaphorical nor vague. Saying that self-knowledge has the special epistemic features that it does in virtue of its being a priori is quite different from saying that it has the special features it does in virtue of being from the same point of view as the judgements they are about. The latter is a vague metaphorical claim of limited explanatory value. But the a priori is a legitimate, well-defined epistemological category.

Claiming that self-knowledge is a priori suggests that we can use the paradigmatic cases of direct a priori judgements as a model for self-knowledge. We know what it is that we currently believe, intend, or want in *just the same way* that we know that nothing is both green and red all over at the same time. We are authoritative about such matter precisely because it is only us who have access to our mental states in this way: we are the only ones to whom such judgements are self-evident.

Appealing to the a priori thus provides a model by which we can understand how self-knowledge works. It is thus illuminating in a way that appealing to points of view is not.

At this point however, one might object that it is not been properly explained what it is that *makes* self-ascriptions of this sort self-evident from the first-person perspective. In virtue of what is this case? Or—in other words—*how* is it possible that we have a priori knowledge of our own thoughts?

It is true that the account I am peddling does not provide a substantive answer to that question. Yet it is, I submit, far from obvious that this is a challenge that my account is required to answer.

I have thus far offered some reasons in favour of thinking that a certain range of self-ascriptions—the sort typically involved in critical reasoning—are best

understood as being self-evident. They must be so understood on pain of major upheaval of our epistemic practices. I have further suggested that this self-evidence legitimates the decision to characterise them as belonging to a pre-existing epistemic category: the a priori.

Effectively I have suggested that understanding self-ascriptions as a priori removes the epistemological puzzle surrounding self-knowledge. The question 'how could a type of belief be direct and groundless and yet still be knowledge?' has an easy answer: that type of belief could be a priori.

I take this account to be a sufficiently detailed answer to the general 'How do we know?' question that epistemology is interested in.

Now it is important to acknowledge that this account certainly does not apply to *every* item of knowledge of our own minds. This is because self-knowledge is not epistemically homogeneous.

My discussion of self-knowledge thus far has been admittedly narrow. My attention has been focussed on what I take to be the central cases of self-knowledge: cogito-like thoughts and our judgements about our presently held beliefs, desires and intentions.

However there are a number of ways of knowing about ourselves that look very different to the way we come to know about our minds in the cases I have focussed on. That is, there are ways of acquiring self-knowledge that are not direct, non-inferential or authoritative. A clear example of this sort would be our knowledge of our own character traits. Knowledge of this sort is typically inferential rather than direct. Coming to know whether we are brave or cautious, generous or prudent, is something that requires time and experience. It is not something that is just immediately obvious to us in the way that the thoughts we are currently thinking are obvious to us.

Roughly speaking, there are two broad categories of self-knowledge: self-knowledge that is 'special' and self-knowledge that is mundane in the sense that it is very much like the knowledge we have of other people's minds.

But there are also what we might think of as borderline cases: cases which in some ways resemble the special kind of self-knowledge and in other ways resemble the mundane kind. One example of such is our knowledge of the *causal processes* that lead up to the decisions and beliefs we arrive at. As Nisbett and Wilson point out, people tend to think of themselves as being in an epistemically privileged position with regard to these types of judgement, but there is good psychological evidence that suggests that this confidence is misplaced. Such judgements are assumed to be authoritative, so they are like 'special' self-knowledge in that sense, but self-ascribers are not genuinely in an epistemically privileged position with regard to those judgements. So they are like mundane self-knowledge in that sense.

Another good example of a borderline case is the one Crispin Wright finds in Jane Austin's *Emma*. Wright picks up on a passage where Emma comes to believe herself to love Knightley based on how she finds herself reacting to a friend's declaration of love for him. As Wright puts it:

> [N]ow she realises that she strongly desires that he marry no one but her, and she arrives at this discovery by way of surprise at the strength and colour of her reaction to Harriet's declaration, and by way of a few minutes reflection on that reaction. She is, precisely, not moved to the realisation immediately; it dawns on her as something she first suspects and then recognizes it as true. It explains her reaction to Harriet. (Wright 1998 p. 16)

What this points out is that while Emma is usually very clear (we may assume) about what she desires, this is not *always* the case. Sometimes we might not be able to tell, via the usual introspective methods, that we have the desires that we do.

