Necessary Excuses for Epistemic Self-Trust

Singh, Robert

Awarding institution: King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Necessary Excuses for Epistemic Self-Trust

Robert Singh
King’s College London

This thesis is submitted for the degree of Doctor of Philosophy.
Abstract

We instinctively trust our epistemic faculties. This thesis attempts to explain the reasonability of such self-trust. Such self-trust might be thought to be unjustified because we could have no non-circular evidence in favour of its reasonability; any evidence we had could only have been generated by trusting our faculties. In this thesis I’ll argue that although we are unjustified in trusting our epistemic faculties, we’re excused for doing so. To make that argument I’ll present an account of the nature and normativity of excuses before applying that account to epistemic self-trust.

In the first three chapters I establish the question of self-trust and explain why it poses a genuine problem. In Chapter One I give an account of self-trust in terms of reliance on doxastic practices and higher-order commitment to the truth-conduciveness of those practices. In Chapter Two I consider and reject three ways of vindicating such higher-order commitment when it is construed as justified belief: bootstrapping arguments, a priori arguments from complexity reduction, and practical arguments. In Chapter Three I consider non-doxastic construals of such higher-order commitment, arguing that such non-doxastic attitudes must be sincere if they’re to be epistemically valuable, before presenting an account of attitude sincerity which I argue no attitude can meet without suffering from the same problems as face belief in the truth-conduciveness of our doxastic practices.

In the remaining three chapters I present my answer to the question of self-trust. In Chapter Four I present an account of excuses, arguing that attitudes are excused when they’re formed in accordance with prescriptive or guidance norms without satisfying salient evaluative or constitutive norms. In Chapter Five I present an account of the normativity of excuses, drawing on the account of their nature presented in the previous chapter, arguing that excused attitudes have instrumental deontic force. In Chapter Six I apply this account of the nature and normativity of excuses to epistemic self-trust, arguing that self-trust is a natural consequence of possessing concepts; since concept-possession is necessary, self-trust is excused.
Acknowledgements

Above all I’m grateful to Mark Textor for his insightful, considerate and patient advice in all of our supervisions together. I also want to thank Clayton Littlejohn and Julien Dutant for their thoughtful suggestions and detailed comments which helped with the material on epistemic normativity and excuses in Chapters Four and Five. I’m grateful to Matthew Soteriou for helpful comments which were important for the section on Christopher Peacocke in Chapter Two. This thesis also benefited from supervisions with Mark Van Atten and Maria Rosa Antognazza; I’m grateful for the time they took to read over and comment on work which fed into the final work.

Uriah Kriegel gave important suggestions on the material which forms much of Chapter Three, and I’m grateful for helpful comments from audiences at King’s College London and the Southern Philosophical Forum who heard early drafts of that material. I’m grateful to Oxford University Press for permission to use parts of a forthcoming paper in that chapter.

I’ve discussed parts of this thesis with Asa Mendel, Barnaby O’Callaghan, Fraser Scarlett, Kevin Xu and Nina Van Rooy, and I thank them for their consistently insightful remarks and astute objections.

This thesis was possible as a result of funding through an AHRC studentship, which I gratefully acknowledge.
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>0.1 The Question of Epistemic Self-Trust</td>
<td>7</td>
</tr>
<tr>
<td>0.2 Reid’s Answer</td>
<td>9</td>
</tr>
<tr>
<td>0.3 Foley, Zagzebski and Surviving Self-Criticism</td>
<td>11</td>
</tr>
<tr>
<td>0.4 Alston and Practical Reasons for Self-Trust</td>
<td>16</td>
</tr>
<tr>
<td>0.5 The Hinge Epistemologist’s Answer</td>
<td>18</td>
</tr>
<tr>
<td>0.6 Answering RST2</td>
<td>20</td>
</tr>
<tr>
<td>0.7 Outline of the Thesis</td>
<td>24</td>
</tr>
<tr>
<td>Chapter One  The Nature of Epistemic Self-Trust</td>
<td>26</td>
</tr>
<tr>
<td>1.1 Sharpening the Question</td>
<td>26</td>
</tr>
<tr>
<td>1.1.1 Doxastic Practices</td>
<td>26</td>
</tr>
<tr>
<td>1.1.2 Objections to Doxastic Practices</td>
<td>29</td>
</tr>
<tr>
<td>1.1.3 Some Sharpenings of RST2</td>
<td>34</td>
</tr>
<tr>
<td>1.2 Trust and Reliance</td>
<td>37</td>
</tr>
<tr>
<td>1.3 Truth-conduciveness, Trade-offs and the Value of Engaging in a Practice</td>
<td>43</td>
</tr>
<tr>
<td>Chapter Two  Three Ways of Vindicating Self-Trust</td>
<td>49</td>
</tr>
<tr>
<td>2.1 Reasonability and Justification</td>
<td>49</td>
</tr>
<tr>
<td>2.2 Bootstrapping and Track Record Arguments</td>
<td>51</td>
</tr>
<tr>
<td>2.2.1 Are bootstrapping arguments obviously illegitimate?</td>
<td>51</td>
</tr>
<tr>
<td>2.2.2 Problems with Bootstrapping</td>
<td>54</td>
</tr>
<tr>
<td>2.3 A Priori Arguments</td>
<td>63</td>
</tr>
<tr>
<td>2.3.1 Descartes and the Reidian Challenge</td>
<td>64</td>
</tr>
<tr>
<td>2.3.2 Complexity Reduction</td>
<td>65</td>
</tr>
</tbody>
</table>
2.4 Denying Evidentialism

2.4.1 The Practical Argument for Self-Trust

2.4.2 Threshold Pragmatism

2.4.3 Reasons Pragmatism

2.5 Conclusion

Chapter Three  Non-doxastic Alternatives to Belief

3.1 Introduction

3.2 Sincerity

3.2.1 The Importance of Sincerity

3.2.2 What is Sincerity?

3.3 Sincere Non-Doxastic Commitment

3.3.1 Hope

3.3.2 Acceptance

3.3.3 Assumption in Hinge Epistemology

3.4 Conclusion

Chapter Four  Excuses

4.1 Constitutive Norms, Guidance Norms and Excuses

4.1.1 Introduction

4.1.2 Motivating Excuses

4.1.3 Excuses and Blamelessness

4.1.4 The Derivative Norms Account

4.1.5 The Virtue Theoretic Account

4.2 To be Excused is to Satisfy a Guidance Norm

4.2.1 What is Guidance?

4.2.2 Accounts of Guidance

4.3 Application to Norms of Justification

4.4 Application to Cases

4.5 Objections

4.6 Conclusion

Chapter Five  Excuses, Guidance and Normativity
5.1 Must all norms be guiding?

5.1.1 Some Reasons to Think All Deontic Norms Are Guiding

5.1.2 Reasons to Think Not All Deontic Norms Are Guiding

5.1.3 Arguments for WGN

5.1.4 Arguments against WGN

5.2 Why guidance norms have deontic force

5.2.1 Guidance norms are means to satisfying constitutive norms

5.2.2 Swamping Problems

5.3 Objections
**Introduction**

**0.1 The Question of Epistemic Self-Trust**

We regularly make use of our epistemic faculties: we perceive the piping of the kettle, remember the location of the teabag, trust the testimony of others as to the proper way of brewing a cup of tea, and hence reason that we should retrieve the milk from the fridge. In doing so we employ distinct epistemic faculties of sense perception, memory, testimony and reason. It’s natural to think that when we employ our epistemic faculties we’re thereby committed to the reasonability of relying on those faculties. We can tell this by considering someone who relied on their faculties but denied, or remained neutral, about whether those faculties merited such reliance; such a person seems deeply irrational.

When we have such commitments we’re typically expected to have good reasons for them. There’s a difference between someone who relies on a thermos to keep their soup warm and someone who relies on an ice bucket. Both enter into a commitment to the instrumental value of the thermos or the ice bucket for keeping the soup warm, but the former can appeal to reasons which renders their reliance reasonable, while the latter possesses no good reasons to which they can appeal. So it’s a fair expectation that we have some good reasons for trusting our own faculties. This fair expectation gives rise to two natural questions:

**Reasons for Self-Trust (RST1)** What reasons do we have for our commitment to the reasonability of relying on our epistemic faculties, if any?

**Reasonability of Self-Trust (RST2)** In what way, if any, is our commitment to the reasonability of relying on our epistemic faculties reasonable?

We know what a soup-container must be like for our reliance on it, and our commitment to the reasonableness of that reliance, to be reasonable. A soup-container merits commitment to the reasonability of relying on it just if it’s good at keeping soup warm (and clean, readily accessible and so on).

---

1 I use ‘reasonability’ as a placeholder for whatever the right normative status turns out to be; I don’t want to decide here whether ‘reasonability’ should be cashed out in terms of justification, rationality, rightness or some other normative status.
Similarly, we also have some intuitions about what an epistemic faculty would have to be like to merit commitment to the reasonability of relying on it. Soup-containers are good insofar as they’re good at containing soup; epistemic faculties are good insofar as they’re good at helping us attain epistemic goods. That can help us understand the reasonability of self-trust without deciding just what the epistemic goods are, or if there’s lots of them or just one fundamental epistemic good. Whatever epistemic goods are, to be properly epistemic they must have some tight connection with truth. It looks plausible that an epistemic good must make some contribution towards raising the likelihood of attaining truth, even if that isn’t all there is to being an epistemic good. So to merit trust a faculty has to be *truth-conducive*, in the sense that it has to produce a sufficient ratio of true to false belief.

That places a constraint on a good answer to RST1 and RST2. Just as when we rely on a thermos to keep our soup hot we’re committed to the capacity of that thermos to preserve heat, so too when we rely on our epistemic faculties our higher-order commitment to the reasonability of that reliance has a particular content: we’re committed to those faculties being truth-conducive. If we want to answer RST1 and RST2, then, we need to discover what sorts of reasons we possess, if any, for being committed to the truth-conduciveness of our epistemic faculties.

In this thesis I want to pursue one sort of answer to these questions, the sort of answer which is present in a tradition running from Thomas Reid to William Alston.\(^2\) That answer has two parts. The first is a denial that we have any non-circular reasons for self-trust; if good reasons must be non-circular, then, we lack any good reasons for self-trust. The second is to affirm a positive account of the way in which self-trust is reasonable. On this line of thought, self-trust is permissible because we’re *blameless* for trusting our epistemic faculties, and we’re blameless in so trusting because we don’t have any alternative to doing so.\(^3\) There are some important questions about this sort of answer:

\(^2\) I see this tradition as including not only Alston and Reid but also Hume, at least on interpretations of Hume considered in Qu 2019 p.307, Wilson 2008, Avnur 2016, 2019, Wright 2004a, 2014; for more on Wright and Hume see Williams 2012. A similar argument can be found in Greco 2012. It plausibly also includes Wittgenstein 1969 and the hinge epistemology tradition, for which see Coliva 2012a, 2012b, 2015, Pritchard 2016, Wright 2004a, 2004b.

In what sense do we lack any genuine alternative to self-reliance? Does this mean that we couldn’t, as a matter of psychology, fail to rely on our own faculties? Or does it mean that we would run into undesirable practical consequences if we tried to do so, like falling down holes or spoiling cups of tea? There are lots of ways to cash this out, and it’s not obvious if any of them will be successful or satisfy the doubts or sceptical worries which can motivate the question of epistemic self-trust.

In what sense do we lack any genuine alternative to our higher-order commitment to the reasonability of relying on our own faculties? Is this the same sense in which we have no genuine alternative to relying on our faculties?

What does it mean to be blameless? In what sense, if any, would showing that we’re blameless be a good answer to RST2? What normative status does blamelessness have, and is it plausible that this is the same status as reasonability? Does blamelessness entail epistemic justification? If it does, how does this square with some of the other constraints that we might want to put on being justified – complying with norms which are truth-conducive, for example, and not just making the best possible attempt at so complying? If it doesn’t, can we be satisfied with an answer to RST2 which doesn’t leave us justified in self-trust?

In this introduction I want to situate my answers to these three questions in relation to a number of other views on self-trust which also have plausible answers to those questions. This should make it clear why my position is both novel and plausible.

### 0.2 Reid’s Answer

Here’s how I think Reid would have answered these questions. First, Reid would have said that one way of vindicating our reliance on our faculties is by arguing from our inability to eschew reliance on our faculties. Reid thinks he ought not to attempt to cease believing the deliverances of his senses in particular because:

“[I]t is not in my power: why then should I make a vain attempt? It would be agreeable to fly to the moon, and to make a visit to Jupiter and Saturn; but when I know that Nature has bound me down

---


5 I’ll use ‘vindicating’ to mean ‘explaining the reasonability of’.
by the law of gravitation to this planet which I inhabit, I rest contented, and quietly suffer myself to
be carried along in its orbit. My belief is carried along by perception, as irresistibly as my body by
the earth. And the greatest sceptic will find himself to be in the same condition.”

That Reid implicitly acknowledges that self-mistrust might still, for all this line of argument has to
say, be agreeable, is one indication that the kind of vindication we have for self-reliance leaves us
blameless without being justified. I don’t mean here to import an idea of justification alien to Reid’s
philosophy, but rather to point out that there is a distinction here between what would be reasonable
abstracted from human constraints, and reasonable conditional on human constraints, and that
Reid’s vindication of self-reliance is concerned with the latter.

But the circumstances under which the difficulty of meeting an epistemic standard vindicates us in
failing to meet that standard remain obscure on this view. The laws of gravity which prevent Reid
from flying to the Moon are natural laws which hold as a matter of physical necessity. On Reid’s
view, our basic psychological dispositions also obey natural laws with nomological necessity. It’s
not obvious that nomological necessity is essential for blamelessness; it seems plausible that we can
be blameless for failure to perform feats which are psychologically possible but rarely or never
accomplished. Perhaps there are some foibles which all adult humans possess, but one raised from
birth to avoid them could do so; it’s not clear whether the possibility of being raised in such
different circumstances is enough to guarantee blameworthiness for those not raised in such
circumstances. Or perhaps blameworthiness has nothing to do with species membership, but is a
feature of individuals; perhaps we all have individual foibles which we can’t easily rid ourselves of
even if many of our peers aren’t troubled by them, and we are blameless for such failings while our
peers wouldn’t be. Reid doesn’t give us a theory of blameworthiness, so it’s difficult to decide these
cases on Reid’s behalf.

Moreover, we lack an account of what normative status we have when we’re blameless. Is Reid
being deflationary here, arguing that there’s nothing more to be said in our favour than that it isn’t
worth criticising ourselves for our self-reliance? If so, it looks like he might be conceding more to
the sceptic than we’d like – Reid seems to be conceding that we’re unjustified in our self-reliance,
and asserting further that we’re incorrigibly unjustified. On its face this isn’t very comforting. If
being unjustified is bad, being incorrigibly unjustified sounds even worse. That’s a problem because

---

6 Reid 1997 p.169
7 Wolterstorff 2001 pp.49-50
it's not clear how this can give us a *vindication* of our self-reliance; it isn’t clear what the link between what is reasonable and what we’re able to do is supposed to be.

One way of understanding Reid would have him deny that it’s possible for standards of justification to outstrip what we’re capable of – that is, we would interpret Reid as endorsing an ought-implies-can principle for justification. But there ought to be something concerning in the view that there are no standards by which we can judge ourselves to fall short when we rely on our faculties without sufficient reason. While Reid needn’t deny that there are some such standards, once the existence of such standards is conceded, it is no longer clear that Reid can rule out the view that we have done wrongly or done as we ought not to do if we have failed to meet those standards.

So something further must be said about why Reid has offered a genuine vindication. Unless there’s some connection between what we’re able to do and what we’re reasonable in doing, Reid’s argument seems to miss the target set by RST2.

---

0.3 Foley, Zagzebski and Surviving Self-Criticism

0.3.1 Foley on Intellectual Self-Trust

One general strategy for vindicating self-trust is to adopt a criterion of reasonability such that a subject S is reasonable just in case S has resolved all criticisms which S can level against their own beliefs, where such criticisms can be resolved by either replacing the criticised belief or undermining the grounds for criticism of that belief. Richard Foley and Linda Zagzebski have both pursued strategies along these lines, and have done so in such a way as to place them in the tradition of vindicating self-trust by appeal to what we can and can’t achieve. I’ll now look at each of their accounts and suggest how they would answer the three questions raised above.

On Foley’s view⁸, self-trust is a kind of intellectual faith which we have little choice but to adopt, given the failure of other responses to scepticism. Let’s set aside why those responses fail for the moment; what motivates Foley’s answer to RST2 is what it’s reasonable for us to do given that they do fail. If straight responses to the sceptic fail, we should consider alternative normative standards⁹

---

⁸ Foley 1993, 2005

⁹ This isn’t to say that Foley is recommending revising our normative standards, only that the impossibility of satisfying sceptical demands suggests that our standards were never as high as the sceptic was presupposing.
which accept that we lack non-circular reasons for self-trust but which deny that self-trust is unreasonable.

Foley suggests one such standard which I’ll call Foley-rationality or rationality$_F$. A belief is rational$_F$ just in case it can survive all possible scrutiny which can be levelled against it, just drawing on the resources which are internally available to us: rationality$_F$ is “a matter of having opinions that are capable of standing up to one’s own, most severe scrutiny”.$^{10}$ Your perceptual belief that there’s a purple balloon on the chair is rational$_F$ because it can survive whatever internal criticisms you can level against it; there aren’t any signs of hallucination or illusion, the balloon is clearly discriminable from any similar objects, you’ve heard nothing of anyone trying to trick you by attaching a fake balloon to the chair or plant purple lights near you. Similarly, you’re rational$_F$ in trusting your epistemic faculties, because (1) you naturally form a commitment to the truth-conduciveness of your faculties and (2) that commitment can withstand all criticism levelled against it relying on internal resources alone.

This puts Foley sufficiently within the tradition I’m concerned with that we can think of how he would answer (1)-(3). We have no genuine alternative to self-trust, on Foley’s view, because we’re bound to be in the position of having to trust our own faculties without being able to provide any non-circular reasons for doing so; the difficulty in which we find ourselves is one which is necessary for any epistemic subject, and so we have no genuine alternative in the sense that there isn’t a better epistemic standard which we can live up to than the one that Foley is offering us. That’s Foley’s answer to (2) but similar things could be said to (1). Any given operation of our epistemic faculties, if it survives the internal process of self-critique which Foley has described, is rational just because there’s no real alternative to relying on our faculties under such circumstances; what more could we expect of the rationality of self-reliance than being able to survive such self-critique? Finally, Foley’s answer to (3) is that we’re blameless in the sense that we’re rational$_F$, and that reasonability, our placeholder for whatever normative status attaches to self-trust, turns out to be rationality$_F$.

Note that since the goal of our epistemic faculties, according to Foley, is generating not only accurate but comprehensive beliefs, there is no danger that this account of reasonability will leave us fixed in our beliefs, unwilling to gather any further intellectual resources for fear they will (unbeknownst to us) contradict the beliefs we already have. If that happened, we’d be irrational$_F$, on

---

10 Foley 2001 p.28

12
Foley’s view, so it might look like the best thing for a rational subject to do is avoid collecting any more beliefs, once they’ve successfully become rational once. But to do so would mean failing to implement their epistemic faculties, faculties which have just been secured as rationally trusted, because our epistemic faculties are directed towards forming comprehensive beliefs; to refrain from gathering new material would require some reason, and no reason could be given that didn’t cast doubt on what we have just secured.

As Catherine Elgin points out, this account would declare even someone completely delusional to be reasonable, because their beliefs could withstand all possible internal scrutiny even if they wouldn’t withstand our scrutiny. Even someone not subject to delusions could avoid being appropriately receptive to the world by believing that anything besides some arbitrarily privileged source of knowledge, such as religious scripture, was more open to doubt than that privileged source. If we took scripture as indubitable then our trust in it could survive any internal scrutiny we levelled against it.

Foley would probably say that an account of reasonability which didn’t have such consequences would be one that hasn’t adequately faced up to the failure to provide a straight response to scepticism. In the absence of such a response, there isn’t a strategy that we can adopt that will differentiate us from the seriously deluded or the privilegers of scripture. Elgin is tacitly demanding something which (Foley contends) we can’t have, for just the same reasons as we can’t have an argument against scepticism. We’re in the same position as the deluded person because we can’t help but do our best to be rational with the resources available to us; while we could try to provide a track record argument for the truth-conduciveness of our faculties, the deluded person could do just the same, so there’s no strategy any of us can draw on which differentiates between us which will convince any of us that we aren’t radically deluded. If there were such a strategy, we could use it to differentiate ourselves from the radically deluded, and we would have provided a successful argument against scepticism. In the absence of such an argument, why hold ourselves to a standard of rationality that can never be met? Better, from Foley’s point of view, to be content with a standard of rationality which is achievable.

But there’s a concern underlying Elgin’s arguments which this response wouldn’t touch. Whatever the plausibility of loosening standards of reasonability in the face of the failure to find a straight solution to sceptical arguments, whatever standards of reasonability we arrive at had better be ones

---

11 Elgin 2004
which match the felt prior need for the concept. Before sceptical worries were introduced, we were using the concept of reasonability so as to exclude the delusional. It’s a defect of Foley’s account that he has to give up that use of the concept. I think that what Foley would say is that we shouldn’t put too many burdens on the concept of reasonability – there are lots of epistemic goods, and consequently lots of ways the delusional person may have gone wrong, and we needn’t be concerned to stretch the concept of reasonability to accommodate them. The problem with such a response is that it would seem to abandon the tight connection between what we ought to do and what it’s reasonable to do; Foley would now be conceding that there are more ways we can fail to do what we ought than by being irrational. But in that case introducing Foley’s rationality appears to just be changing the subject.

0.3.2 Zagzebski on Epistemic Self-Trust

Like Foley, Zagzebski supplies a sharpening of reasonability which is intended to render self-trust reasonable. On Zagzebski’s view, to be rational is to do what we do naturally, only better; call rationality in this sense Zagzebski rationality or rationality. Zagzebski claims that we have a natural desire for truth, and a natural expectation that that desire can be satisfied. We naturally trust our epistemic faculties to be good for satisfying that desire. One thing we do naturally is minimise dissonance, where dissonance is some sort of conflict or disharmony between our mental states. It’s not always rational to minimise dissonance – as Zagzebski says, some dissonance, like that between a belief and desire, shouldn’t be reduced by modifying either our belief or our desire, but accepted as the product of all of our faculties acting properly. But often dissonance is a sign of our faculties operating improperly, of a mental state’s not being properly directed towards its object. In such cases we naturally minimise dissonance by modifying one of our mental states. The questions like RST2 can produce such dissonance, because we find ourselves trusting in our faculties but not being able to give a vindication of that trust.

12 Foley makes this point about rationality and knowledge (Foley 2001 pp.21-2) and the point is naturally extended to other epistemic goods.
13 Zagzebski 2012
14 I’ll say more on the natural shortly, but as a first stab we can say that a property or state is natural just in case it is universally possessed pre-theoretically by normal, healthy creatures; natural properties are therefore those we can expect a subject to possess prior to encountering challenges to self-trust.
15 One worry we might have is whether we have a natural desire for truth; a natural desire for knowledge is the more traditional formulation. But I don’t think Zagzebski is committed to the view that we have a natural desire for true belief; according to Zagzebski, what we have is a natural desire to have our questions answered. It might be that acquiring knowledge, not just true belief, is necessary to have had your question answered. Zagzebski needn’t take a view one way or the other on this question.
16 Zagzebski 2016
Zagzebski’s account is similar to Foley’s insofar as it treats reasonability as the result of surviving internal self-critique, although Zagzebski resists Foley’s framing of reasonable epistemic self-trust as a reaction to scepticism. Scepticism is, on Zagzebski’s view, one way of bringing out the phenomenon of self-trust, but self-trust is something that we must have from the beginning of our epistemic lives, long before sceptical questions can even be raised; self-trust is pre-reflective, and scepticism only “forces us to confront the prereflective trust we have in ourselves at the reflective level”\(^\text{17}\). According to Zagzebski, the problem of whether we have any reasons to trust our epistemic faculties is a “problem [that] arises as soon as a person reflects upon her desire for truth and carries reflection upon that desire as far as she can...She feels dissonance when she lacks full reflective justification for her beliefs, and that is a problem even if she does not fear skepticism or even pay any attention to skepticism”\(^\text{18}\)\(^\text{19}\).

Epistemic self-trust is rational\(^\text{2}\) because any state which minimises dissonance will lead to trusting our faculties. We can conceive of someone trying to avoid believing that their faculties are trustworthy while also relying on their faculties, but such a state would be dissonant and it would be rational\(^\text{2}\) to avoid this: “to treat something as deserving of trust without believing it is deserving of trust creates dissonance in the self that becomes noticeable once we reflect upon it. When I become aware that I treat myself as epistemically trustworthy I feel pressure within myself to either believe that I am trustworthy or to stop treating myself as trustworthy”\(^\text{20}\). It’s rational to minimise this dissonance either by refraining from such reliance, or by continuing with reliance together with self-trust, but to accept some middle way would be to invoke dissonance. Here we can see that Zagzebski’s answer to (2) follows from an answer to (1): given that we are rational\(^\text{2}\) in relying on our faculties, it’s rational\(^\text{2}\) to have a higher-order commitment to the rationality of relying on that faculty because the failure to have that higher-order commitment would create unnecessary dissonance.\(^\text{2}\)

\(^{17}\) Zagzebski 2012 p.43
\(^{18}\) Zagzebski 2012 p.41
\(^{19}\) At times, Zagzebski seems to suggest that on Foley’s view the attitude of trust itself is something that only comes about as a ‘fallback’ position, once straight solutions to scepticism are seen to be unsuccessful; this would be a misreading of Foley, because Foley is clear that even pre-theoretically, “Most of us do intellectually trust ourselves by and large. Any remotely normal life requires such trust” (Foley 2001 p.3). The role of philosophical investigation intro trust, and the use of scepticism, is in bringing out the question of whether we’re reasonable in having such trust. That use of scepticism is where the line between Foley and Zagzebski ought to be drawn, not on the question of whether the attitude of trust itself obtains prior to confronting scepticism or not.
\(^{20}\) Zagzebski 2012 p.43
\(^{21}\) Fricker 2016 argues that this argument doesn’t go through unless we have a good account of dissonance-minimisation; since Zagzebski concedes that dissonance must sometimes be accepted rather than eliminated, it’s not clear whether we ought to have self-trust to minimise dissonance or if we should eschew self-trust and accept the resulting dissonance. We can’t answer this without knowing when we ought to accept and when to eliminate dissonance.
We still need to answer (1), and explain why it’s rational to rely on our faculties. Couldn’t we also minimise dissonance by being fully sceptical, and refraining not just from higher-order commitment but also from first-order reliance on our faculties? Zagzebski would say that such a subject wouldn’t be rational because they wouldn’t be doing what we do naturally: we naturally rely on our faculties, and in refraining from such reliance such a subject wouldn’t be doing what we do naturally better, but rather failing to do what we naturally at all.22

It’s not clear how to understand naturalness in this sense. The first thought might be that the natural is the states and properties we come with pre-theoretically, before exercising our executive authorities over our own selves to re-shape ourselves rationally. But as with Foley, this threatens to break the connection between reasonability and what we ought to do; Pritchard and Ryan point out that there is no obvious reason to expect a link between the natural and the normative, given that natural states and properties are the products of whatever heritable epistemic habits have proved to be reproductively beneficial to our ancestors. Why think that there’s any special link between this property and doing as we epistemically ought? Even if we think that there’s more to being pre-theoretically possessed than being a reproductively-beneficial property of one’s ancestors, it’s still not clear how being pre-theoretically possessed should have anything to do with truth-conduciveness; this modified question is just as difficult as Pritchard and Ryan’s original.

Both Foley-rationality and Zagzebski-rationality run into difficulties with answering (3). That’s because they sharpen reasonability in ways which struggle to maintain the connection between reasonability and normativity. In the next two sections I’ll look at two approaches which can maintain that connection.