The same is true, I take it, also of intentions and beliefs: while it is often perfectly clear to us that we have such mental states when we do, for some of our beliefs, desire and intentions, however, even careful introspective scrutiny might

174

fail to reveal them. The second point it raises is that while what we desire is *usually* obvious to us, it is not always. Sometimes, even states that we are normally perfectly authoritative about are unclear to us.

What I want to suggest is the a priori account of self-knowledge is well equipped to deal with a surprisingly wide range of cases like this.

As my earlier discussions of the a priori pointed out, it is possible for people to take certain propositions to be a priori obvious when they are not. In cases like this they *mistake* a non-self-evident proposition for a self-evident one. This mistake leads them to—perhaps blamelessly—fail to comply with the relevant epistemic norm.

In chapters two and four I suggested that one way this might happen was if the believer in question had developed a recalcitrant intuition in virtue of having held some *theoretical* view for a long time. How it intellectually seems to us can be very strongly influenced by our background theoretical beliefs, opinions, inclinations, and so forth. If these intellectual seemings are very strong, we might consider it a priori *obvious* that some claim is true when it really is not. That is to say, false background beliefs and theories can place us in a position in which we systematically misidentify the antecedent conditions of norm (10).

Or, to phrase it the way I did in the last chapter, we might say deviant philosophical theories, folk theories, unwarranted presumptions, and even semantic intuitions can place us in epistemically non-ideal circumstances.

We are thus well equipped to argue that we can sometimes be likewise in epistemically non-ideal conditions with respect to our own minds. For if there can be circumstances that prevent us from recognising analytic propositions as true, then it is hard to see why we would want to deny that they could be analogous circumstances that sometimes prevent us for recognising our own beliefs, desires and so on.

Consider those people who think it is *obvious* that they chose the rightmost item of clothing because it appears to be of higher quality. Now if we take Nisbett and Wilson's explanation of the error seriously then what we can say of them is that their *folk theory* of privileged access has placed them in epistemically non-ideal

position: a position where they mistakenly take the cognitive processes resulting in their decision to be self-evident when they are not. Their over-generalizing about the situations in which they are authoritative has led them to systemically misidentify the antecedent conditions of norm (10). When considering the processes that led up to a decision, they mistaken take the answer to be self-evident to them when it is not.

We could, not implausibly, adopt a similar line with regard to Emma. That is, we could insist that she has been placed in an epistemically non-ideal position. There are a number of possible explanations to choose from. Perhaps Emma has deviant semantic intuitions about love. Or perhaps there is some sort of Freudian repression mechanism or self-deception at play. Perhaps Emma has *inconsistent* beliefs or *conflicting* desires about Knightley. All of these seem to be *prima facie* plausible explanations of how Emma could have been mistaken or unaware of her own feelings towards Knightley. And I see no implausibility in the claim that conditions like these can—like theoretical beliefs and semantic intuitions—place us in an epistemically non-ideal position.

The point here is that while I have focussed primarily on a narrow range of cases of self-knowledge and spent very little time considering the plethora of mundane or borderline cases, the account developed here is flexible enough to apply also to *some* of the more complicated borderline cases. While I think it is important to recognize that these special, central cases of self-knowledge are not entirely representative of self-knowledge as whole, I also think that this account developed here does apply to, and make sense of, a wider range of self-ascriptions than we might have suspected.

5.7   Self-knowledge as a subset of the a priori

Thus far I have argued that taking (some) self-knowledge to be a priori is plausible *as an account of self-knowledge*. What I have not yet discussed, however, is whether taking our knowledge of our own contingent states of mind to be a priori lends itself to a plausible account of the a priori. It is this question I briefly turn to in this final section.

I take it that the most fundamental concern is that what I believe, experience, desire or intend is contingent, whereas the paradigmatic examples of propositions knowable a priori are necessary truths.

The intuition that the a priori is knowledge of necessary truths is a central one, and can be traced to all the way back to Kant. However, it is not clear how seriously we ought to take it, given that Kripke has given widely respected reasons to think that there *are* possible examples contingent propositions that are knowable a priori.