0.4 Alston and Practical Reasons for Self-Trust

On Alston’s view we have no non-circular epistemic reasons for self-trust.24 Hence we can’t answer RST2 by construing reasonability as epistemic justification; justification requires sufficient epistemic reasons, and circular reasons can never be sufficient. We have no non-circular epistemic

---

22 Zagzebski 2012 p.45
23 Pritchard and Ryan 2014; Longworth 2014 points out that if natural properties are those which are deemed so by our best theories of natural science then it’s not obvious what a priori constraints there could be on naturalness, so there seems little hope of finding any a priori connection between rationality and epistemic normativity.
24 Alston 1991 p.103, p.143
reasons because any attempt to acquire such reasons would involve employing our epistemic faculties, and hence any such reasons be circular.  

Instead of construing reasonability as epistemic justification, Alston construes it as practical rationality. We’re practically rational in believing p when we’re (1) doing as we ought in believing that p while (2) (1) doesn’t obtain in virtue of possessing sufficient reasons speaking in favour of the truth of p. As characterised, practical rationality doesn’t automatically have anything to do with responsiveness to practical reasons; to leave open the relationship between practical rationality and responsiveness to practical reasons, I’ll call practical rationality as it has just been characterised Alston-rationality or rationality A.

Alston’s thought is that we’re rational A in being in a state when there’s no alternative to being in that state. Scepticism isn’t a genuine alternative so there is no alternative to engaging in self-trust, and self-trust is rational A.

Alston is appealing to a general constraint on what we ought not to do: we ought not to φ only if there is a real alternative to φ-ing. Where does this constraint come from? A natural way of interpreting Alston is to take his talk of practical rationality at face value; that would suggest that the constraint comes from having the right responsiveness to practical reasons. This would give a clear answer to (1) and (2): we have no alternative to relying on our epistemic faculties because there are no practical reasons underwriting a switch to a different practice, and we are rational A in higher-order commitment to the rationality of relying on our epistemic faculties because we can come to see clearly the condition in which we find ourselves, the condition in which we have no alternative but to rely on our epistemic faculties. And it would give an answer to (3): the sort of reasonability at play here is responsiveness to reasons, including practical reasons.

Interpreting reasonability as responsiveness to practical reasons risks irrelevance, because the question of self-trust can be raised about the epistemic reasonability of trusting our own faculties. It sounds strange to say that I’m not doing as I epistemically ought to in trusting my epistemic faculties, even if I’m doing as I ought all things considered. Such strangeness might indicate that there are good practical reasons not to engage in self-trust; if engaging in a practice is impermissible by its own lights, doesn’t that give me practical reasons not to engage in that practice? I’ll say more

25 Ibid pp.107-8
26 Ibid. p.168
27 Ibid. p.150, p.168
about this way of narrowing RST2 to epistemic reasonability in Chapter Two. But if Alston-rationality isn’t a matter of responding to practical reasons, it’s not clear what it is. It’s hard to see how it could be responsiveness to reasons of any kind, if it’s neither responsiveness to epistemic nor to practical reasons. But then where does the real alternatives constraint come from?

Alston’s own view later became that any appeal to practical reasons is unnecessary.\(^{28}\) Self-trust is reasonable because there is no alternative to engaging in unjustified self-trust if we’re to engage in epistemic inquiry at all.\(^{29}\) On that view we would have epistemic reasons for self-trust even if we lacked evidence for the truth-conduciveness of our faculties. This is the same answer as that given by hinge epistemology, which I’ll now turn to.

### 0.5 The Hinge Epistemologist’s Answer

The hinge epistemologist says that certain propositions ‘hold fast’ for us because they are necessary as background assumptions for any investigation at all.\(^{30}\) On Crispin Wright’s reading of Wittgenstein, some particular propositions must be held fast because they “play a pivotal role in our methodology of empirical investigation and thereby contribute to the background necessary to make cognitive achievement possible, a background without which the acquisition of knowledge would be frustrated by a lack of regulation”\(^{31}\). To have an empirical investigation at all we need some standards or “principles of appraisal”\(^{32}\) by which we can judge clusters of experiences to be veridical or not. These principles can be encoded in propositions, and assuming these propositions will be necessary if we’re to engage in any inquiry at all. So engaging in epistemic practice requires having commitment to certain propositions.\(^{33}\)

That our faculties are generally trustworthy seems to be one such proposition, since self-trust appears necessary for any inquiry at all. So here we have one sort of answer to (1) and (2): reliance and higher-order commitment are necessary because any creature engaging in an inquiry at all must rely on and trust their faculties.

---

28 Alston 2005 p.220  
29 Ibid. p.221  
31 Wright 2004b p.288  
32 Ibid. p.291  
33 Ibid. p.293: “The overarching idea is, roughly that empirical practice – having a ‘life’ - presupposes un-earned certainties.”
Wright also has an answer to (3) in his distinction between entitlement and justification. Since most of our epistemic lives involve beliefs for which we either have justification, or where it’s appropriate to demand justification if it’s lacking, it’s easy to conflate justification with doing as we epistemically ought – in Wright’s terminology, it’s easy to conflate justification with warrant.

But there is another way of being warranted, which Wright calls entitlement. We’re entitled to assume a proposition when it’s necessary to do so to have an epistemic life at all. Entitlements don’t bear on the truth of such propositions, because we can’t offer any non-circular evidence for them. Still, there is something epistemically reasonable about assuming a proposition to which you’re entitled, and we ought to ensure our understanding of reasonability is sufficiently expansive that it can accommodate this notion of entitlement. The right sharpening of ‘reasonability’ is warrant, where to be warranted is to be either justified or entitled, and the sense in which self-trust is reasonable is the sense in which we have an entitlement to self-trust.

Where does entitlement get its normative force? It’s something like this: when we’re entitled to believe that p, believing that p is epistemically valuable because it enables us to engage in inquiry, inquiry which will lead to further epistemic rewards. Perhaps, then, Wright is endorsing a sort of consequentialist view of epistemic normativity, one on which we could have general reasons to endorse the truth-conduciveness of our faculties prior to any attempt to account for self-trust because doing so will promote further epistemic goods. But that can’t be reasonable in general. Consider:

**Sick Mathematician** A mathematician, Oleg, is more likely to survive a serious illness if he believes he will survive. If Oleg survives, he will prove a variety of important and novel theorems. Oleg believes that he will survive because he wants to secure the epistemic value that would come with proving many new theorems.

---

34 Wright 2014 p.214
35 Ibid.
36 Wright 2004a p.167-8
37 For criticism of the notion of entitlement, see Jenkins 2007.
39 This example is taken from Firth 1981 via Berker 2013.
Intuitively, Oleg isn’t merely unjustified: his belief that he will survive isn’t epistemically reasonable at all. So we aren’t in general reasonable in believing for the sake of acquiring further epistemic goods.

Two responses suggest themselves. We might think that the problem with Oleg is that he believes that he will survive, rather than accepts, assumes or in some way act as though he will survive.\(^{40}\) I’ll look at problems with this response in Chapter Three. We might also think that the problem with Oleg is that he believes that he will survive when doing so isn’t necessary for epistemic inquiry. Oleg could survive without having that belief, so it isn’t necessary for their own inquiry that they do so. The problem with this response is that it doesn’t explain why being necessary for an inquiry confers any normative status. We wanted to explain that by appeal to general epistemic consequentialist principles. If those turn out to be bad principles, we can’t appeal to more narrow principles which only endorse epistemic consequentialism where it’s necessary, because the narrowing of general principles can’t automatically confer any normative status which those general principles don’t already confer. So attempts to sidestep Sick Mathematician risk failing to address the question of the normative status of entitlement.

0.6 Answering RST2

The discussion above has suggested a number of desiderata for a good answer to (1)-(3).

**Correct Extension of Reasonability** We shouldn’t adopt a theory of reasonability, together with a theory of the way in which we can’t help but engage in self-trust, which would entail that very many other states, intuitively unreasonable, would be declared reasonable on the ground that subjects in those states can’t help but be in those states. The problem of delusional subjects was raised by Elgin against Foley, and there are other cases we want to avoid declaring reasonable, such as cognitive biases.

**Minimal Resources** We should attempt to minimise appeals to further concepts, especially concepts which are at least as obscure as that which we want to explain. We want to avoid appealing to unanalysed notions of ‘naturalness’ as Zagzebski does.

\(^{40}\) This would be the approach of non-belief hinge epistemologists like Coliva and Pritchard mentioned in fn.31 above.
Axiological Consequences A good answer to (3) has to explain what would be valuable about being blameless for self-trust, as opposed to being both unjustified and blameworthy. One of the problems we found with Reid was that he presented the fact that we were unable to refrain from self-trust without explaining the axiological significance of that fact.

Deontic Consequences A good answer to (3) also has to explain the deontic significance of the fact that we’re unable to refrain from self-trust. In particular, it has to explain how it can be permissible to ϕ when, plausibly, to be permitted to ϕ just is to be justified in ϕ-ing. In general, a good answer to (3) ought to explain how the permissibility at issue relates to what we ought to do in general.

Relevance A good answer must avoid changing the subject. In particular, it must either explain why the axiological and deontic components of blamelessness are distinctively epistemic (as opposed to practical), or explain why an account of blamelessness can be relevant while denying that blamelessness has any distinctively epistemic value or deontic consequences. We saw that this was a challenge for one way of interpreting Alston, since it wasn’t clear whether an account of blamelessness in terms of practical rationality was an answer to RST2 at all. It was also not clear whether Foley satisfied this desideratum; an account of blamelessness which renders blameless the seriously deluded is arguably not answering the question asked.

Principled Distinctions We saw that one problem with Wright was the risk of positing a normative distinction to account for the reasonability of believing hinge propositions without giving general reasons for upholding the distinction in question. The normative notions at play in an answer to (1)-(3) ought to have wider applicability beyond accounting for self-trust.

Here’s how I want to answer (1)-(3).

(4) There’s no genuine alternative to self-reliance because we’re necessarily bound to rely on our own faculties. That’s because being a candidate for epistemic evaluation at all entails possessing some concepts. Possessing concepts entails either self-reliance, or having akratic attitudes towards our own concepts.41 Some concepts, like tonk, are illegitimate because, if they were legitimate, they would permit transitions between beliefs which generate contradictions.42 By possessing any

---


concepts at all, we have some sort of commitment towards the concepts involved in those attitudes being legitimate. Were we to refrain from such a commitment, we would have an akratic set of attitudes; if pushed we would have to endorse: ‘I entertain that p, but I don’t believe that C is legitimate’ for some proposition p containing C. But ensuring that our concepts are legitimate involves reliance on our faculties. So if we’re to avoid having such akratic pairs of attitudes, we must rely on our faculties. We’re bound to rely on our faculties because our particular nature puts us in a dilemma between relying on our faculties and being epistemically akratic. I disagree with the position found in Reid and early Alston that we have no genuine alternative to self-reliance for any contingent psychological reasons, nor for any practical reasons. But I also disagree with the sort of reasons given by Foley, Zagzebski and Wright, who hold that we’re blameless in self-reliance because self-reliance without non-circular reasons is a situation we necessarily find ourselves in in virtue of engaging in an epistemic life at all. Instead we’re bound as a matter of necessity to incur some epistemic obligations, in virtue of possessing concepts, just through having the potential to engage in epistemic projects, even if that potential is never actualised.

(5) We’re rational in self-trust because the failure to have a higher-order commitment to the reasonability of that reliance would involve akrasia. Epistemic akrasia may not always be unreasonable; non-ideal subjects in particular, it might be thought, should sometimes be epistemically akratic when their first-order evidence departs from their higher-order evidence. In Chapter Six I argue that such level-splitting views about evidence which generate reasonable epistemic akrasia don’t apply with respect to excused self-trust.

(6) We’re blameless in the sense that we have an excuse. Excuses arise from the inability to comply with norms which tell us what we ought ideally to do. Where it’s impossible for a subject S to comply with such a norm as a result of some defect of S’s, we are excused for failing to comply with that norm. Excuses aren’t just reasons not to criticise S, or practical reasons not to worry about an epistemic failing. To be excused is to have complied with a norm which S can comply with and which stands in some appropriate relation to the norm which S ought ideally to comply with. That excuses can play this role is independently plausible; someone who has broken the law inadvertently is excused because they have complied with a norm which it is possible for them to comply with – φ only if you believe φ-ing to be legal, for example – and which stands in the right relation to complying with the ideal norm. The independent use of excuses, and the intuitive links that such independent uses have with attempting to comply with a set of norms, suggest that such appealing to excuses in defence of self-trust wouldn’t be *ad hoc*. 

22
Here’s how (4)-(6) satisfy the six desiderata outlined above:

**Correct Extension for Reasonability** On this view self-trust is excused because we’re necessarily unable to refrain from self-trust. That doesn’t say anything about the normative status of other defects which we don’t possess as a matter of necessity, such as cognitive biases or being massively deluded due to mental illness. To some extent these other defects are excused to the extent that we can’t, as a contingent matter, help but have them. But they seem to be weaker excuses than that for self-trust.  

If a weaker excuse makes a state less reasonable, then delusions and biases will be less reasonable than self-trust.

**Minimal Resources** Excuses have general applicability and we would (I’ll argue) need to appeal to make sense of a range of normative phenomena. This account of the reasonability of self-trust doesn’t require any extra resources than we wouldn’t also need in other areas.

**Axiological Consequences** Whatever is valuable in following ideal epistemic norms, being excused puts us in the best possible position to achieve that good. Following Julia Staffel, I’ll argue in Chapter Four that epistemic excuses still enable us to recoup some of the benefits of doing as we ideally ought; complying with non-ideal norms still means we’ll form many true beliefs.

**Deontic Consequences** The normative status of excuses is derived from the normative status conferred by compliance with ideal norms. I’ll argue in Chapter Five that we ought to comply with non-ideal norms as a means to complying with ideal norms. Such an account of the normative force of a set of norms is vulnerable to a swamping problem and the problem of the value of fulfilling the means when achieving the end is impossible; both vulnerabilities will be discussed in that chapter.

**Relevance** Excuses are clearly rooted in distinctively epistemic concepts; excuses are accounted for as features of non-ideal norms corresponding to ideal norms, where it is specifically ideal *epistemic* performance which is at issue. There is no risk here of bringing in practical reasons to vindicate self-trust.

**Principled Distinctions** The theory of excuses that I’ll present arises from general notions, that of ideal and non-ideal norms, which we have reason to be committed to prior to any investigation of

---

43 I give an account of the relative strength and weakness of excuses in Chapter Six.
self-trust. Excuses in particular have a role in other areas, such as legal and moral contexts, as well as other problems within epistemology.

0.7 Outline of the Thesis

In Chapter One I will get clear about what epistemic faculties are and what it is to be committed to their truth-conduciveness. I’ll do so by following Alston in construing our epistemic faculties as doxastic practices composed of transitions which output beliefs and belief-like attitudes (1.1). I’ll argue that self-trust has two components: reliance on our doxastic practices, and a higher-order commitment to the truth-conduciveness of those practices (1.2-1.3).

In Chapter Two I’ll consider three ways of answering RST2, each of which understand ‘reasonability’ as justification: the argument from bootstrapping (2.2); the a priori argument from complexity-reduction (2.3); and the argument from practical reasons (2.4). I’ll argue that each of these avenues fails and hence the question of epistemic self-trust is a genuinely pressing question – we really do have an unjustified higher-order commitment when we engage in self-trust.

In Chapter Three I argue that the problems raised in Chapter Two persist even when we weaken higher-order commitment from belief to some alternative non-doxastic attitude. I’ll argue that all attitudes with epistemic value must be sincere, arguing that to be sincere requires being held consistently across contexts (3.2). I’ll then consider a range of alternative attitudes to belief, concluding that they either fail to meet the sincerity constraint or face the same problems as were found for belief in Chapter Two (3.3).

In the next three chapters I present my positive answer to RST2. In Chapter Four I give an account of the nature of excuses, arguing that existing alternatives failure to capture the normative dimension of excuses (4.1), before presenting excuses as arising from compliance with guidance norms without successful satisfaction of their corresponding constitutive norms (4.2), arguing for an account of guidance which explains independent phenomena and hence which give an independent basis for positing the existence of guidance norms, before showing that guidance norms give a good account of epistemic excuses (4.4).

In Chapter Five I give an account of the normativity of excuses, arguing that since guidance norms are necessary if constitutive norms are to have deontic force for us (5.1), guidance norms have
instrumental normativity (5.2); they serve as means to the end of satisfying constitutive norms. The account of guidance norms which I presented in Chapter Four is able to handle swamping problems (5.2.2) and so this instrumental account of their deontic force isn’t vulnerable to the central objection to instrumental accounts.

Finally, in Chapter Six I argue that excuses can vary in quality depending on the closeness of the world at which they don’t obtain (6.1), then go on to argue that our excuse for self-trust is the best available to us because it obtains at all worlds at which we’re a candidate for epistemic evaluation; that’s because it obtains whenever we possess concepts, and we’re bound to possess concepts (6.2). Self-reliance is a natural consequence of possessing concepts because it’s necessary to avoid conceptual akrasia: entertaining contents with conceptual contents while suspending judgment about whether those concepts are legitimate (6.3). Since reliance is excused, self-trust is excused if we’re to avoid epistemic akrasia (6.4). So both reliance and our higher-order commitment to the reasonability of that reliance is excused.
Chapter One
The Nature of Epistemic Self-Trust

1.1 Sharpening the Question

1.1.1 Doxastic Practices

Recall RST2:

RST2 In what way, if any, is our commitment to the reasonability of relying on our epistemic faculties reasonable?

In this chapter I’ll explain how I understand RST2. I’ll first present an account of epistemic faculties, before presenting an account of trust in those faculties.\textsuperscript{44}

Faculties are the general capacities which enable us to form and maintain beliefs. Someone with the faculty of memory has a general capacity to form and maintain beliefs \textit{via} memory.\textsuperscript{45} Faculties are supposed to be the sorts of things we possess just in virtue of how we’re constituted. Someone in a deceptive or inhospitable environment can still possess the very same epistemic faculties as someone in a hospitable environment. A bad way of understanding faculties would be this: faculties are the processes, whatever they turn out to be, that lead to your forming the beliefs that you do in fact form. That would include very many processes as faculties which don’t remain constant between different environments. A good account of faculties should have as a consequence that faculties supervene on the constitution of individual subjects.

I will focus on William Alston’s ‘doxastic practices’ framework as a way of clarifying the notion of an epistemic faculty. On Alston’s model, we can think of our standard epistemic faculties – sense perception, memory, reason, self-knowledge and so on – as sets of transitions with both doxastic and non-doxastic inputs and exclusively doxastic outputs; on this view an epistemic faculty is “a

\textsuperscript{44} For some background on faculties and self-trust, see Alston 1985, Plantinga 1988, McCraw 2015, Dormandy 2020.
\textsuperscript{45} Capacities might be blocked, so someone can have this capacity even if, in very many cases, they are unable to exercise their capacity.
system or constellation of belief-forming habits or mechanisms, each realizing a function that yields beliefs with a certain kind of content from inputs of a certain type”\textsuperscript{46}.

A paradigm case of a belief-formation mechanism with doxastic input is the disposition to reason in accordance with \textit{modus ponens}. Take Gurinder, who is conscientious about believing the immediate consequences of her beliefs. Gurinder is disposed to believe q when she believes that if p then q and believes that p. In the language of doxastic practices, the doxastic mechanism Gurinder exhibits takes the belief that if p then q and the belief that p as inputs and outputs a belief that q.

A paradigm case of a doxastic mechanism with non-doxastic input is sense perception. When Yael is in a certain phenomenal state – the state in which it visually seems as though there is an apple before her – Yael is disposed to form a belief in the content of that seeming – that there is an apple before her. Here Yael exhibits a doxastic mechanism with a visual seeming as input and a belief in the content of that seeming as output. If it is plausible that such visual seemings aren’t beliefs\textsuperscript{47}, then it’s plausible that this is a case of a belief-forming mechanism with a non-doxastic input. Following Alston, I’ll call practices involving mechanisms with non-doxastic inputs \textit{generative}.

Doxastic practices can capture the way that the faculties we possess supervene on how we’re constituted by restricting the available inputs. Restrict inputs to just the mental states of a subject. If S and S* differ in a doxastic practice, they differ in at least one doxastic mechanism. So there is at least one mental state M such that S is disposed to form attitude A in M while S* is disposed to form the distinct attitude A*. Different dispositions with respect to one’s own mental states can be explained only by (perhaps minor) differences in constitution.

This is neutral between internalism and externalism about justification and about mental states. Externalists about justification will say that there’s more to a subject’s justificatory status than is determined by their mental states. But it’s the justificatory status which varies between the hospitable and the inhospitable environment, not the doxastic practice: two subjects can have identical doxastic practices but different justificatory status.

Externalists about mental states will say that the inputs of our doxastic practices might have contents which are properly construed along externalist lines. Suppose Yair, a subject in an

\textsuperscript{46} Alston 1991 p.155
\textsuperscript{47} See eg. Tolhurst 1998; for an alternative view see Taylor 2015.
inhospitable environment, has a misleading appearance which is indiscriminable from a veridical experience of an apple. Then the doxastic mechanisms which Yair and Yael exercise are different, because Yair exercises a mechanism with a non-veridical seeming of an apple as an input while Yael exercises a mechanism with a veridical appearance of an apple as input. But Yair and Yael can still share the same doxastic practice, because both subjects have the same mechanisms in their doxastic practices even if different mechanisms are exercised in different environments. So the supervenience of doxastic practice on the constitutions of subjects doesn’t beg the question against externalism.

There might be different doxastic mechanisms from those with visual seemings as inputs involved when Yael forms her belief that there is an apple before her. Some mechanisms might take as inputs states of which Yael isn’t phenomenally conscious. 48 Perhaps some sub-personal process can function as an input here; since sub-personal processes aren’t beliefs, this is also a mechanism with a non-doxastic input. 49 Nor are sub-personal processes access conscious. 50 But whether or not we agree with Alston that S’s belief that p is justified only when it was formed through a mechanism with an internally accessible input 51, that is something that ought to be argued for within the framework of doxastic practices, not an assumption built in from the outset. If the doxastic practices model is to be a fair way of reading RST2, it ought to be as neutral as possible on such questions. We ought provisionally to be prepared to accept sub-personal inputs insofar as they prima facie play some role in a subject’s faculties.

I should note that on Alston’s view, doxastic practices are strictly speaking the “exercise[s]” 52 of such systems of belief-forming mechanisms, not the systems themselves. That suggests that ‘doxastic practice’ isn’t quite the right clarification of ‘epistemic faculties’, because we could have faculties that are never exercised, but not practices which are never exercised. Since I take it for granted that we do, in fact, exercise our faculties – we do form beliefs on the basis of visual experience, memory and so on, even if we’re wrong to do so – then in practice this distinction won’t be of great significance. But for the avoidance of doubt, I take the faculties to be identified with the system of dispositions, not the exercise of those dispositions, and I will generally take practices to be those dispositions, too. If you think it’s essential to a practice that it be exercised, then doxastic

51 Alston 1986a, 1988, Swain 1988
52 Alston 1991 p.165
practices in my sense need not be practices, although in actuality they almost always are, and ‘doxastic practice’ is a label which has been retained only to show the connection with Alston.

Finally, it’s an advantage of the doxastic practices framework that we can accommodate some of our intuitions about nested faculties. We have a faculty which can recognise faces, and that faculty is part of our general ability to form beliefs based on visual perception. The doxastic practices framework can account for this: a faculty is part of another faculty just in case the doxastic practice corresponding to the former has transitions which each occur within the doxastic practice corresponding to the latter. We can also speak of fundamental doxastic practices; a doxastic practice D is fundamental just in case D is part of no faculty except itself.\(^53\)

1.1.2 Objections to Doxastic Practices

1.1.2.1 Non-doxastic Outputs

It might be thought that doxastic practices are too restrictive because they require the outputs of transitions to be beliefs rather than other affirmatory attitudes. Perhaps we should also be concerned with the reasonability of engagement in our acceptance-forming practices, for example.\(^54\) Although I will argue that we ought to be sceptical about such alternative affirmatory attitudes\(^55\), there is no need to write such scepticism into our definition of doxastic practices; it is better to think of doxastic practices as involving transitions to any affirmatory attitude, with belief-formation as a paradigm case. Although I’ll typically talk as though the outputs of transitions are always belief, this should be read as the clumsier ‘belief-like affirmatory attitude’ unless it is clear from the context that such alternatives are being explicitly ruled out.

We should also assume that suspended judgment\(^56\) and disbelief can be the outputs of transitions constituting a doxastic practice; again, although I will talk of transitions outputting belief, I intend also for outputs to include these attitudes. I take it that these, unlike alternative affirmatory attitudes,

\(^53\) It’s tempting to guess that fundamental doxastic practices correspond to basic faculties like perception, memory and so on; I consider this in 1.1.2.3 below.


\(^55\) See Chapter Three, especially 3.3.2.

\(^56\) I take suspended judgment to be a third attitude, one which is distinct from merely refraining from taking a view; I suspend judgment on whether Labour will win the next election, because I have considered the matter and take the view that the evidence neither supports it nor decisively speaks against it, but, until I just now considered it, I refrained from taking a view on whether Labour has an even number of members. For more on suspended judgment see Friedman 2013, Masny 2020, Rosa 2021, Atkins 2017.
are doxastic attitudes, in the sense that they’re governed by the same norms as belief is; disbelief that p is reasonable when belief in \( \sim p \) is reasonable, and suspended judgment is reasonable when neither belief nor disbelief is reasonable.

1.1.2.2 Non-perceptual Practices

Does the input/output structure skew things towards a misleading epistemology for non-perceptual knowledge? It seems likely that at least one of the faculties of reason, testimony-receptivity, self-knowledge and mathematical intuition would be inappropriately construed as analogous to sense perception. Each of these practices seem to lack non-doxastic inputs which have some phenomenal character which has some bearing on the reasonability of forming beliefs on the basis of those inputs. But talk of non-doxastic inputs, explained by appeal to the paradigm of sensory experience, leaves it obscure how we are to understand these other practices if not by analogy with sense perception. Alston, for example, often talks of “memory impression[s]"\(^{57}\) of the non-doxastic input for self-knowledge as being “a current conscious state"\(^{58}\), which risks presenting all generative practices as having their own special non-doxastic inputs, each with a distinctive phenomenal character, so that we would have memory-intuitions, introspection-intuitions and so on. And that isn’t an assumption that we should build into our account of epistemic faculties.

We could say that non-perceptual practices are non-generative, and instead have doxastic states as inputs. But that would again seem to decide too much at this early a stage; it’s controversial whether doxastic states are inputs for the transitions in each of our non-perceptual practices.\(^{59}\)

To address this problem we should adopt weak criteria for being a non-doxastic input. In principle, we ought to allow the obtaining of any mental state to count as a non-doxastic input. We can, then, be undogmatic about the sorts of non-doxastic inputs we might have in non-perceptual generative doxastic practices; even if it’s controversial that there is some introspection-sensation with a distinctive phenomenal character, it’s perfectly neutral to say that we form beliefs about ourselves only under certain mental conditions, and those conditions (whatever they may be) can therefore serve as non-doxastic inputs for introspection.\(^{60}\) In general, so long as we can agree that a good

\(^{57}\) Alston 1991 p.156
\(^{58}\) Ibid.
\(^{59}\) Chudnoff 2013 would deny this, for example.
\(^{60}\) One worry is that introspection might not be formed like this at all, because introspective beliefs might stand in constitutive relations towards their contents. If that’s right, then the epistemology of introspection will be very different to that of other faculties, and requires separate treatment. I’ll limit my focus here to beliefs which don’t
account of our epistemic faculties involves specifying mental circumstances under which we will form a belief and those under which we won’t, we will be able to accommodate all generative practices under a neutral doxastic practices model.

1.1.2.3 Joint Faculties

Some of our transitions appear to function by the simultaneous exercise of two distinct faculties. Suppose Esther sees her friend in the pub and forms the belief that her friend is in the pub; Esther relies here on memory as well as sense perception, because she needs to recognise her friend to believe they are in the pub. Or take Danny, who forms a belief that Arsenal have won their match because Rebecca has told him as much and Danny knows how reliable Rebecca is about Arsenal; Danny here applies both his receptivity to testimony and uses reason to take into account his assessment of Rebecca’s reliability in this domain. Or suppose Sarah comes to believe that she’s undergoing at least three distinct modalities of experience right now; Sarah seems to come to her belief through the joint exercise of her faculties of self-knowledge and mathematical intuition.

In each of these cases, it’s plausible that there’s no intermediate belief – it’s not that Esther forms a belief that there’s a certain person over there who has certain perceptible qualities, then subsequently checks those qualities against her memories of her friend. Danny doesn’t register that Rebecca has asserted \( p \), assign that an initial credence, and then adjust up or down based on his reasoning about Rebecca’s reliability. Sarah doesn’t form a belief that she has this modality and these modalities and then count the modalities of her experiences, she apprehends immediately that she’s undergoing at least three different modalities of experience. When Esther, Danny and Sarah form their beliefs they do so through some transition, but that transition doesn’t appear to be a member of any traditional doxastic practice. So the doxastic practices framework can’t explain which faculty Esther, Danny and Sarah employ when they form their beliefs; the only answer that framework can supply is to assign a doxastic practice to each faculty, but since that’s impossible in these cases, the doxastic practices framework has no good account to offer.

The doxastic practices framework appears to fail here because it’s tempting, on that framework, to treat the body of our practices as partitioned into traditional practices like sense perception, memory, reason and so on. On this tempting view, each partitioned practice would take the outputs stand in these constitutive relations to their contents; I take it that most beliefs and most faculties aren’t like introspection in this respect. For more on this debate see Coliva 2009, 2016, McGlynn 2019.
from the others as inputs, but there would be no beliefs which aren’t formed iteratively, on the basis of belief formation which at any given stage is solely governed by a single partitioned practice. And that tempting view looks wrong, because Esther’s belief that her friend is in the pub wasn’t formed iteratively in this way.

We should distinguish between the necessity of having some partitions or other, and the temptation to identify those partitions with the particular epistemic faculties found in ‘folk’ epistemology. We can’t be anti-realists about partitions of transitions, resting content with framing RST2 in terms of the total body of transitions, and treating talk of sense perception, memory and so on as convenient shorthand for particular regions of the space of transitions. We can’t do that because we need to be able to explain why it’s right for someone with normally functioning sensory faculties to trust those faculties, even if they also have a body of systematically unreliable transitions elsewhere among their doxastic practices.

But even if some partitions are necessary, they need not be those present in folk epistemology. We should be open to revisionist construals of doxastic practices, such that rather than sense perception, memory and so on we might have quite different ways of partitioning our doxastic transitions. Such revisionist partitions might emerge because in partitioning our practices we’re trying to get at what makes those practices worthwhile, and we need not be deferential to tradition in doing so. What makes those practices worthwhile, if anything does, is their role in securing a special sort of good – that is, epistemic value.61 What’s interesting about doxastic practices – and what makes them useful clarifications of epistemic faculties – is that they can explain why you ought to trust a given transition: you ought to trust a given transition because you trust the practice as a whole. And it only makes sense to trust the practice as a whole if the transitions constituting that practice are roughly homogeneous with respect to their capacity for securing epistemic value. Alston makes the same point, although he here assumes that the epistemic value secured is truth: “From an epistemic point of view, we want to work with practices that are as homogeneous as possible with respect to reliability...For we want practices to be subject to epistemic evaluation in terms of reliability; and unless the constituent mechanisms of a practice are close in degree of reliability, the practice as a whole cannot be judged markedly reliable or unreliable”62. We could swap out reliability, parsed as high likelihood of securing truth, and replace truth with whatever we think is the fundamental epistemic value – knowledge, justified belief or whatever – and Alston’s point would still go

61 I discuss this at greater length in 1.1.3.
62 Alston 1991 p.166

32
through. Let a practice be truth-conducive if it produces a high ratio of true to false belief, knowledge-conducive if produces a high ratio of knowledge to ignorance and so on. Then we can cluster doxastic practices by truth-conduciveness, knowledge-conduciveness or whatever value we care about while avoiding the problem of joint faculties.

Talk of likelihood might introduce a further problem here. Let \( M_1, M_2 \) and \( M_3 \) be transitions in a proposed doxastic practice \( D \), and let \( p_i \) be the proposition that a given transition \( M_i \) is reasonably engaged in. Then on the above view, we ought to construct doxastic practices such that \( P(p_i \mid p_j) \), for any \( 0<i!=j<4 \), is sufficiently high, and such that \( P(p_i \mid p_j \& p_k) \) is in general not much bigger than \( P(p_i \mid p_j) \).

Here’s the problem: if we’re formulating partitions of doxastic practices as a way of understanding RST2, how can we help ourselves to the detailed knowledge about conditional probabilities within a given practice? Aren’t we in danger of explaining why we’re reasonable to trust our epistemic faculties by appealing to detailed empirical information about the common structure behind perceptual judgments that this apple is red and this pear is green? No, because the suggestion isn’t that anyone engaging in a doxastic practice ought to have such detailed knowledge, but rather that we ought to theorise the sort of practices they’re engaged in according to such knowledge. Moreover, epistemic subjects can still exploit the fact that folk epistemological and revisionist views of practices are likely to overlap in extension; epistemic subjects can commit to the trustworthiness of a transition on the basis that it belongs to an extension of a folk epistemological concept, an extension shared with a revisionist concept which can itself explain why there would be homogeneity of trustworthiness among such transitions.

Note that even on this revisionist conception of doxastic practices, it seems unlikely that there will be any dramatic changes in the extension of what we identify as our fundamental practices. Our habit of clustering visually-based transitions under one practice – that of visual perception – is rooted in those transitions sharing a common causal structure, and we ought to expect such similarities of structure to correlate with success in securing epistemic value.\(^{63}\) Perhaps there’s always a reliance on memory in our visual perceivings; if so, the commonality of structure behind our visual transitions will differ from the folk epistemological view, and we’d do better to speak of the fundamental doxastic practice here as being strictly ‘visual perception in conjunction with

\(^{63}\) Note that Yair will have the same clusters as Yael, because for any two transitions within one of Yael’s clusters, their truth-conduciveness is still highly correlated for Yair, even if in Yair’s inhospitable environment neither are in fact truth-conducive.
memory' rather than ‘visual perception’. But there would be no revisions to the extension of the practice here, and the assumption that members of that extension have a common structure which explain their homogeneous success in securing epistemic value would be vindicated.

The transitions which Esther, Danny and Sarah undergo in forming their beliefs will cluster into some doxastic practice or other with transitions which share similar likelihood of being truth-conducive. Most likely those transitions will share a common causal structure, but we don’t need to know anything about that causal structure to be confident that they will cluster in some way. So the doxastic practices framework can accommodate joint faculties.

1.1.3 Some Sharpenings of RST2

Doxastic practices are a sensible way of reading epistemic faculties as they appear in RST2 because they capture what is troublesome about that question. They capture what is troublesome about SQ because they codify the question in the following form:

**RSTDP** Why is it reasonable to think that a high enough proportion of transitions in our doxastic practices are reasonable, such that continued engagement in that practice is reasonable?\(^{64}\)

As noted in the Introduction, I’m using ‘reasonable’ as a placeholder for whatever the right normative concept to plug into RSTDP is. I don’t mean to suggest here that that ‘reasonable’ ought to have the same substitution at every place in RSTDP; perhaps the first ‘reasonable’ ought to be ‘permitted by norms of justification’ and the second ‘obligatory on norms of justification’. RSTDP isn’t intended to decide such questions at this stage.

RSTDP should be readily interpretable by any suitable alternative way of capturing the question, since any suitable alternative will have to express the idea that we have belief-forming capacities which operate on the basis of prior doxastic and non-doxastic states, and that the reasonability of

---

\(^{64}\) The intended reading of RSTDP is that we think that our continued engagement in our doxastic practices is reasonable because a high enough proportion of the transitions composing those practices are reasonable; the intended reading is not that continued engagement is those practices is reasonable because we’re reasonable to think that a high enough proportion of transitions are permitted. RSTDP asks why the former is the right stance to take.
the exercise of those capacities is open to question; such capacities will be readily-interpretable\textsuperscript{65} using talk of transitions, so long as they are suitably determinate.\textsuperscript{66}

RSTDP sharpens RST2 successfully because we have a grip on the general notion of the reasonability of continued engagement in a practice.\textsuperscript{67} Practices can be reasonably engaged in when it is compatible with and helpful for achieving some good; it is reasonable to engage in the practice of carpentry when doing so achieves some good, such as providing you with furniture. Engagement in a practice will typically only be reasonable if engaging in that practice will secure the goods aimed at by that practice. Outside of exceptional circumstances, it’s only worthwhile engaging in the practice of carpentry if doing so produces chairs, tables or whatever. Similarly, it’s only worthwhile engaging in our doxastic practices if doing so secures the sort of good that those practices aim at. Doxastic practices are epistemic practices, so it’s plausible that doxastic practices are aimed at securing epistemic value. If you think that the best way to secure epistemic value is to hold true beliefs, then RSTDP also lends itself to a further sharpening, spelling out reasonable transitions in terms of promoting true belief:

\textbf{RSTT} Why is it reasonable to think that a high enough proportion of transitions in our doxastic practices produce true belief, such that continued engagement in that practice is reasonable?

You might think that the goal of our doxastic practices isn’t producing true belief but \textit{justified} belief. When we engage in a doxastic practice we want to do well, from an epistemic point of view. Doing well, epistemically, isn’t exhausted by having true beliefs, because we can have unjustified true beliefs.\textsuperscript{68} And, perhaps, it isn’t necessary to have true beliefs.\textsuperscript{69} Perhaps someone who believes just what their evidence supports has discharged all their epistemic obligations, but if it’s possible to

\begin{itemize}
\item \textsuperscript{65} In the case of non-doxastic inputs, we’re not interested in merely those cases where the content of that input is true. We don’t want to say that sense perception is reliable just in case it reliably outputs true beliefs so long as the content of the sensory experience is also true. Not only does this controversially assume that sensory experience always has a truth-bearing content, but it would make sense perception vacuously reliable even in sceptical scenarios. For generative practices, what’s important is that the practice in fact does generate a sufficiently high proportion of true beliefs, regardless of the truth of the content of the input.
\item \textsuperscript{66} Zalabardo 2012 p.2 talks about cognitive devices rather than faculties or doxastic practices; cognitive device-talk ought to be interpretable in terms of doxastic practices-talk to the extent that our cognitive devices have inputs and outputs along the lines discussed for doxastic practices.
\item \textsuperscript{67} Note that although doxastic practices need not be practices on my terminology, once a doxastic practice is engaged in then it certainly is a practice; hence we can still fruitfully apply our grasp of the general notion of a practice to the case at hand. I take it we have a general grip on the reasonability of a practice in part because we very often know the right normative concept to substitute for ‘reasonability’in particular cases, not that we have a general grip on reasonability; since it’s a placeholder concept there’s little chance we could have such a general grip on it.
\item \textsuperscript{68} Although for dissenting views see Sutton 2007, Littlejohn 2012, Steglich-Petersen 2013a.
\item \textsuperscript{69} For defences of this position, see Eder 2019 and, from a pluralist perspective, Perrine 2020. For consideration (and rejection) of this point of view, see Gibbons 2013.
\end{itemize}
believe just what your evidence supports without believing what’s true then discharging your epistemic obligations doesn’t guarantee truth.

A natural candidate for epistemic success in this sense is *justification*.\(^{70}\) I’ll assume that there are both substantive and structural constraints on justification.\(^{71}\) Structural constraints on justification forbid certain combinations of attitudes; paradigmatically, they forbid akratic attitudes like ‘p and I ought not to believe p’. Substantive constraints on justification forbid believing p without sufficient reason to believe p; paradigmatically, they forbid believing without sufficient evidence. I won’t assume that evidence is the only reason that can bear on having a doxastic attitude, although I’ll argue for this in Chapter Two.\(^{72}\) But I will assume that evidence is probability raising.\(^{73}\)

If holding justified belief is the best way to secure epistemic value, RSTDP ought to be sharpened to:

\[
\textbf{RSTJ} \quad \text{Why is it reasonable to think that a high enough proportion of transitions in our doxastic practices are justified, such that continued engagement in that practice is reasonable?}
\]

Because justification seems related to securing true belief, if the fundamental epistemic value is true belief, RSTT and RSTJ are both *prima facie* plausible ways of sharpening RSTDP. But our view of the correct sharpening will vary depending on what we think of as the fundamental epistemic value. If you think it’s knowledge, then the better sharpening will be:

\[
\text{RSTK} \quad \text{Why is it reasonable to think that a high enough proportion of transitions in our doxastic practices are justified such that continued engagement in that practice is reasonable, and specifically that a proportion of the transitions are true beliefs?}
\]

---

\(^{70}\) Sometimes ‘justification’ and ‘rationality’ are used interchangeably, and sometimes they’re distinguished. I’ll use ‘justification’ to refer to compliance with substantive and structural epistemic norms, paradigmatically compliance with evidentialist and enkratic norms. Sometimes these constraints are applied to rationality; I’ll use ‘justification’ such that it refers to rationality in this sense.

\(^{71}\) The motivation for the distinction is the distinction between cases where we have reasons to take an attitude towards a proposition, and cases where we ought or ought not to have combinations of attitudes. For more on the distinction between substantive and structural constraints on justification, see Worsnip 2018, 2019, Bradley 2021.

\(^{72}\) I defend a limited form of evidentialism in 2.4.

\(^{73}\) In particular, I’ll assume that if evidentialism captures the substantive constraint on justified belief, then for a belief to be justified is for it to be probabilistically coherent, given a subject’s evidence and other beliefs. To be probabilistically coherent at least involves satisfying the Kolmogorov axioms of probability theory, as well as Lewis’ Principal Principle (Lewis 1980, Bovens and Hartman 2003, Hartman and Sprenger 2010, Williamson 2010, Strevens 2004, Weisberg 2013). I’ll leave open which further principles, if any, S’s credences need to satisfy. There are lots of problems which a defence of this view of evidence would have to address, but although some of them may have good answers once excuses have been introduced – such as the problem of logical omniscience – I won’t be offering a direct defence of this view. On this view E is evidence for p if P(p|E)>P(p). Not all understandings of being probability raising need be committed to this background Bayesian picture of justification. There are other plausible views of evidence which aren’t committed to the probability raising view: evidence for p might be what p best explains (Conee and Feldman 1985, 2004); perhaps also our evidence must be what we know (Williamson 1997); see also McCain 2018. In general the distinctions between these views won’t matter for my argument, but where I need to pick a particular account of evidence I will be assuming the Bayesian understanding of evidence.
RSTK Why is it reasonable to think that a high enough proportion of transitions in our doxastic practices produce knowledge, such that continued engagement in that practice is reasonable?

In the next section I will look at how we might decide between these sharpenings.

1.2 Trust and Reliance

I will argue that trusting in our epistemic faculties involves two commitments. First, I’ll argue for:

Reliance If we trust an epistemic faculty, then we’re disposed to engage in the doxastic practice(s) corresponding to that faculty.

That is, if you trust your faculty of memory, then you tend to engage in the memorial doxastic practice; when there’s an applicable transition in that practice, you typically will have the belief which is the output of that transition. That doesn’t mean you’ll always do so, because beliefs outputted by one practice can be defeated by those outputted by another. But if a doxastic practice would incline us to believe p on the basis of some input, and there is no defeater available to us, then we do believe p. Conversely, where we don’t trust a doxastic practice, we don’t typically have the beliefs which are the outputs of its transitions (unless it is an output shared with transitions in a doxastic practice which we do trust). Most of us don’t trust the doxastic practice of astrology, because we don’t believe many of its typical outputs – we don’t, on learning that the stars are in such-and-such an alignment, come to believe that it is a propitious day for such-and-such an activity.

Reliance seems true because it explains some obvious features of trust. If you regularly refuse to believe that there is an apple before you when it visually seems that there is an apple before you and with no defeating evidence, then you don’t trust your faculty of visual perception, however much you may insist in general terms that you believe the judgments and transitions which fall under that faculty to be rightly engaged in. That’s an argument about the ordinary meaning of ‘trust’. But there’s a substantial point there: if you affirm the reasonability of engaging in the practice of visual perception, but fail to rely on that faculty, then you’re being unreasonable. You’re being
unreasonable because you’re being akratic; you endorse the reasonability of holding that p (that there’s an apple before you, say), but fail to endorse p.74

There is potential for ambiguity in the notion of reasonability that I should mention here. I’ve said that RST2 is concerned with the reasonability of trusting our faculties, in the sense that RST2 is concerned with the way in which in trusting our faculties we’re doing as we ought, however we ought to understand ‘ought’ here. In general, all it takes to do as we ought is to do just what we’re permitted to do. But in the above argument from akrasia, I’ve treated ‘reasonable’ as though it meant obligatory, and that seems to be trading on the ambiguity of the term. But because doxastic practices output beliefs, there’s an implicit uniqueness claim when we claim that engaging in a practice is reasonable. Suppose S judges that a practice D1 is reasonably engaged in, and that, were S to rely on D1, S would believe p, while in fact S relies on D2, and as a result disbelieves p. Then whatever the fundamental epistemic goal is, from S’s point of view anyone engaging in D1 is doing a worse job of securing that goal than they could be – they could replace whatever transitions in D1 which generate belief that p with the transitions in D2 which generate disbelief that p.

Can such a swap always be achieved? Perhaps to preserve consistency there will no simple exchange of D1 transitions bearing on p to D2 transitions; the transitions will be so enmeshed with others in D1 that many others would have to be changed if the resulting practice is to be coherent. But some such changes will surely be possible – after all, with sufficient changes we could bring it about that D1 is identical with D2. Perhaps we need not go that far, but from the point of view of one engaged in D2, any consistency-preserving transition swaps which bring D1 closer to D2 will make engagement in D1 more valuable.

We can’t refrain from engaging in a doxastic practice governing p once the question of p is raised; the attempt to do so will only result in a practice consisting of transitions with any number of inputs and the single output of suspending judgment in p. So anyone judging that a doxastic practice D1 is reasonably engaged in but refraining from engaging in that practice themselves will be making that

Sometimes such akrasia may be reasonable. If Owain has been given misleading evidence about the reasonability of the baby-killing practice, then it might be reasonable for Owain to think the baby-killing practice is a good one without it being reasonable to kill any babies. If that’s not right, then we must either think that Owain can never have misleading evidence about the reasonability of a practice, or that having misleading evidence can never make it reasonable to believe that the baby-killing practice is reasonable, or that it can be reasonable to kill babies if you reasonably believe killing babies to be good. Since these all seem implausible, akrasia must be reasonable. But even if that’s right, there’s clearly something wrong with akrasia such that in many cases it would be deeply unreasonable to be akratic. Our position with respect to our epistemic faculties doesn’t seem much like Owain’s with respect to the baby-killing practice, so we should still assume akrasia would be unreasonable in our case.

74
judgment from the point of view of a competitor practice D₂, and will judge that D₂ is superior to D₁. And when the question is whether to believe or disbelieve p, falling short of a competitor is doing something impermissible; if D₂ is doing well in generating a disbelieve that p, then D₁ can’t also be doing well by generating a belief that p. In this respect doxastic practices aren’t like carpentry; you can judge that someone is reasonably engaged in such a practice while refraining from engaging in carpentry. But we can’t so refrain from engaging in doxastic practices⁷⁵, and if we judge that one doxastic practice is reasonably engaged in, we must judge that competitors aren’t rightly engaged in, so we can never judge that another practice is rightly engaged in without engaging in it ourselves – or at least, we can never do so without being akratic.

The sort of subject I have in mind here is one who engages in some doxastic practices; such a subject must, once the question of p is raised, have some view or other on p, even if that is a matter of suspending judgment. There is another option: a subject who refrains from engagement in any doxastic practice because they refrain from taking any attitude towards p, even suspending judgment; when the question of p is raised they refuse to take any view at all. But that isn’t the sort of subject that can be in question here, since the sort of subject we’re interested in is one who is prepared to make judgments about the reasonability of engaging in certain practices, and who hence must be engaged in at least one doxastic practice already.

It might be that we have no obligations to believe at all⁷⁶; all we have is an obligation not to believe if our belief would fail to meet certain conditions. If that were the case, it might be permissible to refrain from taking a view on p, because there is never an epistemic obligation to take a view on p. That might be right, but even on such a view it’s open whether it’s permissible to: refrain from believing that p and believe that p is reasonably believed. Perhaps such a combination is among those forbidden. And the akratic character of such a position suggests that it should be among those forbidden. So while I’m open to the view that we have no obligations to believe, that doesn’t mean we have no obligations to believe conditional on our already having other beliefs.

The first-order attitudes which we form when we engage in reliance are sometimes taken to characterise a minimal sort of trust. In this sense we trust a shelf to hold books without collapsing when we place books on it, we trust others to tell us the truth when we believe what they say, and

---

⁷⁵ I don’t mean here that we can’t refrain from reliance on our ordinary doxastic practices – that’s to be proved – but that we must at least engage in the minimal doxastic practice of suspending judgment on all propositions for any input.

we trust an epistemic faculty to deliver reasonable outputs when we form the beliefs which are the outputs of the transitions constituting the practice associated with that faculty.

Animals can rely on certain faculties without forming any higher-order judgments about their own reliance, because they are (I assume) unable to form any such higher-order judgments at all. But we are able to form such judgments, and hence a failure to form such a judgment when we’re relying on a faculty leaves us vulnerable to the charge of being akratic. I’ll therefore argue for:

**Higher-order Commitment (HOC)** In trusting an epistemic faculty, we are committed to the reasonability of relying on that faculty.

HOC says that in holding beliefs because they’re the outputs of a doxastic practice we enter into a higher-order commitment to the reasonability of engaging in that practice. I also want to argue for

**Higher-order Commitment to Truth-conduciveness (HOCTC)** In trusting an epistemic faculty, we are committed to that faculty’s being conducive towards forming true commitments.

I will say more about what it means to have a commitment in Chapter Three; for now, it’s enough to rely on the intuition that you couldn’t fully trust in a practice, even if you engaged in it, were you disposed to deny that such engagement is conducive towards the relevant ends which engaging in that practice is supposed to further. It’s impossible to fully trust in the practice of dieting as something that will help you to lose weight even if you stick to the letter of a diet if you are inclined to deny that such a practice will further your goal of losing weight. It’s impossible to fully trust in a practice while denying that it is reasonable to engage in that practice as a way to attain the relevant ends which are supposed to be furthered by that practice, so if we want to find a good answer to RSTDP, we have to restrict our attention to defending practices which we trust in this sense, in the sense which involves a higher-order commitment to the reasonability of trusting in those practices.

But must RSTDP be asking about full trust? We might think that our everyday trust is a sort of ‘in-between’ trust77; we have in-between trust when we exhibit some of the dispositions associated with trust, but not all of them, and not in all contexts. HOCTC could be true for full trust but not for in-

77 In-between trust should be thought of by analogy with in-between belief; see Schwitzgebel 2001, Borgonni 2016, 2018, Archer 2017.
between trust; perhaps we have in-between self-trust by relying on our faculties but without being committed to their truth-conduciveness.

Someone doing this would be behaving in ways which they’d be prepared to deny were reasonable. That would be behaving akratically, which means they wouldn’t be behaving as they ought. Since RSTDP is asking about what’s reasonable, it ought to be about the sort of self-trust which *prima facie* could be reasonable. Since in-between trust is automatically unreasonable because it’s akratic, it shouldn’t be considered when answering RSTDP.

Note that this doesn’t mean that in relying on our faculties we’re treating

**Truth-conducive (TC)** Our basic doxastic practices are truth-conducive, in the sense of forming a high ratio of true to false beliefs

as a premise in any sort of inference. It might be that we form beliefs *via* our epistemic faculties immediately, and we shouldn’t attribute even implicit inferences based on TC to someone who sees a red apple and thereby comes to believe there is a red apple before them. Even if you endorse a non-inferentialist constraint on accounts of sense perception, you ought to deem someone who relied on an epistemic faculty and denied TC with respect to that faculty to be irrationally akratic.

It does mean that we must be committed to the truth-conduciveness of our own faculties in an exclusive sense. Suppose we rely on D. Then there are very many practices similar to D which differ only with respect to some transition T. Consider some list of these practices D₁, D₂ and so on. Then if D has a high probability of forming true beliefs, any Dᵢ will also have a high probability of forming true beliefs. If we’re committed to the truth-conduciveness of D, we might also be committed to the truth-conduciveness of D₁, D₂ etc. But commitment to the truth-conduciveness of D ought to be commitment to the truth-conduciveness of D in particular, not a family of related practices. That’s because we’re committed to D by our possession of the dispositions constituting D, and only to D₁, D₂ etc. by their similarity to D. Each time we form a belief by relying on D, we ought to be committed to the truth-conduciveness of D in that particular instance. But we need not be committed to the truth-conduciveness of any Dᵢ in that particular instance, since they may differ in a relevant transition in that instance. So HOCTC attributes to us a stronger commitment than commitment to the high probability of D’s being truth-conducive, but also a commitment to each
transition constituting $D$ to be truth-conducive. In doing so it can capture the way in which we’re particularly committed to $D$ in a way which doesn’t include $D_1$, $D_2$ etc.

Here are two reasons to deny HOCTC. First, it doesn’t seem to capture the self-trust of children and animals, because children and animals can’t form higher-order commitments – or, at least, children and animals can’t form many sorts of higher-order commitments, in particular those which require more than lacking the disposition to deny that our faculties are truth-conducive. Anything which requires having an attitude with the content that one’s epistemic faculties are truth-conducive risks attributing conceptual resources to animals and children which they don’t possess.

When we’re interested in what a concept like self-trust is, we’re interested in what a mature creature such as us engages in. I don’t mean to deny that children or animals engage in doxastic practices in just the same sense as we do; what I do want to say is that a mature creature like us can’t do so without taking some sort of view on the status of those practices, if only implicitly. In this respect self-trust is like belief. It’s plausible that children and animals have beliefs without being able to form higher-order attitudes about those beliefs. But subjects who are able to form such higher-order attitudes must sometimes do so, when those attitudes are made salient and are necessary for their other attitudes to be reasonable.

Second, it’s sometimes denied by reliabilists\(^78\) that we need have reliable grounds for belief in the reliability of our doxastic practices for it to be reasonable to engage in those practices; if all it takes for it to be reasonable to engage in a practice is just for that practice to be reliable, it’s not obvious why there should be any requirement that we have reliable grounds for the reliability of that practice.\(^79\)

If this means that we don’t need prior evidence for the reliability of a faculty before it’s reasonable to rely on that faculty, the reliabilist may be right. But the reliabilist still needs some sort of story about why akrasia is in general unreasonable. For the reliabilist, believing that $p$ and believing that believing $p$ is unjustified is unreasonable in just the same way as believing that $p$ and believing that the belief that $p$ is formed through an unreliable process is bad. The former pair of attitudes seems unreasonable to have, so the latter must also be unreasonable to have. That’s true even though we

---

\(^78\) Zalabardo 2005

\(^79\) Not that all reliabilists would take this line – someone like Goldman (Goldman 1967) would say that we also require second-order reliable processes which determine our reliance on our first-order processes, and which would generate higher-order commitment.
can form the belief that p through a reliable process without having a second-order process of checking whether our first-order processes are reliable.

Perhaps the reliabilist could say that reliance on our faculties provides evidence for the reasonability of relying on them, and hence justification ‘bubbles up’ from first-order beliefs to higher-order commitment. Perhaps they could say that there is some different normative status, besides justification, which attaches to higher-order commitment. However that goes, some such story must be available if reliabilists aren’t to routinely admit akratic attitudes.

1.3 Truth-conduciveness, Trade-offs and the Value of Engaging in a Practice

Even if it’s granted that a commitment to the reasonability of engaging in a practice is a prerequisite for trusting in that practice at all, why should that notion of reasonability of engagement be cashed out in terms of truth-conduciveness? Why, if we’re committed to the reasonability of relying on a practice, must we be committed to the truth-conduciveness of that practice?