One line of thought might be that this only works because the propositions Kripke had in mind ('the metre stick is one metre long') are very unusual kinds of propositions. Despite its contingency, there is still some sort of 'semantic guarantee' of its truth. Following Gareth Evans, Hawthorne describes cases like this as *superficially contingent*:

> A true sentence is superficially contingent just in case the function from possible worlds to truth-values associated with that sentence reckons it false at some (non-actual) world. A deeply contingent true sentence is one for which there is no semantic guarantee that there actually exists some verifying state of affairs. (Hawthorne 2002, p. 247)

Now a sentence like 'I believe that p' is not like 'the metre stick is one metre long' in that there are no semantic guarantees that it is true in the actual world. It is in this way *deeply contingent.* On what grounds might we deny that we can plausibly have a priori knowledge of deeply contingent truths? Hawthorne suggests (but does not endorse) the following intuitive argument for this, an argument he reconstructs from his reading of Evans:

> Suppose that having understood some sentence *s*, one does not thereby obtain some guarantee of a verifying state of affairs. One will in that case find it perfectly conceivable that the actual world enjoys

a distribution of objects and properties that falsifies *s*. But now, it seems, one will need to do some empirical investigation to figure out whether the actual world is a verifier or a falsifier of *s*. There thus appears to be a straightforward argument against the possibility of deeply contingent a priori knowledge. (Hawthorne 2002, p. 248)

This argument does, I take it, capture the reasoning behind the general suspicion that only very unusual contingent propositions are knowable a priori. However, as Hawthorne notes, this reasoning is only plausible if we already *presuppose* that self-knowledge is a posteriori.

> Some writers include introspective knowledge under the term 'a priori'. But it would be rather a cheap shot at Evans to so define 'a priori' and to then claim that such sentences as 'I have a headache now' can express deeply contingent a priori truths. (Hawthorne 2002, p. 248)

I am in full agreement that it would be a cheap shot to *define* 'a priori' such that self-knowledge counted as a priori and use that as a basis to dispute Evans' claim. However, what I have tried to do here is more than simply stipulate that the a priori includes self-knowledge. I have offered grounds for thinking that there are substantive epistemological commonalities between paradigmatic cases of a priori knowledge and self-knowledge.

I take Hawthorne's impatience with writers who take self-knowledge to count as a priori to indicate a suspicion on his part that taking self-knowledge to be a priori is just to be *pedantic* about how we understand the sense in which the a priori is 'independent of experience': that this way of lining up the distinction simply obfuscates any possible philosophically interesting differences between the categories.

Insofar as one offers, as I have done, a rationale for taking self-knowledge to be a priori, this suspicion should be allayed.

178

Furthermore, there is also an advantage for an account of the a priori that allows that even deeply contingent propositions may count as a priori. Namely, it reinforces the impression that the a priori/a posteriori distinction is fundamentally an *epistemological* distinction: not a modal, semantic, or metaphysical one. We are talking about different types of epistemic warrant here, rather than about propositions with a different modal status, or which are made true by different kinds of features.

5.8 Conclusion

In this chapter I have argued that we have excellent reasons for thinking that it is not only analytic propositions that are self-evident. Interestingly, some propositions about our current beliefs, desires, intentions, and so on are also self-evident. I have also argued that, in virtue of this self-evidence, self-knowledge is best understood as substantively a priori. This presents us with an unconventional but unusually neat and clear answer to the Problem of Self-Knowledge.

If the conclusions of the last two chapters are correct, then self-evidence is *far* from an empty category. Not only is there reason to suspect that those trivially true propositions involving bachelors and vixens count as self-evident, but there is reason to think that more substantive philosophical claims like the Gettier Claim might also plausibly count as self-evident. Moreover, as I have argued in this chapter, a great many commonplace claims about our own beliefs, intentions, desires and emotions will also count as self-evident. Self-evidence is in fact an *abundant* property: a great many of our beliefs possess it.

Seeming-Independence, then, cannot be appropriately criticised for its reliance on the notion of self-evidence.