The initial thought is that doxastic practices are for something; they’re supposed to get us epistemic value, just as carpentry is supposed to get us furniture. When we trust a practice, we’re trusting it to do what it’s supposed to do.81 If epistemic value has something to do with truth, doxastic practices are doing what they’re supposed to do when they get us true beliefs.

Here’s one problem with this argument. Suppose the structure of our doxastic practices looked like the following: we have a constellation of doxastic practices D_1, D_2...D_n such that each D_i took as inputs either outputs from D_i itself or outputs from a core doxastic practice I.82 For example, we might have a core doxastic practice, I, comprising transitions of the form: From the input: it seems to me that p, output: belief that p. I would convert all seemings into beliefs, while various specialised doxastic practices would develop these seeming-beliefs within specific domains – sense

80 Tal 2017
81 We can also trust practices to achieve other goods, as when we engage in artistic practices to build social status; in general trust has a tri-partite structure where S trusts a process or agent to secure a good (Baier 1986, Zagzebski 2012). In such cases we needn’t be committed to those practices being conducive towards the goods which they ostensibly aim at. But trust in a practice with no caveats is trust that that practice does what it’s supposed to do.
82 This need not be the empirical structure of our doxastic practices for the point to go through. Fodor (Fodor 1983) might say that this structure is exactly inverse from the structure discovered by empirical psychology; in that structure, we have lots of partitioned practices which feed into one practice which exhibits executive function. But that this structure is _a priori_ possible is enough to cause problems for the argument given.
perceptory practices would provide transitions for the various seeming-beliefs derived through sensory seemings, memory practices would provide further transitions for seeming-beliefs derived from memory seemings, and so on. Such doxastic practices would exhibit a ‘hub and spoke’ structure, with I as a central practice and all other doxastic practices exclusively connected to I (and themselves) but no other doxastic practice. Now suppose you thought that it’s reasonable to engage in a doxastic practice in virtue of that engagement maximising expected epistemic utility\textsuperscript{83}; on this view, it’s reasonable just when you expect to have the best possible ratio between true beliefs and false beliefs\textsuperscript{84}. Suppose also we have reason to think that each $D_i$ is truth-conducive while having no reason to think that I is truth-conducive. This may be because each $D_i$ gives independent confirmation of the truth-conduciveness of every other $D_j$ by delivering the same verdict on overlapping contents, but because of the hub and spoke structure, no other doxastic practice can give independent confirmation of the truth-conduciveness of I itself, as each practice relies on I for its own outputs. Then it might be reasonable to engage in I for the purposes of maximising epistemic utility because engaging in I is a prerequisite for engaging in every other doxastic practice, though we wouldn’t want to be committed to the truth-conduciveness of I itself. Perhaps what we’re interested in are the final outputs of each $D_i$ practice, and we’re concerned with I only insofar as it plays a distant role in generating these outputs. Then we could be committed to the reasonability of engaging in I without being committed to the truth-conduciveness of I.

Here’s one initial response: such a possibility is illusory because it is only reasonable to engage in I for the purpose of feeding its outputs into the other doxastic practices if we think those outputs will tend to be true. If our concern is with getting true beliefs out of each $D_i$, we need to be concerned with what we’re feeding in as inputs, and if we enter an input into a transition contained in some $D_i$ we will only be behaving reasonably if we do so while committed to the likely truth of that input; assuming we are committed to the truth-conduciveness of $D_i$, we are committed to the proposition that $D_i$ typically outputs true beliefs when true beliefs are inputted, so we can be committed to the truth-conduciveness of $D_i$ only if we are committed to the truth-conduciveness of I.

But this may miss the point of this approach. The thrust of the initial objection is that the strategy of trusting in the outputs of I dominates not trusting in those outputs when trying to maximise epistemic utility. Trusting dominates not trusting because if I is in fact truth-conducive, then trusting

\textsuperscript{83} Pettigrew 2013

\textsuperscript{84} ‘Best possible ratio’ need not mean ‘more true than false’; there might be extra disutility that comes with having a false belief when compared with the utility that comes with having a true belief, so in some cases it may be better to minimise belief.
in I is the best way to maximise true beliefs, while if I isn’t truth-conducive, the potential for forming true beliefs at all is so limited that trusting is no worse a strategy than not trusting. Part of what this objection gets at is that you appear to be able to act on a provisional basis as though I is truth-conducive without, it seems, having fully committed to I’s truth-conducive, at least not in the same way that you will have fully committed to D,’s reliability.

This is a sort of dominance argument: you can either trust or not trust the outputs of your faculties as a body; however the world is, trusting dominates not trusting with respect to maximising true belief, because if the world is cooperative, trusting leads to more true beliefs, while if the world is uncooperative, we ought to be indifferent between trusting and not trusting; hence we ought to trust our faculties. This argument fails because it’s not plausible that it’s reasonable to maximise true belief. If it were, you ought to believe every proposition and its negation, since doing so guarantees a true belief per proposition. Rather, you ought to maximise epistemic utility, accounting for the disutility which comes with having false beliefs. But in that case the dominance argument fails, because the epistemic disutility that comes with false beliefs means that, in the case where the world is uncooperative, we shouldn’t be indifferent between trusting and not trusting; in such cases, not trusting is the preferred strategy because it minimises epistemic disutility.85

So as long as the fundamental epistemic good involves truth, a doxastic practice can be reasonable to trust only as long as that doxastic practice is itself truth-conducive, not because it promotes truth-conduciveness in general. That’s because there’s no way for a doxastic practice to promote truth-conduciveness in general without itself being truth-conducive. Because it’s reasonable to trust a doxastic practice only if it’s reasonable to be committed to that doxastic practice’s being good at promoting the fundamental epistemic good, it’s reasonable to trust a doxastic practice only if it’s reasonable to be committed to that practice’s being truth-conducive. To be good at promoting a set of goals is to be satisficing with respect to those goals; expecting optimal strategies that would maximise those goals seems too much to ask of a practice. There’s no way for a doxastic practice to be satisficing with respect to promoting the fundamental epistemic good without outputting a sufficient ratio of true to false beliefs, so there’s no way for a doxastic practice to be good at promoting the fundamental epistemic good without being truth-conducive.86

85 This argument is made in Littlejohn 2012.
86 This is true when considering the value of doxastic practices synchronically. At any given time, a doxastic practice promotes the fundamental epistemic good if it either outputs a sufficient ratio of true to false beliefs, or enables a sufficient ratio of true to false beliefs from other practices. Treated synchronically, a doxastic practice can enable a sufficient ratio in other practices only via its outputs feeding into those other practices as inputs, and it can do so only if its outputs have a sufficient ratio of true to false belief. Treated diachronically, this doesn’t hold, because
Although I’ve been assuming that truth is the fundamental epistemic good, I don’t need to assume that for the argument for HOCTC. To see that let’s first look at some reasons to doubt that true belief is the fundamental epistemic good. We might think that instead knowledge is the fundamental epistemic good, as Clayton Littlejohn has argued. One reason Littlejohn has offered is that we can have accidentally true beliefs, but we don’t seem to accord them the same epistemic value as we accord to knowledge. This can be brought out by considering the accidentally true beliefs which may be formed if you entered one of Nozick’s experience machines. You may form the belief that your sister had received a job offer on the basis of the experiences you have while in the machine, while your sister has, as it happens, really received a job offer outside of the machine, too. We don’t feel inclined to give that accidentally true belief any more epistemic credit than all of the false beliefs you’d have formed in the machine. That suggests that what we really value in epistemic contexts is knowledge, not true belief.

Rather than argue against Littlejohn’s knowledge-first conception of epistemic value, I’ll try to sidestep the problem. HOCTC just says that in trusting in a doxastic practice we must have a commitment to its truth-conduciveness. But this is compatible with a choice of knowledge as the fundamental epistemic value over true belief. In either case, to pursue the value will involve attempting to form true beliefs, since the best way to try to achieve knowledge is to engage in doxastic practices which are conducive to forming true beliefs. Conversely, the best way to form true beliefs is to engage in just those practices you would expect to generate knowledge, since knowledge that p has features – modal anti-luck conditions, excluding relevant alternatives where p is false – which make practices aimed at generating knowledge particularly likely to generate true belief. That would be true even if we were in an experience machine. So the question of which is the fundamental epistemic value can be left to one side for the purposes of formulating and responding to RST2.

---

doxastic practices can enable a sufficient ratio of true to false belief in other practices without having truth-conducive outputs. Suppose, for example, that S engages in a doxastic practice D which involves receiving testimony from others; when engaging in D, S typically believes whatever they are told by those they trust, even if believing what they’re told has bad consequences for S. Then the practice D*, formed by replacing transitions in D relating to S’s acquiring a dangerous illness with a conviction that S will survive whatever illness he might have, will be relatively truth-maximising. (Or it will on the assumption that S, like us, is more likely to survive their illness if they believe they will survive, and will then go onto acquire many more true beliefs.) Norms of belief seem to be different from norms of inquiry, so it would be changing the subject to treat the reasonability of self-trust diachronically, but it’s worth noting that the argument I’m making is conditional on the assumption that we’re interested only in synchronic reasonability.

87 See Littlejohn (forthcoming).
In general, for any plausible fundamental epistemic good E, it must be the case that possessing E either entails having a true belief, or significantly raises the probability of having a true belief. For a good to be distinctively epistemic it must bear on the truth of some belief in some way, however indirect; any plausible way of so bearing will significantly raise the probability of having true beliefs. That might be as indirect as forming many true beliefs because you exemplify many epistemic virtues, where those virtues are the fundamental epistemic good. But if those epistemic virtues didn’t raise the probability of acquiring a high proportion of true beliefs, they wouldn’t be counted as peculiarly epistemic virtues. That isn’t to say that there aren’t other epistemic goods which we also value. We value originality, relevance, insightfulness, explaining many phenomena, having many interesting consequences and serving as the basis for an extended research program; concomitantly, we value the processes, dispositions and virtues which, directly or indirectly, tend to produce these epistemic goods. But since forming true beliefs is a distinctively epistemic good, any plausible fundamental epistemic good must be conducive towards forming such attitudes. If profundity is also an epistemic good, then any plausible fundamental epistemic good would have to be conducive towards profundity, too.

HOCTC therefore seems plausible not as a characterisation of what we’re fundamentally concerned with in epistemic matters, but with a lowest common denominator of what we’re fundamentally concerned with. Since any fundamental epistemic good must be conducive towards the formation of true beliefs, if we’re to be reasonable in trusting our epistemic faculties we must at a minimum be concerned with truth-conduciveness. There might be further questions of why it’s reasonable to rely on our faculties for other epistemic goods beyond truth, but HOCTC raises enough problems on its own to deserve special focus. I take it that Reliance and HOCTC are therefore both necessary conditions for reasonable trust in our epistemic faculties.

Are they sufficient? It’s sometimes thought that there’s more to trust than reliance and higher-order commitment, at least with respect to epistemic trust in others. In particular, it’s sometimes thought that trust in others as sources of testimony requires an affective component. This might be thought necessary to explain the normative judgments we make of others when they succeed or fail in living up to the trust placed in them by others; a good account of epistemic trust has to account for the outrage we feel when others betray our trust as against our response when a machine fails to work as expected. We don’t blame a thermometer for giving a misreading, so our trust in others must be more than treating them like a thermometer, and it’s plausible that that extra component is affective.

---

However, although the phenomenon of self-blame and failure in one’s duty to oneself is interesting, I don’t think the affective component of trust is particularly important here. Failure in one’s duty to others must be a violation of the reasonable expectations which they have in you, and an affective component may be a vital part of accounting for those reasonable expectations, but the question of what reasonable epistemic expectations you can have in yourself seems far closer to the reasonable expectations that can be placed on you in doing your epistemic duty at all. So although I don’t want to deny that there may be an affective component to self-trust, I don’t think it will be important for the case that I want to make for the reasonability of self-trust, since the alleged unreasonability of self-trust, and the response to that unreasonability, seems to be adequately accounted for by the reliance and higher-order commitment components.

1.4 Conclusion

In this chapter I’ve argued that we ought to understand RST2 in the following way: in asking us why it’s reasonable to trust our faculties, RST2 is asking why it’s permissible to engage in the doxastic practices which correspond to our faculties. I’ve argued that to engage in a doxastic practice is to trust it, in the sense of relying on that practice and holding a higher-order commitment to that practice’s successfully achieving the sort of goods which it aims at. RST2 can therefore be sharpened into a question about the reasonability of our reliance and our higher-order commitment to the reasonability of that reliance.

Although both aspects of the question will matter, the reasonability of higher-order commitment is easier to get a grip on, because we have a plausible candidate notion ready to hand: a commitment is reasonable just when it is a justified belief. In the next chapter I will look at one major problem with this plausible candidate, and three attempts to avoid that problem.
Chapter Two
Three Ways of Vindicating Self-Trust

2.1 Reasonability and Justification

I’ve argued that trusting one’s epistemic faculties requires a higher-order commitment to those faculties being truth-conducive; to avoid akrasia, we have to have some higher-order commitment to the reasonability of relying on those faculties, and it can only be reasonable to do so if they can be expected to secure some fundamental epistemic good. And that fundamental good must, at a minimum, must be conducive towards truth (possibly indirectly and among other epistemic goods).

If that’s right then RSTDP looks like a pressing question. That’s because it’s plausible that

Doxastic Commitment (DC) The commitment we have when we have an HOC in TC is belief.

If DC is true, what would make commitment to our own trustworthiness reasonable? It’s natural to think that beliefs are reasonable if they’re justified. To be justified is to obey substantive and structural norms regulating belief – that is, to respond appropriately to reasons which bear on having a belief, and to avoid holding sets of attitudes which exhibit the sort of incoherence typified by epistemic akrasia. Being justified seems like a good way of trying to have true beliefs and minimise false beliefs. So if true belief is part of the fundamental epistemic good, and being reasonable is doing well at achieving that good, then it’s plausible that being reasonable is just being justified:

Reasonability Entails Justification (REJ) It’s reasonable for S to believe p only if S is justified in believing p.

We now have three options. We could deny DC; if REJ makes it impossible to vindicate self-trust, that might give us more options to vindicate self-trust by appealing to some understanding of reasonable commitment which doesn’t involve justified belief. I’ll argue in the next chapter that this approach fails. We could endorse DC and accept REJ, trying to explain how an HOC is can be justified. That’s the approach I’ll look at in this chapter, again concluding that it fails. Finally, we could deny REJ and try to find some other way of understanding ‘reasonability’. That’s the
approach I’ll take in the remaining three chapters, arguing that HOCs can be reasonable when they’re excused.

I’m going to call those reasons which bear directly on the accuracy of a belief *evidence*. If the only reasons involved in substantive constraints on justification are those which bear on the accuracy of a belief, then a subject can only believe p with justification if their evidence supports p. Supporting p plausibly involves speaking in favour of p to a sufficient degree, with no defeaters or underminers cancelling that support.\(^89\) If that’s right, it supports the following principle:

**Evidentialism** S is justified in believing that p only if S has sufficient evidence for p.\(^90\)

If this view of justification is correct, then for self-trust to be justified we would have to possess sufficient evidence speaking in favour of the truth-conduciveness of those faculties. This raises a problem of circularity:

(C1) Either the evidence we have for the truth-conduciveness of a doxastic practice has been generated by reliance on that practice, or it has been generated by reliance on another doxastic practice.

(C2) If evidence for the truth-conduciveness of a practice has been generated by reliance on that practice, that evidence is circular.\(^91\)

(C3) If that evidence has been generated through relying on another doxastic practice, that evidence is undefeated only if we have an HOC in the truth-conduciveness of that practice.

We can now re-run (C1)-(C3) with respect to our HOC to the truth-conduciveness of this new doxastic practice. There are few enough basic practices that trying to provide further justification quickly ends in a very short loop; we justify sense perception on the basis of evidence from memory, for example, but, if we can give evidence for the truth-conduciveness of memory at all, it seems likely we will have to do so by appealing to sense perception. So any evidence for the truth-conduciveness of our basic doxastic practices is circular. But circular evidence can’t support a proposition, because it can’t bear on whether a proposition is true or not. Calls this the *argument from circularity*: we can’t be justified in trusting our faculties because doing so would require non-

---

\(^89\) For example, see Munoz 2019.


circular evidence for the truth-conduciveness of those faculties, evidence which we can’t provide since any evidence we generate is generated via reliance on those same faculties.

The argument from circularity made certain assumptions:

**Anti-circularity** Circular evidence can’t support a proposition.

**Anti-privilege** There are no privileged epistemic faculties which can serve as a foundation to demonstrate the truth-conduciveness of the other faculties

**HOC-evidentialism** An HOC is justified only if we possess sufficient evidence that our faculties are truth-conducive.

In this chapter I want to consider three responses to the argument from circularity, each denying one of these assumptions. First, we can deny Anti-circularity; in particular, we might think that track record arguments provide sufficient evidence for trust in our own faculties. I discuss this in 2.2. We might deny Anti-privilege; in 2.3 I discuss one way of doing so by looking at an *a priori* argument for self-trust given by Christopher Peacocke. Finally, in 2.4 I discuss rejecting HOC-evidentialism, in particular by permitting practical reasons for belief.

### 2.2 Bootstrapping and Track Record Arguments

#### 2.2.1 Are bootstrapping arguments obviously illegitimate?

A track record argument for the truth-conduciveness of a practice D is an argument of the form:

(T1) p (generated by D)

(T2) D generates a belief that p.

(T3) So D has generated a true belief.

(T4) There are many cases like T1-T3, and few cases where D generates p which are not like T1-T3.
(T5) So D very often generates true beliefs and rarely generates false beliefs.

(T6) So D is truth-conducive.

Although track record arguments can be circular, this is not so in all cases. We very often employ track record arguments to determine whether non-fundamental doxastic practices are trustworthy.\textsuperscript{92} We can do that when we can independently verify p, generated by D. When we can do so a track record argument provides non-circular evidence for the truth-conduciveness of D.

But we want to show the truth-conduciveness of our basic faculties. To apply a track record argument to those cases we would need to independently check whether the outputs of our basic faculties are true. The argument from circularity showed that this results in a loop, because we’d need to appeal to another basic faculty, one which would also need vindication. Applying track record arguments to our basic faculties only produces circular evidence for their truth-conduciveness. I’ll call any track-record argument which is circular in this way a \textit{bootstrapping} argument.\textsuperscript{93}

Since bootstrapping cases are just track-record arguments which happen to be circular, unless there is some good argument against that circularity the default assumption ought to be that bootstrapping arguments are good arguments. As David Papineau argues, when we are asked to show that a method is reliable, we have recourse to standard methods of investigation, the same methods we use to inquire into pretty much anything; if we are asked the question “are inductive inferences always reliable?”, then “what is more natural than to try to resolve this question by means of our normal procedures of investigation, which include, as it happens, our inductive procedures?”\textsuperscript{94}

\textsuperscript{92} Alston 1993 pp.12-3

\textsuperscript{93} As Jonathan Weisberg (Weisberg 2012.) has argued, we require minimal assumptions to produce bootstrapping cases. Bootstrapping need not arise only for those who think we’re justified in relying on our faculties before we have a justified belief that our faculties are truth-conducive. It’s also a problem for someone who already had a justified belief in TC, because even they could bootstrap their way to further conclusions – for example, that their faculties were much more truth-conducive than is necessary for their being justified in relying on them. For more on bootstrapping, see Vogel 2000, Weisberg 2010, 2012, Brueckner 2013, Wedgwood 2013, Cohen 2010, Kornblith 2009, Melchior 2020b, Jackson 2016, Douven and Kelp 2013, Nair 2019. Bootstrapping arguments are generally taken to be problematic, even if pinning down the way in which they’re problematic is hard, but Van Cleve defends them as good responses to scepticism; see Van Cleve 1984, 2003, 2015.

\textsuperscript{94} Papineau 1992 pp.14-5

\textsuperscript{95} Bergmann 2004 makes the similar point that track-record arguments seem in general unobjectionable and hence the presumption ought to be in favour of bootstrapping arguments. Bergmann would claim that although such arguments are in general sound, they need not confer justification on their conclusion, because justification transmission is context-sensitive; in particular, it obtains only when the premises of an argument are unquestioned.
We can’t object to all arguments which are rule-circular, because, you can rely on your own memory of winning a memory game to assure yourself that you have a good memory, this is rule-circular, but seemingly sound and even informative. Since bootstrapping arguments are rule-circular track-record arguments, and rule-circularity and track-record arguments are not, in general, objectionable, we can’t assume that bootstrapping arguments don’t confer justification on their conclusions.

Finally, that there is something intuitively problematic about bootstrapping arguments isn’t enough to show they don’t confer justification, because it’s open to the bootstrapper to claim that the intuitive defectiveness of their arguments is because they’re dialectically ineffective, especially when levelled against the sceptic. Brown and Pryor both argue that bootstrapping arguments are faulty only in that they would fail to convince the sceptic; but an argument can convey justification without being able to convince the sceptic, so being dialectically ineffective doesn’t prevent an HOC being justified by bootstrapping. If bootstrapping doesn’t confer justification then we need something more than intuitions here.

Bootstrapping arguments work only on the assumption that its premises, which have been acquired through the faculty whose truth-conduciveness is to be proven, are justified. Why are the premises justified? The defender of bootstrapping arguments can appeal to doctrines about justification, whether dogmatist, reliabilist or whatever, which render it plausible that our first-order beliefs are justified prior to forming any higher-order beliefs about their justificatory status. Bootstrapping arguments wouldn’t even get off the ground if the bootstrapper didn’t have some general story about how we’re justified in believing the premises of bootstrapping arguments. What the bootstrapper must deny is that the reasonability of reliance is explained by the prior reasonability of

---

96 This example is taken from White 2006 p.530.
97 Brown 2005
98 Pryor 2004. Coliva 2015 p.63 considers Pryor’s argument and points out that there is no reason for the sceptic to deny the premises if they are merely agnostic about their own truth-conduciveness, so this explanation of its defectiveness doesn’t go through, given Pryor’s commitment to perceptual dogmatism. But probably the sceptic that we have in mind when we intuitively judge the bootstrapping argument to be defective is one who has thought through the consequences of their agnosticism and is doing their best to be enkratic; hence the sceptic we have in mind will refuse to endorse the premises to preserve enkrasia.
99 Barnett 2014 argues that bootstrapping arguments are defective precisely because their premises are unjustified; when an evidential source’s trustworthiness is in question then we aren’t justified in any belief formed through relying on that source.
commitment to the truth-conduciveness of our faculties, the conclusion of the bootstrapping argument.

I’ll now consider three arguments against bootstrapping: violation of sensitivity, violation of disconfirmability, and failure to provide Bayesian evidence. I’ll argue that the first two fail to be convincing, except insofar as they implicitly appeal to the third. I’ll then defend the Bayesian argument against objections.

2.2.2 Problems with Bootstrapping

2.2.2.1 Discriminibility and Sensitivity

We’ve seen that there’s nothing obviously wrong with being rule-circular. But there is something worrying about a method which would give the same result even when applied to an unreliable doxastic practice. That isn’t the case with someone who relied on their memory to prove they had a good memory – people with bad memories often know that they have bad memories on the basis of remembering cases where they have been forgetful, so rule-circularity doesn’t entail indiscriminability. On the other hand, someone who relied on a practice of crystal ball gazing would also be able to give a bootstrapping argument for their HOC to the truth-conduciveness of that practice, and hence to offer a bootstrapping argument for our basic faculties is in Alston’s words “to offer a stone instead of bread” since “we can say the same of any belief-forming practice whatever, no matter how disreputable”.

Perhaps we can cash indiscriminability out in terms of sensitivity; the reason bootstrapping isn’t discriminating is because the conclusion of a bootstrapping isn’t sensitive, in the sense that any belief formed through a bootstrapping argument would still be held even if it were false. But

100 Alston 1993 p.17; see also Alston 1991 p.148.
101 This isn’t decisive against the bootstrapper because there are powerful general arguments against sensitivity. That sensitivity lacks the closure property (Kripke 2011) and that it struggles to accommodate believing on the basis of probabilistic evidence (Vogel 1987) suggests sensitivity doesn’t give strong reasons against bootstrapping arguments, and that we seem unable to have a sensitive belief that we don’t believe p falsely (Vogel 2001) all mean sensitivity can’t be uncontroversially assumed. Still, being insensitive is generally evidence that something has gone wrong, outside of special cases, so insensitivity does give some reason to doubt the legitimacy of bootstrapping. This is why sensitivity is still important even for conferring justification; if a belief can’t be
insensitivity can’t be what’s wrong with bootstrapping arguments for HOCs. That’s because such commitments couldn’t possibly be sensitive. Take Yael’s commitment to the truth-conduciveness of sense perception. The closest possible world where Yael’s practice of sense perception fails to be truth-conducive is one where Yael is systematically deluded; Yael regularly has sensory experiences which are wildly at odds with reality. But if Yael were systematically deluded in this way, then Yael would still maintain her commitment to the truth-conduciveness of sense perception. So the problem doesn’t arise from the bootstrapping argument, but rather with any argument which has as its conclusion that our doxastic practices are truth-conducive.

Instead the objection ought to be that bootstrapping arguments don’t confer justification because they can’t make a belief sensitive which, had it been held without being based on that argument, would have been insensitive. If Roxanne believes that her fuel gauge is reliable based on a bootstrapping argument, she is no better off than had she believed on the basis of no reasons at all, because in each case her belief is insensitive. But had Roxanne tested the level of the fuel tank and checked it against the gauge, her belief would have been sensitive, and an argument based on such tests would thus confer justification. The bootstrapping argument in favour of self-trust would thus be defective by exhibiting an argument pattern which is in general not sensitivity-raising.

But it seems implausible that non-defective arguments must be sensitivity-raising. Consider:

**Anti-Hypochondriac**

Ibrahim is considering whether to believe that he has a disease, one which is excruciatingly painful for anyone who has it, and which also causes sufferers to break out in yellow spots. Ibrahim reasons from the fact that he doesn’t have any yellow spots to the belief that he doesn’t have the disease. Ibrahim also isn’t in excruciating pain, but hasn’t considered any argument based on that premise. But if Ibrahim were in excruciating pain, he would notice, and would believe that he was diseased on that basis.

---


103 This example is taken from Vogel 2001.
Ibrahim’s belief that he is undiseased is sensitive, because if he was diseased he would be in excruciating pain, and would believe that he was diseased. Because Ibrahim’s belief is sensitive, the argument hasn’t raised the sensitivity of Ibrahim’s belief. Nonetheless, Ibrahim’s argument from the absence of yellow spots seems like a good one; cases where sensitivity is over-determined are therefore counterexamples to a sensitivity-raising requirement on legitimate arguments.

One thought is that Ibrahim’s use of the argument didn’t raise the sensitivity of his belief in this case, but other arguments of the same form would raise the sensitivity of other beliefs. What does it mean to be of the same form? I take it that shared logical form is a paradigm case of shared argument form, but there may be other cases of shared argument form; perhaps track record arguments also share a common form, given by T1-T6 above.

Perhaps then this argument equivocates between argument-types and argument-tokens, and the requirement is only supposed to be that argument-types are sensitivity-raising, while their tokens need not always be sensitivity-raising. The yellow spots argument is a token of *modus tollens*, and it might be that tokens of *modus tollens* are very frequently sensitivity raising, even if we can construct cases in which they aren’t.

But this is vulnerable to a generality problem, and not one which can be solved with popular solutions for the reliabilist generality problem. Any argument will be a token of many argument-types, and there seems no principled reason to take any one of those types as canonical so as to judge whether it is sensitivity-raising. Bootstrapping arguments are instances of many types – the bootstrapping type, the track record type, the inductive type and so on. Reliabilists might be able to solve their generality problem by appeal to psychological types; the canonical belief-forming process, which is to be evaluated for reliability, is the psychological type a token of which actually produced a given belief. But I can see no obvious equivalent to such psychological types here; perhaps there is one, but unless one is found the problem of sensitivity-raising can’t show why bootstrapping arguments are defective.

104 Comesana 2006, Tolly 2019
Isn’t it a shared presupposition in this debate that there are privileged argument-types the tokens of which ought to be evaluated in part on the basis of their being instances of those types? In particular, doesn’t both the bootstrapper and their opponent agree that there is a bootstrapping argument-type, and tokens of bootstrapping ought to be evaluated in part on the basis of their being tokens of that type? But it’s not obvious that the bootstrapper should concede that, because they might think that the relevant argument-type is that of an inductive argument, or of a track record argument, or some other type which is more general than bootstrapping. So sensitivity-raising requirements don’t seem plausible for explaining what’s wrong with bootstrapping.

2.2.2.2 Disconfirmability

If discriminability is cashed out in terms of sensitivity, then it’s not obvious that the indiscriminate nature of bootstrapping arguments is problematic. But it would be natural to construe indiscriminability in terms of the failure of the process to deliver contrary verdicts rather than a failure of the conclusion formed through that process to be sensitive. That motivates the following principle:

**Disconfirmability** If we know that a certain test cannot yield disconfirmation of our hypothesis, then no result of the test can confirm the hypothesis either.\(^{105}\)

Disconfirmability is about tests, and says nothing about arguments. But presumably arguments can function as kinds of tests. Whether you can provide an argument for a conclusion is a way of testing for that conclusion. And if that argument confers justification, it will confirm the conclusion. But if an argument can’t confer justification, it couldn’t confirm the conclusion – otherwise we could have all sorts of dubious tests which would confirm hypotheses by deploying invalid arguments. So Disconfirmability is implicitly about bootstrapping arguments and conferring justification as well as tests and confirming hypotheses.

\(^{105}\) I take this principle from White 2006 p.544.
It appears that for any doxastic practice we can produce a bootstrapping argument in favour of its truth-conduciveness. So the practice of producing a bootstrapping argument isn’t disconfirmable – we’d never be able to fail to produce such an argument. Weisberg objects that a bootstrapping subject still assigns positive probability to ‘I believe that p and ~p’ – the bootstrapper doesn’t think they’re infallible – and if we framed disconfirmability in terms of how we ought to update on ‘I believe that p and p’ then this would show that ‘I believe that p and p’ has some evidence. But we can just as well frame disconfirmability in terms of how we ought to update on producing a bootstrapping argument, and that process seems undisconfirmable.

Disconfirmability looks intuitively plausible. But as White points out, it doesn’t tell us which bit of the bootstrapping argument should be rejected. Does reasoning from ‘p’ and ‘D generates p’ to ‘p and D generates p’ fail to transmit justification? Then Disconfirmability would either be inviting us to give up the principle that justification is closed under conjunction, or committing us to the view that in such cases we already have prior justification for ‘p and I believe that p’. The latter seems dubious and in any case needs more on what that prior justification could be. The former would mean Disconfirmability is in competition with a highly plausible principle, and the presumption ought to be that Disconfirmability is false, unless we have a clear argument in Disconfirmability’s favour.

But if we don’t deny the closure of justification under conjunction, we also can’t deny that S is justified in believing the long conjunction p = ‘p₀ and I believe p₀ and p₁ and I believe p₁ and…’, since S is justified in believing each conjunction of the form ‘pᵢ and I believe pᵢ’. So S is justified in believing that they have a track record of success in relying on their doxastic practice. Does Disconfirmability then tell us to reject the next step – the inductive move from a track-record of success to the conclusion that S’s practice is truth-conducive? Again, if we’re comparing the intuitive plausibility of induction principles and Disconfirmability, Disconfirmability will lose unless we have some clear argument in Disconfirmability’s favour.

2.2.2.3 The Bayesian Argument against Bootstrapping

106 Weisberg 2012
107 Failure to transmit justification need not entail failure of closure. But it does mean that closure fails unless we have some other source of justification.
I’ll now consider the Bayesian or probabilistic argument against bootstrapping.** Such an argument has been offered against one sort of Moorean argument:

(M1) I have hands.

(M2) If have hands then I’m not being deceived by an evil demon inducing in me the sensation of seeing hands.

(M3) I’m not being deceived by an evil demon inducing in me the sensation of seeing hands.**

The Bayesian argument against this Moorean variation is:

(P1) An argument with the single premise P based on evidence E with conclusion Q confers justification on its conclusion only if \( P(Q|E) > P(Q) \).

(P2) In the Moorean argument, M1 is based on the visual seeming of hands.

(P3) \( P(M3|E) < P(M3) \)

(P4) Therefore, the Moorean argument doesn’t confer justification on its conclusion.

P1 is plausible if Evidentialism is true and evidence must be probability-raising. Being probability-raising need not be sufficient for being evidence in favour of Q for the argument to go through; all that’s required is, if evidence speaks in favour of the truth of a proposition, then at a minimum it raises the probability of that proposition.** If being justified requires having sufficient evidence, then an argument can only confer justification if its premises are evidence for its conclusion, so P1

---


109 For more on Moorean arguments, see Sosa 1999.

110 Moretti and Piazza 2013 (p.2492) point out that this fails when S has strong prior justification for p based on E*, if E* is then defeated by E which itself provides justification for p; E might be weaker evidence than E* while still providing sufficient evidence for belief, so E justifies S’s belief that p without raising its probability. This complication can be ignored for the sake of this argument; P1 could be replaced with P1*: An argument confers justification only if \( P(Q|E) >= P(Q) \), or, where E defeats E* such that \( P(Q|E^*) > P(Q|E) \), \( P(Q|E & ~E^*) > P(Q|~E^*) \). Because the latter disjunct isn’t applicable in the bootstrapping case – there’s no obvious candidate for E* - the substitution of P1* for P1 wouldn’t be important.
gives a necessary requirement for conferring justification.\textsuperscript{111} P2 is assumed by the bootstrapper. P3 is true assuming that E is entailed by ~M3, which is true because part of ~M3 just is having the sensation of having hands. This is plausible only if we assume that our evidence in ~M3 is just the same as in the non-sceptical scenario; call this assumption the \textit{same evidence} assumption. The same evidence assumption is controversial and I’ll say more about it shortly.

The analogous bootstrapping argument is:

\begin{enumerate}
\item[(B1)] By relying on practice D it has often been the case that I have believed p (generated by D) and p has obtained, and rarely the case that I have believed p and p has not obtained.
\item[(B2)] If relying on D frequently leads to believing p when p obtains, and rarely believing p when p doesn’t obtain, then D is truth-conducive, and in relying on D I’m not making systematically false judgments.
\item[(B3)] D is truth-conducive and in relying on D I’m not making systematically false judgments.
\end{enumerate}

Let E* be our apparently successful track record. E* is our evidence for B1. Since ~B3 entails E* and B3 doesn’t entail E*, P(B3|E*) < P(B3). ~B3 entails E* because making systematically false judgment entails having an apparently successful track record. So bootstrapping arguments for B3 don’t confer justification.

Here are two problems for this argument. First, the Bayesian argument might not cover all intuitive cases of transmission failure. Pryor points out that a Moorean argument with the conclusion ‘I have hands and I’m not a disembodied spirit’ doesn’t face the same difficulties, but still seems intuitively problematic.\textsuperscript{112} The Bayesian argument can nonetheless provide an indirect answer to this problem. If such Moorean arguments confer justification, while the Moorean variant discussed above does not, and the premises of both are based on just the same evidence, then a subject can be committed to ‘I’m justified in believing that I have hands and I’m not a disembodied spirit, but I’m not

\textsuperscript{111} Weatherson 2007, Pryor 2013 and Kung 2010 suggest revising our account of evidence such that P1 is false; doing so would be motivated by the need to defend dogmatism, but P1 seems less controversial than dogmatism so all else being equal we ought to preserve P1 at the expense of dogmatism if the two are in conflict.

\textsuperscript{112} Pryor 2013. Pynn 2013 gives another case of transmission failure which can’t be explained by the Bayesian argument.
justified in believing that I have hands and am not a disembodied spirit being deceived into thinking I have hands by an evil demon'. Such a commitment sounds strange, because it sounds strange to be justified in believing yourself not to be in a sceptical situation (being a disembodied spirit) without being justified in believing yourself not to be in very similar situations (being a disembodied spirit deceived by an evil demon.) Since the only cases compatible with your evidence in which you’re a disembodied spirit are also those where you’re massively deceived in some way, it sounds strange to justifiably believe that you aren’t in such a case without justifiably believing you aren’t massively deceived.

Second, while the Bayesian argument seems to be successful against internalist bootstrappers, like Pryor, an externalist bootstrapper like Van Cleve would deny the shared evidence assumption. They’d say that our evidence supporting B1, E*, is that we have had a successful track record, not that we appear to have had a successful track record. ~B3 doesn’t entail E*, because ~B3 entails that we have had a wholly unsuccessful track record. So it’s not the case that P(B3|E*) < P(B3).

But we can adjust the Bayesian argument such that the shared evidence assumption would be acceptable even for the externalist. Consider:

(B1) By relying on practice D it has often been the case that I have believed p and p has obtained, and rarely the case that I have believed p and p has not obtained.

(B2*) If relying on D frequently leads to believing p when p obtains, and rarely believing p when p doesn’t obtain, then D is truth-conducive, in relying on D I’m not making systematically false judgments, and hence I’m not getting extremely lucky in having so many true beliefs.

(B3*) D is truth-conducive, I’m not making systematically false judgments, and hence I’m not getting extremely lucky in having so many true beliefs.

I’m intending B3* to be read such that ~B3* is true if a random sample of my beliefs, those which I’m using as my track record, are true just by luck, as they would be if I were being routinely deceived by an evil demon who occasionally permitted me to have some true beliefs. On that way of reading B3*, the shared evidence assumption would be true, even for the externalist, because the
evidence E I’d have in ~B3* would be just the same as my evidence in B1. My evidence in both cases would be the veridical visual appearance of hands; both would be a sequence of veridical visual experiences, although in the former that sequence would not be possessed by luck and in the latter it would. So again, E would be entailed by ~B3* and the standard Bayesian argument still goes through.

Could the externalist deny the shared evidence assumption even here? Perhaps they could deny that evidence generated by a generally unreliable faculty is the same as evidence generated by a generally reliable faculty, even if that evidence is similar in other respects (such as involving the veridical visual presentation of a hand.) Perhaps they could deny that veridical experiences produced by luck is evidence at all. But this seems to drop the important role of evidence in being shared across different practices, a role which helps us to compare the reliability of those practices. It also seems unmotivated, so this doesn’t seem a plausible route for the externalist bootstrapper to take.

This isn’t decisive, but does push the externalist towards a form of reliabilism such that veridical perceptions can fail to be evidence, and fail to be the very same evidence as introspectively indistinguishable veridical experiences produced by the very same psychological processes. That seems strange. The standard disjunctivist position is that there’s no experience when there’s no object; hallucinations of hands aren’t experiences of hands of the same nature as veridical experiences of hands.113 The standard disjunctivist denies that introspectively indistinguishable experiences must be identical between counterfactuals. But the standard disjunctivist doesn’t deny that introspectively indistinguishable veridical experiences must be. So the externalist bootstrapper can’t take the standard disjunctivist approach here, because even on that approach the deceived subject would, in this case, have just the same experiences in their sample as the undeceived subject. Instead the externalist bootstrapper would be pushed into a more extreme form of disjunctivism on which the same veridical experience can be evidence or fail to be evidence between subjectively indistinguishable counterfactual situations. That more extreme position is unmotivated even if we’re sympathetic to standard disjunctivism, so we ought to reject externalist bootstrapping.

---

113 For more on disjunctivism see Soteriou 2016; for the relationship between disjunctivism and scepticism, see Walker 2019; for the relationship between bootstrapping and disjunctivism, see Stuchlik 2015.
It also fits poorly with at least one popular solution to the generality problem, because on that solution our evidence is generated by a particular psychological process.\textsuperscript{114} It fits poorly with this to say that psychological processes generate evidence when they are generally reliable but not when they aren’t. This would be to import notions of reliability into the conception of a psychological process. But psychological processes are supposed to be prior to judgments about reliability – if they weren’t then this wouldn’t be a solution to the generality problem. So this puts pressure on the externalist – they risk making an \textit{ad hoc} move to avoid the problem, and unless independent motivation can be provided for why they want to disagree with the standard disjunctivist approach or a plausible solution to the generality problem, we should conclude that the Bayesian argument is also effective against the externalist.

The externalist might object that while this would rule out certain bootstrapping arguments, they would still be able to bootstrap their way to an HOC, because they can bootstrap to B3. But on this view the externalist is committed to strange abominable conjunctions – ‘My faculties are truth-conducive and aren’t systematically false, but I can’t say whether my faculties are truth-conducive and aren’t systematically false in ways which mean I’ve acquired lots of evidence for their truth-conduciveness by chance.’ This seems hard to swallow even for those generally prepared to accept abominable conjunctions in cases of closure failure. We’re here evaluating two different sorts of anti-sceptical propositions. Arguments in favour of closure failure between ordinary and anti-sceptical propositions don’t generally extend to closure failure between anti-sceptical propositions, so these abominable conjunctions seem particularly difficult to accept.

To sum up: bootstrapping arguments don’t confer justification because their conclusions have lower probability conditional on the evidence for their premises than they do absent that evidence; this holds even for many externalist positions, and those externalist positions which would avoid this problem seem implausible and unmotivated. So bootstrapping arguments for self-trust don’t confer justification.

### 2.3 A Priori Arguments

\textsuperscript{114} See Comesana 2006, Tolly 2019.
2.3.1 Descartes and the Reidian Challenge

*A priori* arguments for self-trust have to do two things. First, they need to be sound. Consider Descartes’ *a priori* argument for HOCTC:

(1) God exists and is omni-benevolent.
(2) If God exists and is omni-benevolent, our epistemic faculties are truth-conducive.
(3) So, our faculties are reliable.

Descartes’ argument might be defective because it’s unsound, since (1) might be false. But another way in which *a priori* arguments can be defective is if they fail to meet the *Reidian challenge*. An argument for self-trust fails to meet the Reidian challenge just in case we can’t explain why the doxastic practice(s) which generate the premises and transitions of that argument deserve to be privileged over other doxastic practices.\(^{115}\) Suppose we try to argue for the truth-conduciveness of memory by appeal to testimony; many people assert that they have good memories, can recall many true things and so on. Such an argument would only be successful if it can be explained why the doubts which would lead us to require a vindication of trust in memory don’t also require a vindication of trust in testimony; the Reidian challenge asks why, if we’re opening the truth-conduciveness of *some* practices to doubt, we aren’t opening that of *all* practices?

Failure to meet the Reidian challenge is problematic because, without meeting that challenge, an argument can at most show relative truth-conduciveness – it can show that if the doxastic practice of testimony is truth-conducive, so is memory. But to have a justified belief in TC, we need to have a justified belief that sense perception, reason etc. are truth-conducive *simpliciter*, not relative to one another.

\(^{115}\) I take the Reidian challenge from Reid’s famous remark: “Reason, says the skeptic, is the only judge of truth, and you ought to throw off every opinion and every belief that is not grounded on reason. Why, sir, should I believe, the faculty of reason more than that of perception; they came both out of the same shop, and were made by the same artist; and if he puts one piece of false ware into my hands, what should hinder him from putting another.” (Reid 1997 p.169).
In this section I’ll look at Christopher Peacocke’s *a priori* argument for self-trust. The importance of Peacocke’s argument is in the way it supplies an answer to the two challenges outlined above. In particular, Peacocke has an answer to the Reidian challenge because he can claim that while it’s initially conceivable that sense perception could be unreliable, it isn’t conceivable that reason could be; it’s a *conceptual* truth that reason is reliable.

2.3.2 *Complexity Reduction*

Peacocke has an answer to the Reidian challenge because of his general account of the *a priori*. On Peacocke’s account, *a priori* beliefs are those which are guaranteed to be true in virtue of the possession-conditions of the concepts which compose them, as well as the general theory which fixes the semantic values of those concepts. Possession-conditions for concepts give us basic inferences involving those concepts – those inferences which Peacocke calls *primitively compelling*, which we’re disposed to do without appealing to further theses about these concepts. A theory which assigns semantic values to concepts – what Peacocke calls a *determination theory* – tells us the conditions under which a transition involving a concept is true. Determination theories ought to assign such conditions in such a way that the primitively compelling transitions which determine a concept are always truth-preserving.

Those contents which are true because of the primitively compelling inferences which constitute possession of the concepts composing those contents, together with the determination theory which guarantees that those transitions are truth-preserving, can be believed with *a priori* justification. Tautologies are paradigm cases of such *a priori* contents. A tautology like \( \sim(P \& \sim P) \) is guaranteed to be true in virtue of the basic inferences constituting possession of negation and conjunction, together with the truth tables for negation and conjunction. That gives Peacocke an answer to the Reidian challenge: *a priori* beliefs are privileged as a starting point for vindicating self-trust because they’re true solely in virtue of the concepts involved and the general theory which fixes the semantic values of those concepts.

---

120 Peacocke 2004 pp.171-2
Peacocke’s vindication of self-trust relies on:

**Complexity Reduction Principle (CRP)** Other things equal, good explanations of complex phenomena explain the more complex in terms of the less complex; they reduce complexity.  

Peacocke links complexity-reduction to likelihood of truth with:

**Qualified Principle of Sufficient Reason (QPSR)** Other things equal, it is more probable that a complex phenomenon has a complexity-reducing explanation than that it has no explanation, or that it has one that does not reduce complexity.

Peacocke argues that the complexity-reducing explanation for the existence of perceptual experience will be one which appeals to the production of our sensory faculties under pressures which select for truth-conduciveness. Since QPSR and CRP are *a priori*, we can know *a priori* that it’s more probable than not that our faculties are truth-conducive; it’s supposed to be *a priori* that explanations of our faculties which appeal to some form of natural selection selecting for truth-conduciveness are complexity-reducing, so it’s *a priori* probable that our faculties are truth-conducive.

Here are two problems with this argument. First, there is a proliferation of concepts which render both CRP and QPSR true but unhelpful for Peacocke’s argument. ‘Complexity’, ‘probability’, ‘explanation’ are clearly not concepts with readily-available analyses. There are a number of concepts which appear to match (to as close a fit as we can reasonably demand) our actual patterns

---

121 This still might not be enough to convince someone like Papineau. Papineau argues (in Papineau 1992) that the requirement for certainty is unnecessary for knowledge, and that the Cartesian temptation to demonstrate the reliability of our doxastic practices solely through intuition and (deductive) inference can therefore be resisted. Papineau might still agree that such a foundational project could have value, since it would provide greater certainty than we would otherwise have. But it wouldn’t be necessary.

122 Peacocke 2004 p.83
123 Peacocke 2004 p.83
124 “That the explanation of the occurrence of a perceptual experience with the instance-individuated content that produces which most reduces complexity is that the experience is produced by a device which has evolved by natural selection to represent the world to the subject” Peacocke 2004 p.87.
of behaviour and pre-theoretical intuitions surrounding these terms. CRP and QPSR might be conceptual truths when these concepts are fleshed out properly, but doing so is far from trivial. Peacocke has his own accounts of these concepts\textsuperscript{125}. But those accounts need defence and further comment. But that there is (and must be) such further comment is precisely the problem; once we move beyond the initial statements of CRP and QPSR to the defence of certain views of the concepts which constitute their contents, we are necessarily engaged in philosophical argument.

So there are a number of concepts in the neighbourhood of ‘complexity’, ‘probability’ and so on. And where there is disagreement over which concept in that neighbourhood is ‘complexity’, disputes over what’s makes an explanation complexity-reducing have been pushed back to the level of which concepts ought to be given preference. We now have an array of concepts complexity\textsubscript{1}, complexity\textsubscript{2} etc., each with different transitions which are found primitively compelling but which otherwise align well enough with our pre-theoretic notions of complexity. That’s not to say that each of those concepts deserves equal standing. Presumably there’s just one concept which deserves to be considered ‘complexity’. The problem for a Peacockian response to the Reidian challenge is that deciding between these concepts of complexity isn’t an error-free process – much philosophical discussion is testament to that fact, since much philosophical discussion involves disagreement over precisely this, and many participants in such discussions must be wrong. But if it’s not an error-free process then there is no error-free route to knowledge of QPSR or CRP.

Peacocke might want to appeal to implicit conceptions here\textsuperscript{126}. An implicit conception is a proto-concept constituted by prior sets of dispositions which partially determine a concept, as well as a disposition for fillings-out of that partial determination to seem right or wrong. Before we learn the inference rules and truth table for disjunction, we have an implicit conception of disjunction, because we can make some inferences involving disjunction and can recognise those inference rules and that truth table as the correct filling-out of ‘disjunction’. Peacocke might want to say that, although we don’t have a concept of complexity, we do have an implicit conception of complexity, and any correct filling-out of that implicit conception ought to render QPSR and CPR true. But while it’s plausible that validating QPSR and CPR are constraints on filling-out an implicit

\textsuperscript{125} An account of each may not be necessary; Peacocke could also rely on illustrative examples of each. The problem is cutting down plausible rival concepts to get to the right one, the one which makes CRP and QPSR both true and properly formulated. That can be done through providing sufficient examples to eliminate rival concepts even if an explicit analysis of the preferred concept can’t be given.

conception of complexity, it’s not so clear that the same is true of natural selection accounts of
human faculties. It’s at least open to question whether those accounts are complexity-reducing
relative to available alternatives. So even if Peacocke’s argument is successful, it fails to meet the
Reidian challenge.

Moreover, it’s not obvious that his argument is successful. The application of CRP to QPSR for the
vindication of self-trust seems to involve premises which aren’t conceptual truths. It’s not a
conceptual truth that any of our capacities has evolved by natural selection at all, even conditional
upon our actually having sensory experience. Is it a conceptual truth that this is the explanation
which most reduces complexity? It can only be so against a background body of beliefs – that the
Earth is four billion rather than six thousand years old, that there is an effective method of genetic
transmission which is both susceptible to random change and is reasonably accurate in transmitting
fitness-conducive traits; these are empirical and not conceptual truths. But without that background,
or if we had a different set of background beliefs, it might not be true that natural selection most
reduced the complexity of experience. Against a background belief that the Earth is six thousand
years old, natural selection could not explain our sensory faculties and the experiences generated by
them without implausible – and complexity-increasing – assumptions about the rate of change of
living species on this planet which would compare unfavourably with the hypothesis of an
intelligent designer.

2.4 Denying Evidentialism

2.4.1 The Practical Argument for Self-Trust

If it’s not clear how self-trust can be reasonable if we endorse Evidentialism, then perhaps we
should abandon Evidentialism. Consider Alston’s argument that we’re reasonable in carrying on
with our pre-existing doxastic practices for non-evidential reasons:

“Given that we will inevitably run into epistemic circularity at some point(s) in any attempt to
provide direct arguments for the reliability of one or another doxastic practice, we should draw the
conclusion that there is no appeal beyond the practices we find firmly established, psychologically
and socially. We cannot look into any issue whatever without employing some way of forming and
evaluating beliefs; that applies as much to issues concerning the reliability of doxastic practices as to any issue. Hence what alternative is there to employing the practices we find ourselves using, to which we find ourselves firmly committed, and which we could abandon or replace with extreme difficulty if at all? The classical skeptical alternative of withholding belief altogether is not a serious possibility. In the press of life we are continually forming beliefs about the physical environment, other people, how things are likely to turn out, and so on, whether we will or no...Hence we are just not in a position to get beyond, or behind, our familiar practices and criticize them from that deeper or more objective position.”

Alston looks to be defending the following principle:

**Doxastic Practices Epistemic Conservatism (DPEC)** If we’re already engaged in a psychologically and socially embedded doxastic practice, then we have prima facie reason to carry on engaging in that practice, even in the absence of positive non-circular evidential support for that practice.

Why should we believe DPEC? Alston gives the following sort of argument:

(RA1) In the ‘press of life’ we are bound to carry on forming beliefs, so scepticism isn’t a real alternative to continued trust in our doxastic practices.
(RA2) If scepticism isn’t a real alternative, there are no real alternatives to continued trust in our pre-existing practices.
(RA3) If there’s no real alternative to trusting in a practice, then it’s reasonable to trust in that practice.
(RA4) So it’s reasonable to continue trusting in our pre-existing practices.
(RA5) Since we have no positive, non-circular evidential support for such continued trust, such trust is reasonable even in the absence of evidence.

Call this the no real alternatives argument. I take RA5 to have been shown by the discussion in the first two sections of this chapter, and the inference from RA1, RA2 and RA3 to RA4 follows from two applications of modus ponens. If there are problems with this argument, they’re with its three premises.

127 Alston 1991 pp.149-50
There are problems with RA1 and RA2; RA1 needs sharpening to explain the sense in which we’re bound to carry on forming beliefs, and RA2 needs to address the possibility of akrasia, to show that reliance without higher-order commitment isn’t a real alternative. I’ll set these problems aside to focus on RA3. RA3 needs disambiguation: there are many ways in which we can be reasonable in engaging in a practice, not all of them epistemic, and it’s not obvious that when ‘reasonable’ is read in an epistemic sense that RA3 is true.

2.4.2 Threshold Pragmatism

We can read Alston as making a practical argument for trusting our epistemic faculties. It might be that trusting our faculties is reasonable in the sense that doing so involves responding to practical reasons. There are two senses in which we might respond to practical reasons for self-trust. On one view we might be making an argument from *threshold pragmatism*. Threshold pragmatists typically deny Evidentialism and instead endorse

**Credence Evidentialism (CE)** An epistemic subject S ought only to have credence X in p if S’s credence is supported by S’s total body of evidence

and

**Threshold Pragmatism (TP)** The threshold at which a credence X in a proposition p is sufficiently high to merit full belief in p is at least partially determined by practical reasons.

The thought is that CE is enough to do justice to our evidentialist intuitions – your model of the world shouldn’t be subject to reasons which don’t vary with the state of the world – but that those intuitions are best thought of as applying to credences rather than beliefs. That’s because the justificatory status of a belief seems to vary depending on whether important consequences hangs

---

129 I discuss the sense in which we’re bound to form beliefs and the relationship between scepticism and the possibility of epistemic akrasia in Chapter Six.

130 Alston says that “it is rational to engage in such [doxastic] practices...there is no rational alternative to this. It is a kind of practical rationality that is in question here.” Alston 1991 p.168. But we should be careful about how we’re using the concept of practical rationality here; Alston writes in those terms to distinguish the sort of rationality involved from epistemic rationality, where that is “the rationality we would show to attach to a belief if solid grounds for its truth were adduced, or to attach to a doxastic practice if sufficient reasons were given for regarding it as reliable”.

on that belief. In the classic case, whether S is justified in believing that the bank closes at 5pm in part depends on whether you have urgent business that must be completed by the end of the day, or whether your business at the bank can wait a number of days without disastrous consequences.\textsuperscript{132} The pragmatic sensitivity of belief in this respect seems to pose no threat to Credence Evidentialism, because in either case your credence in the bank’s opening hours remains the same, but it does suggest that whether you ought to have an outright belief that the bank won’t close depends in part on the practical consequences of the bank’s closing.

TP can’t be the right way of justifying RA3. That’s because TP seems plausible only when conjoined with CE. So our credence in TC, that our own faculties are truth-conducive, must be determined by all and only the evidence which bears on that truth-conduciveness. But, if RA5 is plausible for belief, then it’s also plausible that there is no non-circular evidence bearing on the justified credence to assign to TC. Presumably there is then no reason to assign a higher credence than the base rate for coherent doxastic practices. I don’t know what that base rate is. Maybe it’s senseless to ask what it might be; if it is, that undermines the attempt to employ TP to understand Alston’s argument, because that attempt is plausible only if it makes sense to assign a credence just based on our available evidence to the trustworthiness of our faculties, and that is plausible only if we can conceive of a sensible base rate for trustworthy faculties.\textsuperscript{133} But given the very large number of possible doxastic practices, and the very few which will be truth-conducive, it seems that our credence in the trustworthiness of the practices we happen to find ourselves with must be very low, if that credence is to be based solely on evidential considerations.