Conclusion

This concludes the case for Seeming-Independence. In this thesis I have offered reasons to think that Seeming-Independence is a potentially highly fruitful, and not obviously implausible, account of the a priori. All told, Seeming-Independence offers to retain some of the key advantages of the traditional infallibilist account of the a priori, while avoiding the major pitfalls of that view.

The approach I have taken here has been largely exploratory. The aim of this thesis has been to map out an unusual approach to the a priori. What I have tried to suggest, in effect, is that Seeming-Independence, taken together with its accompanying account of epistemic norms, is worth *taking seriously* as a broad theoretical model. I have not tried to offer a completely filled-in theoretical picture. As a result, there remain a number of gaps that would merit further research.

The account of normativity I sketch in chapter one for instance, is still in its early stages of development, and as a result I have tried to keep my options as open as possible with regard to how it is to be understood. I have not argued for nor, I hope, illegitimately presupposed any *precise* view about what the fundamental norm of belief is, or even if there is one fundamental norm or more than one. Consequently, I also have not made clear *exactly* how epistemic norms are related to the fundamental norm or norms of belief, whatever it or they are.

I likewise have not committed myself to a view on precisely what *following* an epistemic norm involves, beyond making a few very general observations about the minimum conditions any account of rule-following must include. These are all topics that merit further attention.

I also have not mounted a thorough defence of epistemic analyticity, nor of the a priori account of self-knowledge that I suggest in chapter five. What I have claimed is that the *initial prospects* of both of those theories are good. Neither account, I argued, has been shown to be obviously unacceptable. And both carry

with them some very welcome explanatory advantages. Nonetheless, these are areas that are deserving of further of closer scrutiny.

More significantly, however, the general approach I have taken to the a priori here could very easily be applied to other epistemological questions.

One immediate implication of the account of epistemic norms I endorse is that it opens up the possibility that some beliefs that we intuitively take to be perfectly justified are better understood as *unjustified but blameless*. I take this to be a very useful dialectical option. My discussion of the a priori was very heavily dependent upon the availability of that distinction.

In chapter two, I began my discussion of the a priori by suggesting that a range of beliefs that have been almost universally assumed to be properly justified, might be merely *blameless* instead. And as a result, I argued, a long-dismissed account of the a priori was worth revisiting.

Now epistemology in general relies quite extensively on our intuitions about when a belief is justified or unjustified, warranted or unwarranted, known or not known. An enormous amount of work has been built on such intuitions. Almost all of the seminal, game-changing epistemological papers include thought experiments designed to draw out intuitions about justification or knowledge.

The suggestion I have made in this thesis is that we must be very careful about relying on such intuitions. General misgivings about the reliability of intuitions aside, we have reason to think that they are not *fine-grained* enough to distinguish warranted beliefs from those that are merely blameless.

This gives us the resources to revisit and resist a number of the thought experiments many have assumed to be decisive.

One such example, touched on in chapter one, is the New Evil Demon Problem that Cohen and others have directed against reliabilism. Proponents of this problem use Descartes' original evil demon thought experiment in a novel way. They use it to illustrate that reliability is not a necessary condition for *epistemic justification*.

They do so precisely by encouraging us to imagine exact duplicates of ourselves, like us in every way except that their perceptual beliefs are systematically deceived by the interference of a Cartesian evil demon. What the thought experiment is intended to establish is not Descartes' sceptical conclusion, but the altogether different view that these doppelgangers are still justified in their perceptual beliefs despite the fact that their perceptual faculties are highly unreliable.

Now, as I remarked in chapter one, this conclusion is by no means compulsory. The account of normativity developed here allows us to offer a plausible explanation of why it seems so plausible that our doppelgangers *are* justified, while simultaneously denying the intuition.

This account of epistemic norms, then, promises to be an extremely useful conceptual resource, which admits of a great many potential applications within epistemology as a whole.

Seeming-Independence also offers some new and potentially useful dialectical options in meta-epistemological discussions.

Seeming-Independence could, I think, potentially provide a rationale for a *unified* foundationalist theory of epistemic warrant. While the account developed here was not *intended* as such, it would nonetheless be congenial to such a theory.

Historically, foundationalism has tended to take either one or the other of the following two types of belief to be foundational. Empiricist foundationalist theories take beliefs about *our own mental states,* particularly beliefs about our own sensory experiences, as the foundations upon which the rest of our beliefs are built. And, traditionally, rationalist foundationalists take that axiomatic role to be played by our non-inferred *a priori beliefs*, the sorts of conceptual, logical and mathematical beliefs that strike us as undeniable or 'obvious'.

As it happens, those are precisely the sorts of beliefs I have focussed upon here. And my treatment of these beliefs also in many respects mirrors the treatment they have received from classic foundationalists. I have argued that beliefs like these enjoy an *infallible* kind of epistemic warrant—which is precisely

the quality that classical foundationalists like Descartes claimed it was necessary for foundational beliefs to have.

So the account I offer here might plausibly be taken to be broadly consistent with a foundationalist approach. However, this account offers foundationalists a new explanatory advantage. I have argued that *both* our beliefs about our own minds and our beliefs about logical, conceptual, and mathematical truths are a priori. They belong, in my view, in the same epistemological category, and they do so *precisely because* both are infallibly justified in the sense that I have specified here.

As a result, the account I offer here could provide a rationale for any foundationalist who wished to insist that *both* these sorts of beliefs can play the same foundational epistemological role.

What is advantageous about this is that, on the standard foundationalist view, beliefs about our own minds have little in common with mathematical, conceptual and logical beliefs, *other* than that they both play a foundational role. This would naturally raise the question: 'what is it about these sets of beliefs that *allows* both of them to play that role?'

The account offered here makes this question much easier for a foundationalist to answer. They may both play this foundational role, one might insist, precisely because they have the same type of epistemic warrant: a priori warrant.

So, interestingly, Seeming-Independence, along with its accompanying account of epistemic normativity, could potentially be useful for any philosophers wishing to mount a meta-epistemological defence of foundationalism. It thus not only opens up new dialectical options by which to revisit and resist epistemological counter-examples, but it opens up new avenues of discussion in the debate between foundationalism and coherentism.

It is, I conclude, potentially a very useful epistemological approach, that admits of a range of possible applications.

References

Bach, Kent 1985: 'A Rationale for Reliabilism'. *The Monist*, 68, pp. 246-65.

Bealer George 1996a: 'A Priori Knowledge and the Scope of Philosophy'. *Philosophical Studies*, 81, pp. 121-42.

—— 1996b: 'A Priori Knowledge: Replies to William Lycan and Ernest Sosa'. *Philosophical Studies*, 81, pp. 163-74.

Boghossian, Paul 1994: 'The Transparency of Mental Content'. *Philosophical Perspectives*, 8, pp. 33-50.

—— 1996: 'Analyticity Reconsidered'. *Nous*, 30, pp. 360-91.

—— 2001: 'How Are Objective Epistemic Reasons Possible?' *Philosophical Studies*, 106, pp. 1-40.

—— 2003a: 'Blind Reasoning'. *Proceedings of the Aristotelian Society Supplementary Volumes,* 77, pp. 225-48.

—— 2003b: 'The Normativity of Content'. *Philosophical Issues*, 13, pp. 31-45.

—— 2003c: 'Epistemic Analyticity: A Defense'. *Grazer Philosophische Studien,* 66, pp. 15-35.

—— 2005: 'Is Meaning Normative?' In Nimtz and Beckermann (eds.) *Philosophy – Science – Scientific Philosophy.* Germany, Paderborn.

—— 2008: 'Epistemic Rules'. *Journal of Philosophy*, 105, pp. 472-500 [page numbers refer to the version found at http://nyu.academia.edu/PaulBoghossian/Papers/179733/Epistemic_Rules]

—— 2009: 'Virtuous intuitions: comments on Lecture 3 of Ernest Sosa's *A Virtue Epistemology*'. *Philosophical Studies*, 144, pp. 111-19.

BonJour, Laurence 1985: *The Structure of Empirical Knowledge*. Cambridge, MA, Harvard University Press.

—— 1998: *In Defense of Pure Reason.* Cambridge, Cambridge University Press.