TP tells us that the threshold for belief is at least partially determined by practical considerations, paradigmatically, on the seriousness of the consequences if the proposition were falsely believed. The no real alternatives argument has offered some considerations which would lower the threshold, namely, that living up to a higher threshold would be extraordinarily difficult. But trust in your own faculties lies behind your basis for every action you will ever take. It is hard to see how the consequences of false belief could fail to be maximally serious; if missing the last chance to deposit a cheque at the bank is serious enough to raise the threshold for belief, surely the sum of every action you will ever take or fail to take ought also to raise the threshold to \textit{at least} that level. And while the difficulty – realistically, impossibility, at least for the great majority of epistemic

\textsuperscript{132} This case is taken from DeRose 1992, 2002. It is generally taken to apply to knowledge; if knowledge entails justification, its use here should be uncontroversial.

\textsuperscript{133} There might be other reasons which bear on the initial credence to be assigned to TC. But I can’t see what other purely evidential reasons there could be, at this stage of the argument, beyond considerations of the base rate of trustworthy faculties.
subjects – of failing to achieve that goal surely has some bearing on the threshold, perhaps lowering it a little, it couldn’t lower it as far as the very low credence that we ought to have in our faculties’ truth-conduciveness.

There are other problems. If our credence in TC has to be based on purely evidential considerations, then our overall credence distribution will either be akratic or impossible to maintain in everyday life, cutting against the motivation for RA1. They risk being akratic because they risk assigning a high credence to there being milk on the fridge, on the basis of sensory and memorial evidence, but a low credence to such evidence raising the probability of there being milk in the fridge. Such combinations of credences seems problematic in the same way that akratic combinations of beliefs is problematic. This feeds through into the role of credence in action. We would be prepared to act on the basis of very high credence that there’s milk in the fridge, but very low credence that we’re good judges of whether there’s milk in the fridge. This risks being vulnerable to Dutch book-like sets of bets, and will lead to incoherent actions – will you go up to get milk from the fridge you assign a high credence to there being milk in the fridge, or refrain from doing so because you assign a low credence to your possessing sufficient evidence for there being milk in the fridge?

2.4.3 Reasons Pragmatism

TP doesn’t seem to help the no real alternatives argument. We might instead try:

**Reasons Pragmatism (RP)** S is justified in believing p only if S’s total reasons, including practical reasons, are sufficient for believing p.

I’ll first look at Susanna Rinard’s argument for RP. Rinard has argued for pragmatism on the basis that we ought to treat justification for belief in just the same way as we treat justification for all of our other states. However the justificatory status of our other states ought to be determined, it will have to be such that practical reasons count towards being justified in being in a state. In particular, Rinard endorses

134 See Childers 2013.
135 Rinard 2015, 2017, 2019a, 2019b
**Equal Treatment (ET)** However the rationality of [mowing the lawn, wearing a raincoat, listening to music or playing with your dog] is determined, the rationality of any other state...is determined in precisely the same way.\(^\text{136}\)

So the sort of justification applicable to belief admits of practical reasons; there isn’t a special sort of epistemic justification unique to belief which excludes practical reasons.

Rinard says we ought to endorse ET for two main reasons. First, it presents a simpler, more unified and less fragmented view of justification. The evidentialist has to explain why justification comes in (at least) two varieties, while the pragmatist doesn’t. Second, it accounts for the general truth-conduciveness of reasons bearing on belief – it’s practically useful to have true beliefs, most of the time – but unlike Evidentialism, ET can explain why it sometimes seems justified to incorporate reasons which aren’t truth-conducive, as when we make promises which we lack sufficient reason to think we’ll keep, or when we trust a friend beyond what the evidence would support.

Rinard has responses to some existing evidentialist arguments. Consider the argument from the attitude/content distinction. Pamela Hieronymi\(^\text{137}\) has argued that practical reasons can’t bear on the justification for holding a belief because they bear exclusively on whether to hold an attitude, not on the content of that attitude. This isn’t a distinction drawn *ad hoc* for the purposes of refuting the pragmatist – it’s a distinction we’re compelled to draw when we consider the difference between having a reason to intend to \(\phi\) because \(\phi\)-ing is good, and when we have a reason to intend to \(\phi\) because having an intention to \(\phi\) is good. The two can come apart; so, too, can having a reason in favour of the content of a belief and having a reason to have that belief. A reason in favour of the content of a belief is a reason that bears on whether that content is true, but a reason in favour of the attitude is a reason in favour of thinking that holding that attitude would be good, whether for the subject in particular or good in general or in some other way. When we ask about what kinds of reasons can bear on having an attitude, our question is ambiguous between attitude-related reasons and content-related reasons. Once that ambiguity is resolved, the pragmatist objection to Evidentialism becomes unconvincing, because they are construing that norm as applying to attitude-related reasons rather than content-related reasons.

\(^{136}\) Rinard 2017 p.123; note that I’m here interpreting Rinard’s ‘rational’ as ‘justified’; what’s important here is that both are statuses which arise from being properly based on appropriate reasons.

\(^{137}\) Hieronymi 2005; an argument based on a similar distinction to Hieronymi’s is raised in Danielsson and Olson 2007.
Rinard\(^{138}\) objects to the distinction between content-reasons and attitude-reasons. We can make the same distinction between a reason for wearing wool socks and a reason for taking steps which will cause us to be in a state of wearing wool socks. But the practical reasons in favour of wearing wool socks – their warmth, say – are just reasons for causing us to be in a state of wearing wool socks, too. Unless the evidentialist wants to claim that we don’t have practical reasons to wear wool socks, their argument at this stage must fail.

Rinard’s arguments might establish that practical reasons can bear, all things considered, on the justification for having a belief. But I’ll follow Maguire and Woods in denying that they bear on beliefs insofar as beliefs play a role in our epistemic projects.\(^{139}\) Evidentialists ought to distinguish between the justification for holding a belief for the purposes of an epistemic project and the justification for holding a belief \textit{simpliciter}. The evidentialist should claim that, while practical reasons can bear on justification in the latter sense, they don’t bear on the former. This is plausible for the same reason that it can sometimes be reasonable to fold when you have a royal flush, because it might serve some of your interests outside of the round of poker to do so.\(^{140}\)

This isn’t the same distinction as that between attitude-reasons and content-reasons, but is rather the distinction between reasons all things considered and reasons which bear solely on achieving some good. Asking RST2 automatically introduces the context of engaging in an \textit{epistemic} project. An epistemic project is that pattern of norm-governed habits and dispositions which are aimed at achieving epistemic value. For any domain which we can be curious about, there’s a corresponding project engagement in which is the best way of satisfying our curiosity about that domain; we can engage in projects that will resolve our curiosity about cookery, or physics, or comedy. But we can also engage in projects that will resolve our curiosity about any proposition as such. RST2 introduces a context in which resolving such curiosity is salient, in the same way mentioning folding on a royal flush introduces the context of playing a round of poker without this being made explicit.

Engaging in an epistemic project is a way of satisfying our curiosity; we find ourselves with a desire to have true beliefs, and have true beliefs in a way which is shaped by reasons which are

\(^{138}\) Rinard 2015
\(^{139}\) Maguire and Woods 2020
\(^{140}\) This approach to the right kind/wrong kind distinction was developed by Schroeder in Schroeder 2010, and also figures in Maguire and Woods 2020. For some criticisms of this view, see Côté-Bouchard 2016.
conducive towards having true beliefs. That is, we want to have true, *epistemically* justified beliefs. There’s therefore a clear question which RST2 is raising: in what sense is it reasonable, given our desire to satisfy our curiosity, to engage in epistemic self-trust? Giving a practical reading of ‘reasonable’ won’t answer RST2 so construed.

Epistemic projects give us synchronic norms, compliance with which will satisfy our curiosity. They can do this because they can provide us with norms of the form ‘If C then φ’, where φ-ing when C obtains will satisfy our curiosity; in general, φ-ing when C will be taking the best route to having true beliefs. Such norms have triggering conditions, C; they give states of affairs such that S ought to φ whenever those states of affairs obtain. We can think of such triggering conditions as reasons to φ, if we’re sympathetic to views of reasons on which reasons for or against φ-ing are facts which bear on whether we ought to φ. But practical reasons will never figure in C, because practical reasons will never give us the best route to having true beliefs, even if they give us the best route to satisfying all of our desires. These norms give you the best way of aligning your attitudes with your evidence such that you’re in the best possible position to satisfy your curiosity at any given time; this should be distinguished from the sort of diachronic norms which would give the best possible way of conducting an inquiry to satisfy your curiosity. Both synchronic and diachronic norms are important, but if we’re interested in why we’re reasonable *now* to engage in epistemic self-trust, it’s the synchronic norms which are relevant to RST2.

I’d like to make a terminological point here by distinguishing between *compliance* with a norm and *satisfaction* of a norm. We satisfy a norm requiring φ-ing whenever C just in case it’s not the case that C obtains and we fail to φ. We comply with that same norm when we satisfy that norm *because* the norm requires φ-ing whenever C. Then epistemic projects give us norms which, if satisfied, will help us satisfy our curiosity. But to engage in an epistemic project, we’ll generally have to comply with those norms; if we’re intending to satisfy our curiosity, rather than merely incidentally doing so, by engaging in an epistemic project, we’ll have to satisfy the norms of an epistemic project by φ-ing precisely because φ-ing would satisfy those norms, and not purely by coincidence. So both satisfaction and compliance are necessary for engagement in an epistemic project, but only satisfaction is necessary for satisfying our curiosity.

Call this view
**Epistemic Project Evidentialism (EPE)** S’s belief that p is epistemically justified only if S participates in an epistemic project imposing a norm ‘If C then believe p’, and C.

To be clear, I’m not defending EPE as the only plausible view of the relationship between epistemic and practical reasons. All I want to say is that, even if we concede RP, RST2 introduces a context in which our epistemic projects are salient, and EPE is a plausible way of construing some of our obligations in that context. Because Rinard agrees that evidential reasons can be reasons to believe, the reasons given by an epistemic project will be a fragment of all of the reasons bearing on belief. But we can still attempt to answer RST2 restricting ourselves just to this fragment, leaving aside the possibility that there are more reasons which bear on belief than just this fragment. And answering that question is interesting, because we’re curious creatures and we want to satisfy our curiosity.

EPE might seem conducive to the view, such as Wright’s, that we are entitled to assume certain hinge propositions because they’re necessary to engage in epistemic practices. That sort of entitlement might be practical rationality. But entitlement is supposed to be an epistemic status; were it not, epistemic practices would be internally incoherent, and couldn’t be engaged in without massive dissonance. So taking this view of the relationship between epistemic and practical rationality doesn’t open the door to this notion of entitlement.

EPE invites what Coliva calls the ‘Oblomovian challenge’: what can we say to someone who would just opt out of any epistemic projects? If epistemic justification is just complying with norms which help us satisfy our curiosity, what can we say to someone who doesn’t share our curiosity? Coliva’s answer is that, although we can provide practical reasons for engaging in epistemic projects, the only response to the Oblomovian challenge we need is to point out that in engaging in epistemic projects we violate no epistemic norms. This isn’t available to me because I think we do violate epistemic norms just by engaging in epistemic projects; I’ll argue in Chapter Six that we can’t opt

---

141 Rinard 2019 sees evidence as playing an enabling role for forming beliefs. But there are also practical reasons corresponding to evidence; that p is highly likely to be true is generally a good practical reason to believe p.
144 See section 0.5.
143 Coliva 2015 p.145
144 Coliva 2015 pp.146-7
out of our epistemic projects even if we wanted to; if true, this would be enough to refute the Oblomovian challenge.

Coliva would agree that we need some sort of HOCTC to be rational, but that commitment need not be belief, because beliefs require evidential reasons to be rational. If it were belief, and our reasons for that commitment were practical, then our doxastic practices would involve fundamental inter-level incoherence, and it’s hard to see how it could then be rational to engage in them. That’s why Coliva rightly shifts attention to assumption rather than belief. In the next chapter I’ll consider such alternative attitudes to belief and argue that there is no sincere alternative to belief which isn’t subject to the same evidentialist concerns which we’ve examined in this chapter.

2.5 Conclusion

I’ve considered whether we might be reasonable in our higher-order commitment to the truth-conduciveness of our faculties because we have a justified belief that our faculties are truth-conducive. I’ve argued that three plausible ways of vindicating self-trust each fail on this assumption. Bootstrapping fails because it fails to raise the probability of the conclusion given the evidence on which its premises are based. A priori arguments fail because they can’t answer the Reidian challenge. Practical arguments fail because they ignore the implicit standards imposed by epistemic projects which are salient when we raise the question of RST2. It remains to consider whether we can vindicate self-trust by dropping the assumption that a higher-order commitment must be a belief, which I’ll address in the next chapter.
Chapter Three
Non-doxastic Alternatives to Belief

3.1 Introduction

Perhaps it’s hard to explain why it’s reasonable to believe in the truth-conduciveness of our doxastic practices because it isn’t reasonable to believe that they’re truth-conducive. Although we must have some sort of commitment to the proposition that a doxastic practice is truth-conducive if we are to rationally engage in that practice, it isn’t obvious that that such commitment need be belief.

Perhaps all that’s necessary for such commitment is lacking the disposition to deny that our practices are truth-conducive. This would accommodate the self-trust of animals and children nicely; even those without the conceptual resources to affirm the truth-conduciveness of their faculties can lack the disposition to deny their truth-conduciveness.

There would still be something strange, however, about being disposed to suspend judgment about TC. Someone who suspends judgment about p seemingly can’t appeal to p to justify their behaviour. So we need our attitude of commitment to satisfy the following constraint:

No Denial S’s attitude A to p is commitment to p only if S lacks the disposition to deny or suspend judgment on p.

No Denial seems to be necessary for commitment to TC. Is it sufficient? Children and animals can satisfy No Denial if they lack the concepts which would enable them to hold any attitude to TC at all, even suspending judgment. But we aren’t like children or animals; we’re normal mature adults who can form a judgment about TC. Even those normal mature adults who lack precise concepts about doxastic practices or truth-conducive transitions still have the capacity to form de re beliefs about their own beliefs; when a mature adult believes p, they can also form the belief that they’ve ended up with that belief without doing anything wrong, even if they don’t have precise concepts for the ways in which they end up with beliefs.

145 I’m thinking of suspending judgment as the attitude we form towards p when (1) we neither disbelieve nor believe p and (2) we lack any alternative affirmatory attitude towards p, such as those I’ll explore in section 3.
146 I intend ‘attitude’ to be taken in as thin a sense as possible, such that those who haven’t considered or couldn’t consider whether p can also have an attitude towards p if their behaviour is best rationalised by positing their having some attitude towards p.
So for us it’s not enough to satisfy No Denial in the way that children and animals do. We can’t satisfy No Denial trivially, by lacking the ability to have those dispositions at all. That means we can be forced to take a view on TC. I don’t mean that in some cases we must do at least one of: believe TC, disbelieve TC, suspend judgment on TC. What I mean is that we regularly act and reason in ways which are reasonable only if TC obtains. That means that our attitude towards TC must be different to our attitude towards propositions such as that emus are always heavier than kiwis. Not only do we not generally think about the relative weights of emus and kiwis, we aren’t committed to any view about their relative weights by our actions or reasoning in other areas. So although there may be a fourth option besides belief, disbelief and suspended judgment for our attitude towards TC, it can’t be refraining from taking any attitude at all towards TC, because it can’t be like the attitude we had (before the question was raised) about emus being heavier than kiwis. \(^\text{147}\) We can only satisfy No Denial if we can take some further affirmatory attitude towards TC.

We might try to capture this sense in which TC underlies the reasonability of our actions and reasoning by reference to the fact that we would appeal to TC if pressed. Alston has suggested the following attitude that attempts to capture this thought:

**Alstonian commitment (A-commitment)** A subject S with belief q is A-committed to p just in case S would appeal to p in S’s best available defence of q, were q to be challenged.

We might think that what S would appeal to gives the wrong result. Perhaps S is ill-equipped to judge what their best possible defence is. The best available defence will be the defence which justifies q to the greatest extent possible while minimising implausible premises, and not all of us will be equipped to provide such a defence. So we ought to modify A-commitment such that S is committed to p only if S should appeal to p in defence of q:

**A*-commitment** A subject S with belief q is A-committed to p just in case S should appeal to p in S’s best available defence of q, were q to be challenged.

I will interpret A*-commitment such that, if S does appeal to p, then at that point S must form a belief in p. The thought is that, having formed a belief that q on the basis of your epistemic faculties, were your grounds for that belief to be challenged, you would ultimately have to appeal to

\(^\text{147}\) See Audi 1994
the truth-conduciveness of those faculties in your defence. But where the challenge has not yet arisen, you need not have the belief in question.

So interpreted, A*-commitment isn’t yet enough to solve our problem, since on this view there are still circumstances under which it must be reasonable to believe that our faculties are truth-conducive – namely, those occasions when the question of the truth-conduciveness of our faculties is raised at all, and the subject is obliged to appeal to TC in their own defence.

Moreover, A*-commitment would be unreasonable if such a belief would be unreasonable though never in fact formed. If A*-commitment weren’t unjustified under such circumstances, it would impose an obligation on the subject to behave unreasonably, and entering into an obligation to be unreasonable must itself be unreasonable. If it weren’t, there would be a ready-made excuse available for all sorts of unreasonable behaviour. Simply engage in an obligation to act unreasonably, then proceed to do so, appealing to the prior obligation as an excuse.

Even if A*-commitment isn’t what we want, Alston’s thought could be extended in the following way: grant that relying on the outputs of our doxastic practices requires commitment to their truth-conduciveness in some sense of commitment, but cash that out in terms other than Alstonian commitment or belief. Alston’s considerations make it plausible that we don’t always have the belief that our faculties are truth-conducive – we don’t seem to use it as a premise in reasoning to beliefs that p, where p is an output of one of our doxastic practices, though some commitment to it is implicit in the rules we deploy in forming certain beliefs – so such a project isn’t obviously unfeasible. What we’d like is to find some notion of commitment which both (1) captures the sense that we incur some obligation to stand by the proposition that our faculties are truth-conducive, however we are to make sense of that obligation, and however we are to make sense of this notion of ‘standing by’, while (2) freeing us from some of the obligations placed on justified belief which we saw in the last chapter led us into difficulties.

Call an attitude which satisfies condition (1) an affirmative attitude, and call a non-doxastic attitude which satisfies both (1) and (2), if any exists, an ersatz belief attitude. Although I will argue that there are no ersatz belief attitudes – nothing satisfies our two desiderata – I first want to say why we should initially treat the existence of such attitudes as an open question. It seems plausible that there are attitudes like hope or faith whereby we form some commitment to a proposition but in a way distinct from the kind of commitment that comes with belief, and which may free us of some of the
obligations of belief. It’s easy to think of cases where someone may shape their life around religious belief in various ways – they may spend a good deal of their time thinking about religious matters, acting in accordance with the dictates of their religion, and so on – but lack some crucial features of belief, such as a willingness to assert the content of their religion. Such examples of faith seem common enough, and give some reason to think that there may be ersatz belief attitudes to be found. Perhaps, just as some Christians live in hope that Christ will redeem them, those of us who trust the outputs of our epistemic faculties similarly live in hope that they are truth-conducive, shaping our lives around that assumption, but lacking some crucial component of belief which might commit us to providing evidence for our hope.

3.2 Sincerity

3.2.1 The Importance of Sincerity

Let’s flesh out these two desiderata a little more. When we believe p, we have all sorts of dispositions. We’re disposed to act on p, feel that p is true, reason from p, assert that p, as well as acting, feeling, reasoning and asserting q where this is defensible in virtue of our belief that p.

The thought now is that we can conceive of an attitude of commitment which drops or modifies one or more of these features. There are many possible ways to drop or modify these dispositions, so I’ll first try to find some way of narrowing down our options.

One way of narrowing them down a bit would be to eliminate those attitudes which are insincere. When we talk about sincerity we are generally concerned with how action and assertion relate to belief; we’re talking about whether public signs of belief are genuine, or whether someone really does believe what they suggest they believe by how they are conducting themselves. But there’s also a sense of sincerity which arises, not through the relationship between an attitude and its public expression, but through the intrinsic properties of the attitude itself. Consider the following two cases:

148 Much of this section appears in Singh forthcoming.
149 Buckwalter, Rose and Turri 2015 proposes a division between thick and thin belief, with the former exhausted by taking p to be true, while the latter requires a further emotive or conative attitude towards p; it might be thought that the conditions I’ve listed omit the emotive and conative aspects essential to at least some varieties of belief. I am sceptical that we need such a distinction, and their examples seem well-understood as cases of taking-true without reliance in action, but in any case since thick belief entails thin belief, and the evidentialist worry is with thin belief in the reliability of our faculties, we can assume that any alternative to thin belief can be appropriately modified to serve as an alternative to thick belief.
150 See, for example, the use of sincerity in Wright 2016, where we’re concerned with sincerity in the sense that “A speaker’s testimony that p is sincere if and only if the speaker says that p because she believes that p”.

81
**Devil’s Advocate** In the short-term context of a seminar discussion, Asa commits to \( p \) for the sake of argument in order to tease out a response from others. Asa rightly judges that on the totality of his evidence the right attitude to hold to \( p \) would be to suspend judgment, but is prepared to assert and reason as though he believed \( p \) for the sake of the discussion.

**Iteration** Eli commits to \( p_0 \), believing with justification that \( p_0 \) is false but that after \( n \) iterations of adjustment in the face of objections Eli will arrive at a belief that \( p_n \), with \( p_n \) likely to be true.

Intuitively, the attitudes in Devil’s Advocate and Iteration are insincere; not only are they not belief, they don’t seem to be like religious faith or hope, either. Neither attitude seems to satisfy No Denial, and it’s their insincerity which lies behind that failure – if we’re insincere in our commitment to \( p \), then we’re liable to deny \( p \), at least in some circumstances. Moreover, sincerity looks like it’s important above and beyond being necessary to satisfy No Denial because it seems closely tied with the epistemic value of being committed to \( p \). We’re disinclined to credit the subjects in Devil’s Advocate and Iteration with any sort of epistemic good which obtains just in virtue of their holding that attitude. Were \( p_0 \) to turn out to be true then Eli would feel undeserving if he received credit for his commitment to \( p_0 \) in Iteration, and Asa would not feel entitled to any significant credit in Devil’s Advocate were \( p \) discovered to be true. Moreover, it’s the fact that Asa and Eli aren’t *sincerely* committed to \( p \) or \( p_0 \) in either of these cases which seems to lie behind our intuitions about credit in these cases; that suggests that we ought to expect epistemic value to attach only to attitudes which count as sincere commitment.

### 3.2.2 What is Sincerity?

Our grasp of sincerity is sketchy, at least when we try to apply that concept to attitudes beyond belief. In this section I’ll look at some options and make the case for my own account, which I’ll call the *cross-domain* account of sincerity.

First, sincerity seems closely tied to belief. So maybe insincere attitudes are just those which are compatible with disbelief. The problem with classifying playing Devil’s advocate, say, as sincere is that it looks like you can play Devil’s advocate while disbelieving the proposition. It seems intuitively insincere to do so even when you merely lack belief. But this is unsatisfactory as an account, because (1) it might be that our intuitions about such cases aren’t about the attitude itself
but about the public manifestations of that attitude – the assertions which you make when playing Devil’s advocate, say – which in typical conversational contexts may represent you as believing the proposition you’re defending\textsuperscript{151}; and (2) it isn’t explanatory, because we don’t know why belief (of all sincere attitudes) is special. Is sincerity a concept picked out with belief as a paradigm case and, if so, what are the salient features of belief which we’re generalising from? This sort of account can’t tell us what those salient features are, and so isn’t explanatory.

3.2.2.1 Assertion Sincerity

Let’s look at some ways of accounting for sincerity which are more explanatory. I’ll first consider Sanford C. Goldberg’s account of sincerity. Goldberg characterizes sincerity in terms of conversational norms, offering the following condition:

**Sincere Assertion** An assertion that $p$ made in a conversational context $C$ by $S$ is sincere just in case it wouldn’t give an audience good reason to think $S$ has a doxastic attitude towards $p$ which $S$ doesn’t in fact have, given the mutual expectations any participants in $C$ could reasonably have.\textsuperscript{152}

In considering Sincere Assertion I’m diverting the discussion towards linguistic notions, but the target remains finding a good account of sincerity for certain sorts of mental states. We can use Sincere Assertion to get at such an account because Sincere Assertion induces a corresponding sincerity condition for non-belief commitment:

**Assertion Sincerity** $S$ sincerely commits to $p$ if and only if $S$’s commitment that $p$ wouldn’t dispose $S$ towards insincere assertions.

But Assertion Sincerity has unintuitive consequences. Consider the following case:

**Pro Forma Disclaimer** Simran is a physician who regularly performs a procedure $P$. Simran lives in a highly religious country in which the prevailing religion deems $P$ impermissible and Simran is legally required to claim that $P$ is dangerous—a legal requirement put in place solely to dissuade residents of this country from undergoing $P$. Simran is not an exceptionally dutiful physician and has no belief about whether $P$ is dangerous. Because the disclaimer is known to be merely pro

\textsuperscript{151} See Goldberg 2013.
\textsuperscript{152} Goldberg 2013 p.289
*forma*, mandated for religious and not medical reasons, any audience to which Simran would deliver the disclaimer is unlikely to think Simran believes that P is dangerous.

Pro Forma Disclaimer seems possible; filling in the details in the right way would generate at least one real-world case. And in Pro Forma Disclaimer, Simran doesn’t seem to be sincerely committed to the dangerousness of the procedure. But Simran is asserting that the procedure is dangerous without giving an audience good reason to believe that Simran believes that the procedure is dangerous. So Goldberg’s account seems to give the wrong verdict in this case.

It might be denied that Simran is making an assertion here; perhaps the force of her utterance is analogous to actors’ utterances, and is a sort of fictional performance of an assertion. But we can construct cases where this seems unlikely; consider:

**Notorious Liar** Oishi is known to be a consistent liar, such that when Oishi’s audience hears Oishi assert that p they form the belief that Oishi believes ~p. In fact on this occasion Oishi believes p, and attempts to assert as much by uttering ‘p’.

Notorious Liar is the converse of Pro Forma Disclaimer; here, Oishi’s commitment to p is intuitively sincere, but would be deemed insincere on Goldberg’s account. Moreover, there is no concern here that Oishi is making any sort of fictive assertion, since there are no special conversational norms in place which would make that plausible. Since Assertion Sincerity seems unable to rule out these cases, we should reject Assertion Sincerity.

### 3.2.2.2 Right Reasons

Let’s consider an alternative more clearly directed towards ruling out cases like Devil’s Advocate. Perhaps being sincere means being appropriately sensitive to the right sort of reasons:

**Right Reasons** S sincerely holds an attitude A to a content φ just in case S holds A to φ in virtue of possessing (the right sort of) reasons in favour of holding A to φ.

Right Reasons might be suggested by Wright 2016, in which sincere *assertion* is explained as assertion causally connected to justification rather than causally connected to belief. Perhaps in a similar way, sincere attitudes are those which are based on apprehension of the right sort of reasons.
This makes sense of the tight connection between sincere commitment and belief, and explains how it’s tempting to give an account of the former in terms of the latter: in doing so we’re apprehending the common structure which all sincere committal attitudes have, and treating belief as a paradigm case of that structure, namely, being based on the right sort of reasons.

Which reasons must sincere attitudes be based on? It can’t be just any reasons which S takes to speak in favour of holding A to φ—that could render almost any attitude sincerely held, but the pre-theoretical sense of sincerity I’m trying to get at is one on which it’s never possible to sincerely play Devil’s advocate. It’s tempting to say that they must be based on direct epistemic reasons. Someone who commits to p because it’s socially convenient is being insincere, and being insincere because they’re disregarding any connection between what they commit to and what’s true. Moreover, it’s tempting to think that sincere attitudes must be based on all of the available direct epistemic reasons. Someone who commits to p on the basis of only a proper subset of the direct epistemic reasons can’t be sincere—not if they do so knowingly, anyway. Curation of one’s commitments on the basis of what’s most likely to be true seems to just be what being sincere is. In this sense again there seems to be a tight connection between sincere commitment and believing, because sincere commitment seems to involve the attention to all and only direct epistemic reasons which is characteristic of belief. If that’s right, then all sincere commitments must involve belief as a component; they might have additional affective parts, but all sincere commitment involves commitment to a content just on the basis that the totality of reasons point towards its being true, and that constitutes believing that p.