Brown, Jessica 1995: 'The Incompatibility of Anti-Individualism and Privileged Access'. *Analysis*, 55, pp. 149-156.

Brueckner, Anthony 1992: 'What an Anti-Individualist Knows A Priori'. *Analysis*, 52, pp. 111-118.

Burge, Tyler 1998: 'Individualism and Self-Knowledge'. In Ludlow and Martin (eds.), 1998, *Externalism and Self-Knowledge*, Stanford, CSLI Publications, pp. 111-26. Originally published in *Journal of Philosophy*, 85, pp. 649-63.

—— 1993: 'Content Preservation'. *The Philosophical Review*, 102, pp. 457-88.

—— 1996: 'Our Entitlement to Self-Knowledge'. In Ludlow and Martin (eds.), 1998, *Externalism and Self-Knowledge*, Stanford, CSLI Publications, pp. 239-63. Originally published in *Proceedings of the Aristotelian Society*, 96, pp. 91-116.

—— 2010: *Origins of Objectivity*. Oxford, Clarendon Press.

Clifford, W.K. 1877: 'The Ethics of Belief', in T. Madigan, (ed) 1999, *The Ethics of Belief and Other Essays*. Amherst, MA: Prometheus, pp. 70–96.

Davies, Martin 1998: 'Externalism, Architecturalism, and Epistemic Warrant'. In C. Wright, B.C. Smith and C. Macdonald (eds) *Knowing Our Own Minds,* Oxford, Oxford University Press, pp. 321-61.

—— 2000: 'Externalism and Armchair Knowledge'. In P. Boghossian and C. Peacocke (eds.) *New Essays on the A Priori*, pp. 384-414.

—— 2003: 'The Problem of Armchair Knowledge'. In S. Nuccetelli (ed.) *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge MA, MIT Press, pp. 23-56.

Descartes, Rene, 1641: *Meditations on First Philosophy,* (revised ed.) trans. John Cottingham, Cambridge, Cambridge University Press 1996.

DeRose, Keith 2002: 'Assertion, Knowledge and Context'. *Philosophical Review*, 111, pp. 167-203.

Evans, Gareth, 1982: *The Varieties of Reference*, Oxford, Oxford University Press.

Fricker, Elizabeth, 1998: 'Self-Knowledge: Special Access versus Artefact of Grammar: a Dichotomy Rejected'. In *Knowing our Own Minds*, C. Wright, B. Smith, and C. Macdonald (eds.)Macdonald, Oxford, Clarendon Press.

Gopnik, Alison, 1993. 'How We Know our Own Minds: the Illusion of First-Person Knowledge of Intentionality', Brain and Behavioural Sciences, 16, pp. 1-14.

Grice, H.P. 1989: *Studies in the Way of Words*. Cambridge MA, Harvard University Press.

Grice, H.P. and Peter Strawson 1956: 'In Defence of a Dogma'. *Philosophical Review*, 65, 141-58.

Hawthorne, John 2002: 'Deeply Contingent A Priori Truths'. *Philosophy and Phenomenological Research*, 65, pp.247-69.

——2004: *Knowledge and Lotteries*. Oxford, Clarendon Press.

—— 2007: 'Externalism and A Priority' In Goldberg, Sanford C. (ed.) *Internalism and Externalism in Semantics and Epistemology*, Oxford, Oxford University Press. [my page numbers refer to the online version at http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0008/1160/Exteranlism.pdf]

Jackson, Frank 1998: *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford, Clarendon Press.

James, William 1896: 'The Will to Believe'. In F. Burkhardt *et al.*, (eds.) 1979, *The Will to Believe and Other Essays in Popular Philosophy*. Cambridge MA, Harvard University Press, pp. 291–341.

Kant, Immanuel 1781: *The Critique of Pure Reason*. trans. J. M. D. Meiklejohn, Penn State University Electronic Classic Series, 2010.

Kitcher, Philip 1980: 'A Priori Knowledge'. *The Philosophical Review*, 89, pp. 3-23.

—— 1983: *The Nature of Mathematical Knowledge*. New York, Oxford University Press.