But this sort of account of sincerity would beg the question against the use of ersatz belief attitudes under consideration. We’re interested in whether there might be any sincere committal attitudes which lack all of the norms applicable to belief. Plainly there are very many cognitive attitudes which lack such norms—wishing that p is one, for example. Perhaps some sincere committal attitudes are like wishing, rather than like belief, in being free of those norms. I don’t think it’s plausible to rule out that possibility by stipulating that sincere commitment must involve sensitivity to solely epistemic reasons. In the end it might turn out that sincere commitment must be a solely epistemic attitude—it might turn out that there can be no sincerely committal attitudes which are partially based on practical reasons—but we can’t decide that at the conceptual level when giving an account of sincere commitment, but instead only after we’ve applied substantive norms of justification to an independently formulated account of sincerety.
3.2.2.3 The Partial Sincerity Account

We might think that we can accommodate the intuition behind Right Reasons while weakening the requirement that right reasons be the totality of epistemic reasons directly bearing on the truth of the proposition. Zach Barnett suggests

**Partial Sincerity** S is at least partially sincere in holding A to φ just in case S holds A to φ either on the basis of the totality of S’s evidence, or on the basis of S’s evidence once (some or all) of S’s higher-order evidence has been bracketed.\(^{153}\)

The thought is that sincerity is supposed to tell us how far an attitude is rightly-held by one’s own lights. Consider the following two cases:

**First-order Views (FOV)** Two epistemic peers Rachel and Ruth commit to φ and ~φ respectively for some controversial thesis φ. Each are aware of what the other is committed to and that the other is an epistemic peer. Rachel and Ruth each make their commitment just on how the relevant arguments and intuitions strike them, disregarding higher-order evidence of disagreement.

**First-order Swapped (FOS)** Epistemic peers Samuel and Saul commit to φ and ~φ respectively. Samuel is committed to φ solely on the basis of their knowledge that Saul’s evaluation of the first-order evidence speaks in favour of φ, while Saul similarly commits to φ solely on the basis of their knowledge that Samuel’s evaluation of the first-order evidence speaks in favour of ~φ.

Then, Barnett would say, Rachel and Ruth are partially insincere in their commitment, because their commitment isn’t based on the totality of their evidence. But they are intuitively more sincere than Samuel and Saul. The best way to explain the intuitive gradations in sincerity that can be detected in the difference between FOV and FOS is to endorse Partial Sincerity. If that’s right, then perhaps the notion we’re looking for isn’t sincerity but partial sincerity. The goal is to exclude attitudes like playing Devil’s advocate, asserting that p for the purposes for social kudos, and the like. But that can be done equally well with the notion of partial sincerity. If Asa commits to p for the purposes of playing Devil’s advocate then he does so not on the basis of how his evidence strikes him but for the purposes of having a fruitful discussion, displaying his mastery of the material, or whatever. Hence Asa isn’t even partially sincere in his commitment.

\(^{153}\) Barnett 2018 pp.122-3
As Barnett says, we can set any body of evidence aside and check what our inclinations would be, on that basis. That won’t generally get us even partial sincerity. It seems as though bracketing higher-order evidence generally will, including higher-order evidence of one’s unreliability through drug-taking as well as the sort of evidence of unreliability provided by disagreement. What we’re left with after bracketing (some or all) higher-order evidence is somehow still ‘psychologically gripping’, and ‘we will still have something substantial to say in support’ of the proposition arrived at on the basis of our bracketed evidence; in this way our attitude ‘bears important hallmarks of sincerity’.

The problem is that such a restriction is too arbitrary. Why can’t we arbitrarily discount higher-order reasons from certain sources, with certain contents, or with unfortunate consequences, or in some other way restrict the set of epistemic reasons that directly bear on the truth of the content of a content? Intuitively the resulting attitude wouldn’t just be irrational but often insincere. For example, Partial Sincerity seems to declare that in the following case of commitment is partially sincere:

**Arbitrary Restriction** Leo commits to p on the basis of evidence arbitrarily restricted to his first-order evidence, plus evidence of disagreement from right-handed people.

Similarly, in the following case Marvin appears to be partially sincere:

**Academic Equity** To balance the unfair down-grading of philosophical testimony from a particular marginalized group, a philosopher Marvin endeavours to be committed to views which are likely on the arguments and evidence offered by philosophers who belong to that group. Marvin commits to p in virtue of p’s being likely on the body of evidence restricted to arguments and evidence offered by that group in respect to p, as well as Marvin’s first-order evidence.

I take it that in Arbitrary Restriction Leo is plainly wholly insincere in his commitment. My intuition is that the same applies in Academic Equity, but it’s perhaps not obvious that this is so. But there seems no clear motivation for excluding Arbitrary Restriction or Academic Equity as cases of sincere commitment while retaining Partial Sincerity. Unless we place some constraints on

154 Barnett 2018 p.122
permissible restrictions of evidence, Partial Sincerity doesn’t give us much help in filtering sincere from insincere commitment.

Barnett could supply the following constraint on restrictions to evidence:

**Epistemic Restriction** If S commits to p on the basis of a restricted set of evidence, then S’s commitment to p is sincere only if the reasons for those restrictions were solely epistemic.

Epistemic Restriction is motivated by the considerations Barnett offers in favour of his attitude of disagreement-insulated inclination. We’re interested in holding non-belief commitment for the purposes of securing some sort of epistemic benefit; in the case of disagreement-insulated inclination, the indirect benefit of a diverse epistemic community. There is a sense in which such reasons, although not directly truth-related, are epistemic in a way that holding a commitment for social kudos isn’t.

Because there are no good epistemic reasons to restrict the evidence set in Arbitrary Reasons, Epistemic Restriction rightly declares Arbitrary Reasons a case of insincerity. It also captures the sense in which Academic Equity isn’t such a sharp case of sincerity or insincerity. The case is underdescribed because it doesn’t determine whether the evidence set is restricted for epistemic or practical reasons; Marvin might be taken solely as trying to promote the epistemic goods which would be expected to come with elevating unjustifiably neglected philosophical testimony, or trying to secure non-epistemic moral goods. 155 On the first interpretation Epistemic Restriction would predict that Marvin would be sincere, while on the second it would predict insincerity. So Epistemic Restriction coheres with the intuitive data here.

But Epistemic Restriction is unhelpful for explaining how non-belief commitments can accrue epistemic value. It looked like sincerity might be a promising way to understand how commitments can achieve this. But Epistemic Restriction doesn’t elucidate partial sincerity in a way that would allow us to do this, because it writes achieving epistemic value directly into the analysis of partial sincerity. So even if Epistemic Restriction gets partial sincerity extensionally right, it doesn’t get the explanatory role of sincerity right. It’s not enough to say that partially sincere non-belief

---

155 The case is tricky because it might be thought that in such cases moral reasons also bear on the relevant epistemic reasons. If that’s right this would be a further reason, given Epistemic Restriction, to expect our intuitions to be muddled in Academic Equity, so Epistemic Restriction still makes a prediction which coheres with the intuitive data.
commitments just are those non-belief commitments which we can have and satisfy condition (2) if we want to explain how any non-belief commitment can satisfy condition (2) at all. I’ll turn now to a condition on sincerity which doesn’t presuppose that sincerity is epistemically valuable and so can do that explanatory work.

3.2.2.4 The Cross-Domain Account

What seemed to go wrong for Goldberg’s attitude is that it licensed various dispositions relating to endorsing p only in certain contexts. If you can be sincere in asserting that p because there’s no danger of anyone mistaking your true beliefs with respect to p, and this is all that sincerity amounts to, then it seems that you can sincerely commit to p if you have license to assert p in some contexts but not all. But there seem to be very many cases of licensed assertion of p in some contexts without sincerity. Sincerity has to be more than what Goldberg’s suggesting, and I’d suggest that we take as an alternative account the property which Goldberg’s account seems intuitively to be missing: the property of being treated as a premise in all contexts, not localised to some particular context where special norms of assertion may apply. Call this the cross-domain picture of sincere commitment.

Here are some initial considerations in favour of this picture. If we don’t adopt the cross-domain view of sincerity, we’ll be able to generate all sorts of attitudes satisfying Assertion Sincerity, just by having increasingly Byzantine conversational norms in place. We could equally induce conversational norms by fiat, perhaps in a small group of peers for the purposes of group bonding, or to signal status. We’d like to be able to say that dispositions to make such assertions don’t constitute sincere commitment, but in what could this sincerity consist? Not aligning with a pre-existing psychological attitude—that attitude is just what we’re trying to find, so we can’t appeal to it in our account of sincere assertion. The alternative, I’d suggest, is that what makes that assertion insincere is the fact that outside of specific contexts where unusual conversational norms or presuppositions are in play, we wouldn’t be prepared to make that assertion, even if directly questioned. That you aren’t consistent in your assertions is often taken as a good sign of insincerity; in the absence of a pre-existing psychological attitude to which we can appeal in accounting for insincerity, it seems plausible that inconsistency can also be what constitutes insincerity.

Similarly, if you only need to have a disposition to act on your commitment to p in certain cases—perhaps where the costs are low, and plausibly outweighed by the costs of degrading your reputation as, say, an important philosopher—then even those who are intuitively insincere will be able to pass
the test while acting perfectly rationally. In cases like these, what seems to make the subject insincere isn’t that they’ve failed to live up to a special psychological attitude of philosophical commitment, but that they act as though a proposition p is true in one context but as though it is false in another.

Sometimes we seem to sincerely believe that p but don’t have such general dispositions. For example, we might be prepared to assert that S is tall in some contexts but not in others (for example, when reaching the top shelf is salient versus playing basketball). In such cases we have context-sensitive belief, but we needn’t explain that by denying the cross-domain account. Rather, the concept ‘tall’ is context-sensitive, and hence the believed content is, too, and we can explain context-sensitive belief-dispositions by reference to the context-sensitivity of the believed content. To rule out such cases, I want to make two definitions:

**Non-localised disposition** S has a non-localised disposition to φ just in case S is disposed to φ in any context where S possesses the same reasons which bear on the rationality of φ-ing.

**Concept-grounded context-sensitivity** A content p has concept-grounded context-sensitivity just in case S’s dispositions as a result of holding A towards p are non-localised just in virtue of the concept(s) which partially constitute p.

The account of sincere commitment I’m proposing can now be given as:

**Cross-Domain** S sincerely commits to p if and only if either (a) S has a non-localised disposition to do at least one of the following: assert that p, reason from p, act on the basis that p, and feel that p is true, or (b) S lacks (a) solely in virtue of p’s having concept-grounded context-sensitivity.

Cross-Domain rules out the examples I used originally to introduce the notion of sincerity. It rightly declares that S is insincere in Devil’s Advocate, because Asa fails to have the same dispositions outside the seminar room as inside. Similarly, in Iteration, Eli will drop dispositions when he leaves the context of inquiry, so Cross-Domain declares Eli’s commitment insincere.

This also gets the right results for Pro Forma Disclaimer and Notorious Liar. In the former, Simran is rightly ruled insincere because Simran wouldn’t assert that p outside of the context of giving the disclaimer before the medical procedure. In the latter, Oishi is rightly ruled sincere, because Oishi is
disposed to assert and act on p under any circumstances, even though she lacks a receptive audience.

It can also accommodate the ambiguity found in Academic Equity. In such cases it’s ambiguous whether the commitment in question extends solely to Marvin’s academic work, and hence is contextual in the way ruled out by Cross-Domain, or if Marvin extends his commitment to the philosophy of marginalized groups to his extra-academic ethical and political life. On the latter interpretation Cross-Domain would declare Marvin’s commitment sincere. So Cross-Domain can account for our mixed intuitions about cases like Academic Equity.

One problem with Cross-Domain is this: there seem to be context-sensitive sincere beliefs which lack concept-grounded context-sensitivity. In particular, beliefs which are sensitive to stakes can be context-sensitive without possessing concept-grounded context-sensitivity. For example, we might sincerely believe that the bank closes later than 5pm when the stakes are low, but fail to have that sincere belief when the stakes are high. Our dispositions will then be context-sensitive, because we will visit the bank first on our round of chores when the stakes are high, but postpone our visit if the stakes are low. But this seems to have nothing to do with any concept-grounded context-sensitivity in the content. The context-sensitivity appears extraneous to the content and at least partially grounded in the variability of the stakes. Similarly, it’s plausible that Goldberg’s attitude might be sincere while being context-sensitive, because the stakes of being wrong about philosophical issues will vary between the seminar room and the outside world.

This is a general problem which the Cross-Domain account has to address. As stated, Cross-Domain says nothing about cases where S encounters new reasons which bear on their commitment and adjusts their dispositions accordingly. When we enter a high-stakes situation we gain new reasons not to hold certain beliefs, because we encounter practical reasons which raise the threshold for belief. But the same problem will arrive for encountering new epistemic reasons. If you enter a new context in which you encounter fresh evidence against p, then you will rightly drop your disposition to act on the basis that p, assert that p, and so on.

But this isn’t yet a counter-example to Cross-Domain, because Cross-Domain is couched in terms of non-localised dispositions which (I’ve stipulated) are dispositions which hold in all contexts.

156 On this interpretation S might be irrational, but that doesn’t bear on whether S is sincere.

where an agent’s reasons remain constant. That isn’t the case in bank cases nor in cases where an agent acquires new evidence.

We might worry that if counter-examples can be dodged by introducing new reasons in each new context, Cross-Domain risks becoming empty of content. But Cross-Domain isn’t empty; we’ll see below that it will rule out a number of candidate ersatz belief attitudes.

It might be objected that Cross-Domain is false because it wrongly declares fragmented belief insincere.\(^{158}\) Let a belief-set be *active* in a context C just in case we are disposed to treat its beliefs as premises for reasoning, action etc. while in C. Then our beliefs are *fragmented* just when we have multiple inconsistent belief-sets which are active in different contexts. For example, suppose Geoff is prepared to reason, act etc. on the basis that London Bridge is a short walk from Borough when Geoff is walking around the area, but when Geoff is taking the Underground he doesn’t act on that basis, relying automatically on the Underground map, which misleadingly suggests they are further apart. Then Geoff’s belief-set is fragmented – he believes that London Bridge is a short walk from Borough in some contexts but not in others. Nonetheless, while Geoff might be irrational, Geoff doesn’t appear to be insincere; if Geoff were asked for directions to Borough, Geoff might be liable to give bad directions, but Geoff wouldn’t be lying, and Geoff would be prepared to act on his false belief even in high stakes situations. Cross-Domain seems to say that Geoff is insincere, so Cross-Domain seems to be false.

If we’re sympathetic to fragmented belief, we ought also to be fragmented to fragmented reasons. Fragmented belief is supposed to be the sort of thing which is more justified than wilfully inconsistent beliefs. But that can only be if fragmented belief sets are based on a proper subset of a subject’s total reasons. Those subsets will be available for belief-formation only in certain contexts. So we ought to admit fragmented reasons if fragmented belief is to be plausible. Geoff’s total reasons include his mental map of the streets of London, and the map of the London Underground, but only one of these reasons is active in each context.

Then it isn’t *ad hoc* to adjust Non-localised disposition such that S has a non-localised disposition just in case they possess the same disposition in the same contexts where they possess the same *active* reasons. The intuition behind the non-localised disposition component of Cross Domain is that our dispositions can vary while being sincerely held if our reasons change, because our varying

---

\(^{158}\) For more on fragmentation, see Lewis 1982, Egan 2008, Norby 2014.
dispositions are the result of responding to our changing reasons. But that same intuition means that varying dispositions in the face of varying active reasons ought also to be sincere. In both cases the thought is that when the considerations in favour of φ-ing change, it says nothing about our sincerity if our attitude towards φ-ing changes. So Geoff is sincere because Geoff doesn’t change his fragmented belief about the location of Borough when moving from compartment to compartment, even if a different fragmented belief would become active if he left the Underground entirely and thereby activated previously inactive reasons.

On the face of it there are still some non-doxastic attitudes which satisfy Cross Domain, such as acceptance and assumption. The strategy in the rest of this chapter will be to argue that those attitudes which do satisfy Cross Domain thereby incur an evidentialist obligation. Those attitudes which are able to meet that obligation will be too weak to capture the sort of commitment we have to truth-conduciveness. Those attitudes which don’t meet that obligation will be unjustified for just the same reason as belief is, so the retreat to a non-doxastic attitude will have failed.

3.3 Sincere Non-Doxastic Commitment

3.3.1 Hope

I’ll now look at some candidate ersatz belief attitudes, arguing that they can’t satisfy (1) and (2) because they’re insincere and hence can’t have intrinsic epistemic value. First, I’ll consider hope. Charles Sanders Peirce suggested that hope is the right attitude to capture the sort of attitudes which scientists have towards the arrival of a future consensus in their field.\footnote{See Peirce 1992a, 1992b, 1998.} Consider various scientists, all pursuing various lines of inquiry; some are using the same tools to measure different things, others different tools to measure the same things, and others overlap partially in these respects and differ partially in these respects. So two neo-classical economists might be investigating different subject matters but using the same models, while a behavioural economist might seek to understand some of what the first economist studies and some of what the second, but is sceptical about the applicability of their model, and hence uses different tools to further his investigations. Peirce points out that we are typically prepared to affirm (in some sense) that all these inquiries will eventually terminate in some consensus across all the areas the investigators are interested in; those engaged in such investigations, in particular, generally have some such affirmatory attitude towards a final convergence. There’s some commitment to the thought that, as the investigation progresses,
methods will be dropped outright, others adapted, until there is general agreement about the correct methods and the acceptance of all the results which are generated.

If we were precipitate we might believe that such a consensus will eventually emerge. But Peirce suggests that we can instead hope that that consensus will emerge: “All the followers of science are animated by a cheerful hope that the processes of investigation, if only pushed far enough, will give one certain solution to each question to which they apply it...No modification of the point of view taken, no selection of other facts for study, no natural bent of mind even, can enable a man to escape the predestinate opinion. This great hope is embodied in the conception of truth and reality.”

If hope is a plausible alternative attitude for the view we take towards an eventual consensus, then it could be the sort of attitude which we have towards TC. There seems something phenomenologically right about hope here – it captures the way in which we lack the strength of commitment that comes with full belief, but is sufficiently positive that it could plausibly ground treating TC as a premise in reasoning and action. Call this

**Hopeful Commitment** When S has a higher-order commitment to the truth-conduciveness of their doxastic practices, S hopes that their doxastic practices are truth-conducive.

To evaluate Hopeful Commitment we need to get a sharper understanding of hope. I’ll now look at two ways of accounting for hope.

3.3.1.1 Superficial Hope

There is a minimal notion of hope – Philip Pettit calls it the “superficial” or “lowest common denominator” account of hope – which S exhibits just when (1) S believes that some state of affairs p is (epistemically) possible, but doesn’t believe with certainty that it obtains (2) S desires that p. This might be roughly what Peirce had in mind, since it appears that such an attitude would

---

160 Peirce 1893 CP 5.407. Emphasis mine. Similarly, in Peirce 1998, “The different sciences deal with different kinds of truth; mathematical truth is one thing, ethical truth is another, the actually existing state of the universe is a third; but all these different conceptions have in common something very marked and clear. We all hope that the different scientific inquiries in which we are severally engaged are going ultimately to lead to some definitely established conclusion, which conclusion we endeavor to anticipate in some measure”.

161 Pettit 2004 p.154

162 This is intended as a timeless use of ‘obtains’, so as to include hoped for events in the past and future.
be readily held in the sorts of cases where he wants hope to be held. It seems clear that eventual consensus for any given investigation is at least epistemically possible – if you have an opinion on a topic, and others disagree with you, it seems possible, given everything you know, that those who disagree with you would eventually come to share your opinion, or perhaps that you and everyone else will change opinions to another point of view. So long as holding an opinion is (metaphysically) possible for anyone, it seems (epistemically) possible for everyone, at some indefinite future date. On the other hand, I doubt many people engaged in any inquiry feel confident that it will resolve itself into perfect consensus. Finaly, it seems that investigators do desire eventual consensus. So all the ingredients are there for this minimal attitude of hope to be held quite commonly towards the eventual convergence of opinion by investigators in any given topic.

Superficial hope is also sincere, in the sense defended above; it satisfies Cross Domain so long as the belief and desire which comprise a case of hope satisfy Cross Domain themselves. That suggests that superficial hope can have epistemic value, or at least doesn’t suffer the same obstacles to having epistemic value as we saw in Devil’s Advocate and Iteration.

But Hopeful Commitment must be false on the superficial hope reading because it doesn’t capture the unique commitment which S holds towards the truth-conduciveness of their faculties, a commitment which is supposed to distinguish the attitude S takes to their own faculties from that they take to doxastic practices which they aren’t engaged in. Superficial hope isn’t exclusive; it’s possible to hope that two distinct and mutually incompatible doxastic practices are both truth-conducive. For example, it seems likely that our actual practice of sense perception is unreliable in some ways, in the sense that it generates illusions which it’s easy for the non-expert not to notice. Suppose I’m disposed to make a certain faulty transition T when objects are replaced on my peripheral vision. Then, knowing this, it’s epistemically possible for me that any number of alternative doxastic practices, with T replaced with different transitions $T_1$, $T_2$ etc., could be more reliable than my current practice of sense perception. Call these practices $D_1$, $D_2$ etc.

---

163 This is slightly complicated by ‘trivial’ investigations. If I’m trying to find my keys, do I believe with certainty that this investigation will terminate in a consensus, since the only person involved – me – will eventually come to have a belief about where my keys are? Then I can’t have the hope, and hence my investigation isn’t truth-directed. I suspect this isn’t a real problem for Peirce – where we can have a full belief that investigations will terminate in consensus, that suits his purposes just fine; it’s only when that isn’t the case, as seems to be true most of the time, that we must have recourse to the notion of hope. But hope could be merely a fallback position.

164 I discuss the exclusivity of self-trust in section 1.3.
Then for each $D_i$, I will desire that $D_i$ is truth-conducive, because if $D_i$ is truth-conducive then my own doxastic practice is highly likely to be truth-conducive – my own doxastic practice differs from $D_i$ only with respect to one transition. Not only does that raise the probability of my doxastic practice being truth-conducive overall, it means that for all beliefs I form that don’t involve that transition, those beliefs will be formed through a truth-conducive practice. This is a desirable state of affairs, assuming I want my beliefs to be formed through truth-conducive practices. Note that I don’t desire that $D_i$ is truth-conducive because I think my practice is identical with $D_i$, or because I think it’s equivalent in some way, but rather because $D_i$’s being truth-conducive means that my own, distinct, doxastic practice will be truth-conducive in very many, perhaps all, situations.

Not everyone will go through this process of reasoning, so perhaps not everyone will desire that $D_i$ is truth-conducive. But that a subject could so easily go through such a process of reasoning, and do so seemingly rationally, suggests that superficial hope isn’t the right attitude. Commitment to truth-conduciveness should be an exclusive attitude – it shouldn’t be the case that it’s compatible with commitment to the truth-conduciveness of all sorts of other practices which we don’t engage in. But if we could very easily, and perfectly rationally, form exactly the same commitment to the truth conduciveness of each $D_i$ as to the truth-conduciveness of our own doxastic practice, then our commitment wouldn’t be exclusive. That’s because we can reasonably hope that $D_i$ is truth-conducive, for every $i$ and for every transition in $i$, but as I argued in Chapter One the exclusivity of our commitment to our own doxastic practices requires us to be committed to every transition in our own practices but not to every transition in related practices which differ with respect to particular transitions. So superficial hope can’t be higher-order commitment to truth-conduciveness, because it lacks this exclusiveness requirement.

Not only does superficial hope fail to capture the exclusivity of higher-order commitment, it also fails to satisfy No Denial. It’s easy to have the disposition to disbelieve $p$ while believing $p$ is possible and desiring $p$. While it might not be possible to be justified in so doing, it seems we very often do disbelieve $p$ while acknowledging that there are epistemic possibilities where $p$ which we can’t rule out; still, we regularly disbelieve propositions which are sufficiently unlikely that we expect them to never occur.

Note that the failure of superficial hope to satisfy these two desiderata has arisen from the requirement that it satisfy Cross Domain; to do so we’ve chosen an attitude which has belief as a component, because beliefs generally satisfy Cross Domain. But by including belief as a part of the
attitude, our attitude is subject to the same norms of justification which caused problems in the last chapter; in particular, it’s subject to an evidentialist constraint, one which can only be met by ensuring that the content of our belief isn’t TC but it’s possible that TC. But that it’s possible that TC is just too weak a content to do the job of commitment to TC, and the attempt to satisfy Cross Domain and have a plausible alternative attitude to belief has failed.

This illustrates the way in which a broad class of attitudes won’t work for our purposes. Any alternative attitude which tries to satisfy Cross Domain by including belief as a component of that attitude won’t capture commitment to TC because the content of that belief will have to be different to TC. Any such alternative attitude seems vulnerable to counter-examples – just pick a case where believing TC is importantly different to believing the candidate alternative content. So the strategy behind superficial hope probably can’t be patched up – if any non-doxastic alternative attitude is going to work, it’ll have to be solely at the level of the attitude involved, not through changing the content as in superficial hope.

3.3.1.2 Pettit’s Hope

So it looks like superficial hope isn’t what we want. I’ll look now at Pettit’s alternative. Pettit’s account is more restrictive than the superficial view. Rather than merely believing p is epistemically possible and desiring p to obtain, on Pettit’s account we only hope when our confidence that the content will or won’t obtain is low enough to run the risk that we will fail to act on the basis of our (weak) belief.\textsuperscript{165} Were we to rely on belief alone as the cognitive attitude affirming the likelihood of success in our endeavours, it wouldn’t be up to the task, and we would abandon our efforts. It’s common in such cases to talk of our ‘losing hope’ that we will achieve anything. The risk of such a failure to act prompts the agent to form an affirmative attitude towards the possibility of success, the same attitude of practical acceptance which Michael Bratman\textsuperscript{166} describes, and this attitude is hope. Call this attitude Pettit hope, or P-hope.

Practical acceptance of p, for Bratman, is holding p fixed for the purpose of forming a plan or performing an extended bit of practical reasoning. In forming a plan to go to Southwark, for example, you might form a plan on the basis that certain buses will arrive (roughly) at certain times. Of course what we accept in such contexts might also be what we believe. But it need not. In

\begin{footnotesize}
\begin{enumerate}
\item[165] Pettit 2004 p.157
\item[166] Bratman 1992
\end{enumerate}
\end{footnotesize}
particular, what we accept might differ from belief in four key ways: it need not be context independent; it need not be shaped by evidence or ‘aim at truth’; it can be voluntary; and it need not be able to cohere or be integrated with one’s other views, including other acceptances. Context matters, because, when you return to your job as bus supervisor, you can no longer take for granted as part of your practical deliberations that the buses each arrive within a certain window, because it may be part of your job (and hence part of your practical deliberations) that there is a chance that the buses will not and hence need monitoring.

Bratman’s acceptance isn’t enough to get us trust, because the context dependence of Bratman’s acceptance means it doesn’t satisfy Cross Domain.167 This is why P-hope matters: here, we use Bratman’s acceptance precisely to maintain a general and long-standing commitment to p, and hence one which permits deployment in multiple contexts. That we maintain a fixed commitment to p implies also that we have some obligation to integrate our commitment with our other views, for, given that every view we have has some bearing on some context, we are obliged (for fear of incoherence) to ensure our commitment to p is compatible with each of our other views; the process of making our commitment compatible with all of our other views is one way of satisfying Cross Domain.