Kripke, Saul 1980: *Naming and Necessity*. Cambridge MA, Harvard University Press.

—— 1982: *Wittgenstein on Rules and Private Language*. Cambridge MA, Harvard University Press.

Lackey, Jennifer 2007: 'Norms of Assertion'. *Nous*, 41, pp. 596-626.

Lycan, William C. 1996: 'Bealer on the Possibility of Philosophical Knowledge' . *Philosophical Studies*, 81, pp. 143-50.

McKinsey, Michael, 1991: 'Anti-Individualism and Privileged Access'. *Analysis*, 51, pp. 9-16.

Nisbett, R., and T.D. Wilson 1977: 'Telling More than We can Know: Verbal Reports on Mental Processes', *Psychological Review*, 84, pp. 231-59.

Peacocke, Christopher 2004: *The Realm of Reason*. Oxford, Oxford University Press.

Pryor, Jim 2006: 'Hyper-Reliability and Apriority'. *Proceedings of the Aristotelian Society*, 106, pp. 327-44.

Quine, Willard Van Orman 1951: 'Two Dogmas of Empiricism'. *The Philosophical Review,* 60, pp. 20-43.

Shoemaker, Sydney 1994a: 'Self-Knowledge and "Inner Sense" Lecture I: The Object Perception Model' in *Philosophy and Phenomenological Research*, 54, pp. 249-69.

—— 1994b: 'Self-Knowledge and "Inner Sense" Lecture II: The Broad Perception Model' in *Philosophy and Phenomenological Research*, 54, pp. 271-90.

—— 1994c: 'Self-Knowledge and "Inner Sense" Lecture III: The Phenomenal Character of Experience' in *Philosophy and Phenomenological Research*, 54, pp. 291-314.

Sosa, Ernest 2007: *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford, Clarendon Press.

Stoneham, Tom 2004: 'Self-Knowledge', in Niiniluoto *et al* (eds.) 'Handbook of Epistemology'. The Netherlands, Kluwer Academic Publishers.

Wallace, R. Jay 1994: *Responsibility and the Moral Sentiments.* Cambridge MA, Harvard University Press.

Warfield, Ted 1992: 'Privileged Self-Knowledge and Externalism are Compatible',
Analysis, 52, pp. 232-37.

Weatherson, Brian 2003: 'What Good are Counter-examples?' *Philosophical Studies*,
115, pp. 1-31.

Wedgwood, Ralph 2002: 'The Aim of Belief'. *Philosophical Perspectives*, 16, pp. 276-
97.

Weinberg, Jonathan, Shaun Nichols and Stephen Stich 2003: 'Metaskepticism:
Meditations in Ethno-Epistemology'. In S. Luper (ed.), *The Skeptics*.
Aldershot, Ashgate Publishing.

Weiner, Matthew 2005: 'Must We Know What We Say?' *Philosophical Review,* 114,
pp. 227-51.

Wikforss, Asa 2010: 'Are there Understanding-Assent links?' *The Baltic International
Yearbook of Cognition, Logic and Communication*, 5, pp. 1-17

Williamson, Timothy 2000: *Knowledge and Its Limits*. Oxford, Oxford University
Press.

—— 2003: 'Understanding and Inference' in *Proceedings of the Aristotelian Society
Supplementary Volumes,* 77, pp. 249-93.

—— 2007: *The Philosophy of Philosophy.* Oxford, Blackwell Publishing.

Wright, Crispin 1985: 'Facts and Certainty' in *Proceeding of the British Academy*, 71,
pp. 429-72.

—— 1998: 'Self-Knowledge: the Wittgensteinian Legacy' in *Knowing our Own
Minds*, C. Wright, B. Smith, and C. Macdonald (eds.) Oxford, Clarendon
Press.

—— 2000: 'Cogency and Question-Begging: Some Reflections on McKinsey's
Paradox and Putnam's Proof' *Philosophical Issues*, 10, pp. 140-63.

—— 2003: 'Some Reflections on the Acquisition of Warrant by Inference'. In S.
Nuccetelli (ed.) *New Essays on Semantic Externalism and Self-Knowledge*
Cambridge MA, MIT Press, pp. 57-78.