Moreover, it looks like P-hope avoids the problems with the superficial attitude discussed above. There, the attitude was presented as constituted by having certain beliefs and desires; moreover, the requirements were such that many people will hope for incompatible things just because they have the appropriate beliefs and desires, and hence the superficial attitude couldn’t carry the exclusivity needed by both Peirce and Alston. P-hope isn’t reducible to belief and desire, and rather than being constituted is instead caused by the appropriate belief and desire states. Presumably those states aren’t by themselves sufficient to cause the corresponding hope; otherwise, we would have almost as many attitudes of hope as we did with the superficial account. It seems likely that only under special circumstances will the belief and desire give rise to hope, and in particular it will tend to be only when those states are made especially salient or the agent is consciously aware of them that they will cause hope to arise. Moreover, once hope is in place, it is plausibly exclusive, since it commits one to acting as though its content were likely to be true, and that would be impossible in many cases where two contents are inconsistent. So we could only hope for one kind of doxastic

---

167 This context dependence isn’t solely because whether S Bratman accepts p can vary based on S’s reasons, but because S’s institutional role in itself makes Bratman acceptance reasonable or not. Cross Domain rules out Bratman acceptance as sincere in the same way it rules out Goldberg’s seminar commitment as sincere – if a change in institutional context on its own can change whether it’s reasonable to have an attitude, then it fails Cross Domain.
practice to be truth-conducive at any one time. Putting all this together, since it isn’t the case that hope is generated immediately upon generation of belief and/or desire, and that once a hope is in place, it blocks any incompatible hope from being adopted alongside it at a later date, Pettit’s hope is appropriately exclusive for commitment to TC.

But the tension between the features of Bratman’s acceptance and P-hope seem to make this infeasible as an alternative for belief in trust. Bratman’s acceptance can plausibly drop the commitment to truth-directedness because it lacks the context-independent features of belief. As I will argue in the next section, these arguments still go through for Cohen and Alston’s acceptance. Although they don’t for Bratman’s acceptance, that’s because Bratman’s acceptance, on its own, would lead to a form of trust which inherently leads to insincerity; the attempt to use P-hope to avoid this problem leads to the same difficulties facing other forms of acceptance.

3.3.2 Acceptance

According to Cohen, to accept that p is “to treat it as given that p”. What Cohen means by this is that accepting that p means that you “have or adopt a policy of deeming, positing or postulating that p – i.e. of including that proposition or rule among one’s premises for deciding what to do or think in a particular context, whether or not one feels it to be true that p”\(^{168}\). The contrast with belief is intended to be that belief comes with the disposition to feel true, whereas acceptance only involves the other three of four dispositions with which I began 3.2\(^{169}\).

Could it be that acceptance is reasonable without evidence, in the way that belief isn’t? First, we should note that acceptance isn’t the same thing as hypothesizing. You are reasonably free to hypothesize at will, for the sake of exploring some topic. But acceptances are supposed to be based on reasons in a way that hypotheses aren’t; we can hypothesise without any reason at all, but we can’t accept p for no reason at all.\(^{170}\) But acceptance is still more subject to voluntary control than belief, because for professional purposes, as when a lawyer accepts the innocence of their client without believing in their innocence.\(^{171}\) When there is an evidential gap, it is possible to accept a

---

168 Cohen 1992 p.4
169 “[W]hen one is attending to issues raised, or items referred to, by the proposition that p, normally to feel it true that p and false that not-p, whether or not one is willing to act, speak, or reason accordingly”. Ibid.
170 Hypotheses differ from acceptances because they are “are ways of premissing that do not need to hang on anything beyond themselves”. Cohen 1992 p.13.
171 Cohen 1992 p.20
proposition for the sake of a friendship[^172]. Unlike belief, which is involuntary in the way our dispositions to have feelings are generally involuntary, acceptance is voluntary because it is a choice about what we treat as a premise, and we generally have a choice about what we accept as a premise; we can treat \( p \) as a premise in our reasoning and practical deliberations voluntarily[^173]. There is nothing about the three dispositions of reasoning from \( p \), acting on \( p \) and preparedness to assert \( p \) which would make this voluntariness impossible[^174].

The problem is that, even if Cohen’s acceptance is voluntary, it is still subject to norms of justification. What could make an attitude subject to a norm of justification? I’ll consider two options. First, it might be that an attitude is subject to norms of justification because being subject to the truth norm is constitutive of that attitude, and norms of justification are means to comply with the truth norm[^175]. Second, it might be that belief is subject to norms of justification because they play an important role in epistemic projects[^176]. I’ll look at these in turn.

Bernard Williams gave three criteria for being an attitude constitutively governed by the truth norm. These criteria are supposed to be true of belief. First, belief is rightly assessed by reference to its truth or falsehood: “truth and falsehood are a dimension of an assessment of beliefs as opposed to many other psychological states or dispositions”[^177]. Acceptances too can be rightly assessed in part on the basis of truth and falsehood; if you accept something you know to be false, you’re irrational. That’s true even in the cases of defending a client or trusting a friend; defending someone you know to be guilty is probably wrong and usually illegal, and trusting a friend who you know to be untrustworthy is generally unwise.

The second feature of aiming at truth for belief is that “to believe that \( p \) is to believe that \( p \) is true. To believe that so and so is to believe that that thing is true”[^178]. But to have a policy to use \( p \) as a premise amounts to a policy to use “\( p \) is true” as a premise – at least, the two are readily interchangeable if the Tarski truth schema for \( p \) is also accepted. And if that schema isn’t accepted, Williams’s point won’t apply to belief, either, so on the assumption Williams is right about belief,

[^172]: Ibid.
[^173]: “[A]t bottom [acceptance] executes a choice – the accepter’s choice of which propositions to take as his premises”. Cohen p.22.
[^174]: The contrast with belief lies in Cohen’s sympathy with Williams’ understanding of the involuntariness of belief; following Williams, Cohen says that we “think of our beliefs as states of mind that are normally responsive to the truth, not to our own decisions”. Ibid.
[^176]: See 2.4.3.
[^177]: Williams 1973 p.137
[^178]: Ibid.
his point will still apply to acceptance. It’s also not clear how any attitude could fail to satisfy this criterion; desiring p is also desiring that p is true.

Finally, Williams says that “to say ‘I believe that p’ itself carries, in general, a claim that p is true”. This might be thought to provide a genuine asymmetry between belief and acceptance. While it is strange to say ‘I accept p and p is false’, there seems nothing wrong with accepting p and declining to claim that p is true: ‘I accept that p and p could equally well be true or false’ doesn’t haven’t the same ring of absurdity as it would with belief.179

But this seems well-explained by the fact that assertions are implicitly governed by norms which relate to belief, rather than acceptance. It sounds strange to assert ‘I believe that p and p could equally well be true or false’ because, whatever the fundamental norm of assertion turns out to be, at a minimum you ought to assert p only if you believe that p. To be able to assert ‘p could equally well be true or false’ you have to believe that p could equally well be true or false. But then you can’t, without inconsistency, also believe that p.180

But once we’re allowing alternative doxastic attitudes into the picture it’s not obvious why we shouldn’t expand our norms of assertion or, alternatively, hold that we sometimes engage in shmassertions, where shmassertions are governed by acceptance-grounded norms. For example, perhaps you ought only to shmassert p if you have a true acceptance that p, or know that p, or have a rational acceptance that p. Then we would have just the same phenomenon arising for acceptance and shmassertion and we do for belief and assertion.

Suppose we went with the ‘reasonable acceptance’ norm of shmassertion. Cohen might object that there are cases it can be reasonable to accept that p without reason to think that p is true. The lawyer case, for example, shows how we can have practical reasons which make it reasonable to accept that one’s client is innocent without reasons to think it true that one’s client is innocent. But although that might be true, nothing in the conceptual division between belief and acceptance gives us any

179 For more on Moore’s paradox and its relationship with norms of assertion, see Baldwin 1990 p.226 onwards, Heal 1994, Shoemaker 1995. Shoemaker would argue that the problem isn’t with Moorean assertions but with states like believing p while suspending judgment on p. Shoemaker thinks that’s right because belief is transparent: you can find out whether you believe p by asking whether p obtains. The disposition to feel that p obtains is what distinguishes belief from acceptance, so this wouldn’t apply to acceptance. The problem is that, if the impropriety of Moorean states arises solely from the existence of this disposition, it doesn’t explain why the truth norm would be constitutive of belief; rather, the disposition would explain why we’re tempted to take the truth norm to be constitutive of belief. Dispositions to satisfy a norm aren’t enough for a state to be governed by that norm.

180 Nothing about this last step turns on belief’s having a special relationship with truth; this holds for the same reason you can’t coherently desire that p and not desire that p is true.
new reason to think that it is true. If you think that the nature of belief is that no attitude can be belief if it isn’t governed by the truth norm, then you won’t be impressed with the following thought: perhaps belief has that nature, but what about belief minus – a concept defined as the usual dispositions associated with belief, minus being governed by the truth norm? It’s precisely those usual dispositions – the dispositions to act, assert, and reason as though p is true – which incline us to think that belief aims at truth in the first place.

I now want to consider the second way of being governed by norms of justification, by playing a role in an epistemic project. In the last chapter I conceded to Rinard that there is just one sort of justification, and that practical reasons can bear on whether we’re justified in believing. But when we ask whether we’re justified in self-trust we’re interested in whether we’re justified given a fragment of those reasons – the reasons which bear on whether the higher-order commitment in self-trust is true.

From that point of view, attitudes like acceptance can never dodge the evidentialist obligation, because acceptances are still fit to be evaluated as beliefs would be, at least as long as epistemic reasons can bear on acceptances. There’s no reason to think that they don’t, or couldn’t, and so there’s no reason to think that acceptances are any worse suited for evaluation within an epistemic project. To form part of an epistemic project it’s enough that holding an attitude satisfy our curiosity. And if an ersatz belief attitude is sincere, it must satisfy our curiosity. That’s because satisfying our curiosity means we can cease assessing our evidence about p. When we’ve apportioned our attitudes to the norms of justification guiding an epistemic project, we’ve done the best we can do, synchronically, to satisfy our curiosity. Before we’ve so apportioned our attitudes, we can’t have settled on p. But if we sincerely reason on the basis that p, that means we reason on the basis that p in any context where new reasons haven’t been introduced. But to reason on the basis that p means that we must already have settled on p. That would make it impossible to satisfy our curiosity with respect to p by complying with norms of justification around p, because satisfying our curiosity requires not settling on p until we’ve fully assessed our evidence. So from the perspective of an epistemic project, a sincere commitment to p which involves reasoning on the basis that p is a curiosity-stopper – it’s an attitude which prevents further assessment of our evidence as long as it is maintained.

On the other hand, failing to reason on the basis that p seems essential for the sort of commitment we have towards TC, because that commitment seems to involve being able to appeal to TC to
justify our reliance on our faculties. Without that ability, No Denial would be impossible to explain; it was the fact that No Denial needs some explanation which motivated the introduction of A-commitment. Dropping the reasoning disposition of ersatz belief attitudes would leave our commitment to TC unable to play the role it does in explaining the reasonability of our everyday reliance on our faculties.

But there isn’t anything more an attitude needs to be subject to norms of justification than serving as a curiosity-stopper. On this approach, there isn’t any essential connection between belief and norms of (epistemic) justification – practical reasons can apply to belief. Norms of epistemic justification are just restrictions of those reasons to those which help us satisfy our curiosity. But that restriction serves just as well for acceptance as it does for belief.

Acceptance need not always satisfy our curiosity. But it will only fail to satisfy our curiosity when it fails to be sincere. Acceptance doesn’t satisfy the curiosity of the lawyer about their client’s innocence, and it doesn’t satisfy our curiosity about a friend’s trustworthiness. But that’s because in some contexts – outside of the courtroom, in our friend’s absence – we’re prepared to stop reasoning on the basis that our client is innocent or that our friend is trustworthy. That gives us the opportunity to assess our evidence for these propositions.

But unless our commitment to TC is going to be like playing Devil’s advocate, construing self-trust as insincere isn’t an option for us. Insincere commitment to the necessary presuppositions of a project entails insincere engagement in that project.\(^{181}\) If we’re insincere in our hinge commitment then we’ll have some belief-dispositions associated with belief in TC in some contexts but not in others. Consider the contexts where we lack some dispositions. Then we will either no longer rely on our doxastic practices or we will continue to do so. If we continue to do so we will be akratic because we will be committed to ‘p and I lack commitment to the truth-conduciveness of the process which formed my belief that p’. If we fail to do so, our first-order beliefs generated by our ordinary doxastic practices will also be insincere, because they’ll violate Cross Domain – we will have dispositions associated with holding those beliefs only in certain contexts. Something like this seems to be behind the intuitive idea that sincerity, engaging in a project seriously, matters; the value of that project is diminished if it is engaged in merely provisionally or ironically.

\(^{181}\) Jenkins 2007 makes the similar point that trust in the presuppositions of our epistemic projects is necessary if we are to engage in those projects ‘seriously’ (p.29).
This applies to sincere attitudes beyond acceptance. The Pyrrhonist’s attitude of appearance seems to satisfy Cross Domain with respect to action-guiding dispositions.\(^{182}\) Appearance is supposed to be some attitude short of belief which is still action-guiding but which lacks some of the other dispositions associated with belief. The Pyrrhonist doesn’t act incoherently, acting as though they believed that \(p\) in some contexts but not others, so appearance is sincere, even if the Pyrrhonist doesn’t feel that \(p\) is true or assert \(p\) without qualification.\(^{183}\) Such attitudes are examples of Eric Schwitzgebel’s ‘in-between beliefs’. Schwitzgebel has argued that in cases where \(p\) feels true to \(S\) when directly considered, but \(S\)’s actions seem to betray an implicit belief that not-\(p\), then we ought to view \(S\) as having an attitude of ‘in-between belief’.\(^{184}\) On Schwitzgebel’s view, an account of \(S\)’s relevant dispositions – their dispositions to explicitly affirm \(p\), to treat \(p\) as a premise, and so on – is an account of \(S\)’s belief, and if those dispositions don’t appear to cohere, then we’ll call that a case of in-between belief. Both of these serve as curiosity-stoppers as long as they are sincere with respect to reasoning on the basis that \(p\). Pyrrhonian appearance and in-between belief do seem to satisfy that requirement, and hence are also subject to norms of justification and hence must satisfy Evidentialism.

This demonstrates the failure of a general strategy for vindicating self-trust. We try to work up a new attitude which satisfies Cross Domain by sharing many of the same dispositions as belief, but in some way – perhaps we just stipulate that this is so – this attitude is supposed to be immune from the evidentialist obligation. But it is those dispositions – dispositions which fall out of satisfying Cross Domain – which leave beliefs themselves open to the evidentialist obligation, whether you’re sympathetic to Williams-type arguments for evidentialism or epistemic project-type arguments. So we can’t have a sincere non-doxastic attitude which captures the sort of commitment we have to TC while avoiding evidentialist obligations which are shared with belief.

\(^{182}\) See Perin 2010. I interpret the Pyrrhonist as taking a special attitude towards contents \(p\), appearance; my argument here doesn’t apply to the Pyrrhonist who believes that it seems that \(p\), but doesn’t believe that \(p\). Such a Pyrrhonist has the same attitude as us (belief) but different contents of their beliefs. This sort of Pyrrhonism isn’t much help to us, either, because believing that it seems to me that TC isn’t enough to avoid akrasia if we don’t also believe TC, unless we replace all of our first-order beliefs that \(p\) with beliefs that it seems to me that \(p\). Presumably these beliefs would also be generated by some doxastic practice, the truth-conduciveness of which we must then have some higher-order commitment to, and so the same problem would recur, assuming that my conclusion in Chapter Two that there are no privileged practices was correct.

\(^{183}\) In some contexts the Pyrrhonist may utter ‘\(p\)’ and mean ‘It appears to me that \(p\)’; because their utterance is guided by different norms to ordinary assertion it may be better to think of this as a variant of schmassertion than assertion proper.

\(^{184}\) See Schwitzgebel 2001; see also Sommers 2009, Borgoni 2015, 2016 for other views on in-between beliefs. Sommers would say that we have an in-between belief when we have beliefs of two different classes, which he calls ‘mondial’ and ‘propositional’; Borgoni thinks we should attribute to the subject both a belief that \(p\) and a belief that \(\sim p\).
3.3.3 Assumption in Hinge Epistemology

I’ve said that acceptance plays a role in an epistemic project as a curiosity-stopper. It might still be thought that that role is importantly different from that of belief; roles within epistemic projects might be more fine-grained than whether they stop further assessment of evidence. Coliva’s attitude of assumption might be one such attitude, because it seems to play an importantly different role in an epistemic project from belief while also foreclosing further assessment of our evidence with respect to its content.

Coliva says that our many everyday particular judgments are justified only because we have made certain general assumptions about ourselves and the world, including the assumption that our faculties are truth-conducive. These assumptions are not themselves justified; that is what distinguishes Coliva’s view from Wright’s. But they are reasonable to hold because the particular assumptions which we make about ourselves and the world aren’t merely accidentally related to epistemic justification, in a way which would allow many other such assumptions to fulfil the same role, but instead are constitutive of epistemic justification as such.

On Coliva’s view, my objection to acceptance is like objecting to a referee’s uniform on the grounds that in football the clothes we wear are subject to rules (players must wear the colours of either of the two teams) and the referee can’t dodge this obligation through wearing a special sort of clothes – if he’s wearing clothes, he’s liable to incur the team-kit obligation. This is a bad argument because team-kit obligations apply only to those who are playing the game, and the referee isn’t playing the game, although he is necessary for the game to be played at all. Similarly, playing the game of belief just is seeing what is evidentially supported, once certain assumptions are granted; there is no sense in asking what justifies us in holding those assumptions, because epistemic justification presupposes having just such assumptions. Assuming TC does have a role in an epistemic project, but only insofar as serving as a curiosity-stopper about TC is necessary to engage in an epistemic project at all, and this isn’t the same role as belief plays.

But note how the non-doxastic attitude no longer seems to be playing a role in this story. What matters is the notion of rules which apply to some parts of the game and not others – rules which

185 Coliva 2015 p.119
186 Coliva 2015 p.33
187 Coliva 2015 p.120
apply to our commitments to ordinary propositions but not hinge propositions. Why, then, do we need assumption, rather than belief in hinge propositions?

What Coliva might want to say is that non-doxastic attitudes are necessary because belief couldn’t play that constitutive role. That’s because if belief could play that constitutive role then belief would be playing a dual role: belief would still be open to evidential evaluation, and even beliefs in hinge propositions could still be judged by evidentialist standards. This evidential role of belief would lead to incoherence when paired with the constitutive role of (certain) beliefs. An epistemic project would involve making assumptions which, by that project’s own lights, are not only unjustified but wholly unreasonable. That’s because – for all we’ve said so far – while we’re engaging in an epistemic project we’re narrowing the available reasons which would render a belief permissible to just those reasons which epistemically justify a belief. So not only would hinge beliefs be judged unjustified within the terms of an epistemic project, they’d be judged impermissible within those same terms, too, and engaging in an epistemic project would involve internal incoherence. This is why we’d want to substitute belief in TC with assumption that TC.

But if I’m right in what I’ve been arguing, then assumption, acceptance, or any other sincere attitude is also subject to evidentialist standards, and will generate the same incoherence between their constitutive and evidential roles. Non-doxastic attitudes are essential for Coliva’s account, but they can’t play the role they need to play in that account without rendering an epistemic project incoherent. That’s because evidentialist constraint on norms of justification obtains in virtue of the role it plays in satisfying our curiosity. A commitment to p has to abide by evidentialist requirements if it’s to satisfy our curiosity about whether TC obtains. TC is something that we’re curious about, and we want to satisfy that curiosity, so commitment to TC can’t be exempted from the evidentialist constraint. So sincere commitment, of all stripes risks having the same dual constitutive and evidential role that belief in TC would have.

### 3.4 Conclusion

I’ve examined some alternative attitudes to belief, concluding that each has individual defects which make them unsuitable for capturing the sort of commitment we have to the truth-conduciveness of our own faculties without falling under the same evidentialist strictures placed on belief. This doesn’t show that HOCTC is belief. It might be that, if the nature of belief means that no affirmatory attitude based on insufficient evidence is belief, then HOCTC must be some non-
doxastic attitude, since, as I argued in Chapter Two, we lack sufficient evidence for TC. Nonetheless, I have shown that any such attitude must be sincere, and sincerity incurs similar evidentiary obligations to belief. While alternative attitudes can dodge the problems that might arise from the nature of belief, those aren’t the only problems we have to deal with; if I’m right in my conclusion from Chapter Two, then there are further obligations that result from our normative theory which don’t rely on the nature of belief.
4.1 Constitutive Norms, Guidance Norms and Excuses

4.1.1 Introduction

I want to show that our commitment to TC is reasonable because it’s excused. In this chapter I’ll give an account of excuses. I’ll first motivate the notion of excuses with a number of examples, showing that both internalists and externalists need to appeal to excuses (4.1.2), before resisting the temptation to reduce excused attitudes to blameless attitudes (4.1.3). I’ll next consider and argue against accounts by Williamson (4.1.4) and Littlejohn (4.1.5) which I’ll call the derivative norm and virtue theoretic accounts respectively. I’ll then present and argue for the guidance norm account (4.2). To present that account I’ll discuss the variety of accounts already presented for guidance (4.2.2.1-4.2.2.4) and suggest my own (4.2.2.5). I’ll then show that the guidance norm account can accommodate the examples which motivate the introduction of excuses (4.4) before responding to objections (4.5).

4.1.2 Motivating Excuses

Before I look at particular accounts of excuses I’ll first fix the intuitive idea of excuse which those accounts aimed at. Consider the following three cases:

**Accidental Crime**
Curtis, recently arrived in a new city, drops a cigarette butt on the floor, unaware that in his current location this is a crime. Curtis is used to the laws of his old home, where doing so is legal, and has not thought to check all of the laws again when he moved. Curtis has violated the law, but claims in his defence that he can’t be expected to check every law upon moving, and made the reasonable assumption that such laws are constant between cities.

**Misleading Drink**
Solomon looks at his glass, sees that it contains a clear liquid, and on that basis believes that there is gin in his glass. In fact the glass contains petrol. Solomon violates prudential norms by sipping from
his glass – it’s not prudent to drink petrol – but claims that he can’t be expected to check everything he drinks for petrol, and made the reasonable assumption that undrinkable liquids aren’t generally found in glasses.  

**Impossible Promise**

Rufus promises to help Carla move on a certain day, but has broken his leg and can’t fulfil his promise. Rufus has broken his promise, but pleads that he couldn’t help doing so.

In each of these cases the subject violates some norm. Nonetheless their conduct seems better off than someone who knowingly and voluntarily violated the norm. The mitigating circumstances in each case vary; in Accidental Crime Curtis lacks the time or cognitive resources to sift through every law in his new city, in Misleading Drink Solomon has been deceived by a misleading appearance, and in Impossible Promise Rufus is unable to comply with the norm through circumstances beyond his control. But intuitively the subjects’ mitigating circumstances put them in a better normative position than someone who didn’t have such mitigating circumstances. We can’t say that they are in as good a normative position as someone who didn’t violate any norm – Curtis really did break the law, Rufus his promise, and so on. So we need some middle status to describe the sort of normative position each occupies. Let’s say that a subject S is excused when they occupy such a middle status. The task of an account of excuses is to explain the nature of this middle status.

Excuses also arise with respect to epistemic norms. In particular, externalists are often sympathetic to excuses because they explain away some intuitions which would otherwise be problematic for externalism.  

**Right-Handed Society**

Wilson is raised in a society in which left-handed people are widely said to be gullible. This belief has been formed on no evidence at all, but questioning this belief is subject to serious social sanction, and Wilson comes to share this belief through testimony from childhood and maintains it because he is never given reason to question it.

**New Evil Demon**

---

188 This example is taken from Williams 1981.
189 Littlejohn forthcoming b, Williamson forthcoming
Mavis is regularly subject to misleading visual appearances, and comes to share the same beliefs as her epistemic counterpart, Martha, who holds many justified beliefs, some of which amount to knowledge, on the basis of veridical appearances indiscriminable from Mavis’s own.

**Misleading Library**

Lee believes himself to be in the philosophy section of the library because he can see *Socratic Dialogues* on the bookshelf; he forms the belief that there is a philosophy book on the shelf, and reason that he is in the philosophy section. In fact *Socratic Dialogues* is a collection of interviews with renowned Brazilian midfielder Socrates; as it happens, Lee really is in the philosophy section, and *Socratic Dialogues* has been mislaid on the shelf by a passing football fan.

In each case the subject’s belief seems to have something going for it. From their own point of view, they’ve behaved reasonably, and it’s only their unfortunate circumstances that have led them to have a false belief. One natural thought is that in each of these cases the subject’s belief is justified. But many forms of externalism won’t be able to take that route, because behaving reasonably from one’s own point of view isn’t sufficient for justification. On some versions of externalism, Wilson’s misleading testimony, Mavis’s misleading visual appearances and Lee’s false premise not only prevent their beliefs from amounting to knowledge but deprive their beliefs of justification.

Externalist accounts which have these consequences will have to find some non-justificatory state which captures the sort of status each has in these cases. Again, this non-justificatory state appears to be a middle state between being justified and being knowingly and voluntarily in defiance of norms of justification. So for the externalist, it’s natural to say that Wilson, Mavis and Lee are excused for their beliefs. This might come along with an error theory about our intuitions in these cases; we might think that intuitions about excuses are difficult or impossible to distinguish from intuitions about justification, and so we shouldn’t confuse our intuition that there is something to be said for Wilson, Mavis and Lee, from a normative point of view, with an intuition that their beliefs are justified.

Although excuses are useful for externalists, internalists also need to appeal to excuses. Internalists think that all reasons to believe are mental states. But not every apparent reason will be a genuine reason – at least, there’s no obvious reason why the internalist must be committed to this. Even on access internalism, it’s not obvious that our accessible mental states may not be deceptive, appearing to be reasons to believe without being genuine reasons. Consider:
**Distorting Pill**

Otis is trying to work out whether his evidence supports $p$; Otis knows that $p$ or $q$ and knows that $\sim q$, and so is in a position to reason from his evidence to $p$. Unbeknownst to Otis, he’s been given a drug which makes it strongly seem to him that his evidence supports $\sim p$, and he comes to believe $\sim p$.  

In Distorting Pill, Otis is unable to abide by norms of justification, even internalistically construed; Otis will reach an unjustified judgment about $p$ because he’s failed to assess his evidence properly. Still, Otis has intuitively not behaved as badly as someone who came to the same conclusion without having ingested the pill.

This need not rely on misleading appearances of evidential support relations. Consider:

**Intuitionist Society**

Sam knows that $p$ or $q$ and knows that $\sim q$; apart from this, Sam also has good, defeasible evidence that $\sim p$, although less strong than his evidence for $q$. Sam has been raised in a society of intuitionists who cast doubt on disjunction elimination. The proof of $p$ from $p$ or $q$ and $\sim q$ does not seem valid to Sam and, since Sam’s other evidence supports $\sim p$, Sam comes to believe $\sim p$.

Again, Sam is not as normatively badly off as if he’d been raised in a society like ours. If either of these cases seem plausible then even the internalist needs some account of Otis or Sam’s normative status. So excuses seem to have an epistemic application beyond their helpfulness for the externalist.

Perhaps some internalists would say that Otis and Sam are justified because they have abided not only by their evidence but also believed what their evidence seems to support; some internalists might want to say that what counts for justification is what the evidence seems to support, not what it really does support. If that’s right, then some forms of internalism won’t require excuses. But that some forms do require them tells us that excuses aren’t part of an *ad hoc* move made by the externalist but something necessitated by a wide array of views on epistemic normativity. Even the variant of internalism discussed in this paragraph may still require excuses if we aren’t infallible about our seemings, in particular about what our evidence seems to support; consider the variant

---

190 See Christensen 2010.
case where Otis swallows a pill which subtly distorts how his seemings about evidential support seem to him. Unless this is impossible, this will be a case where this variant on internalism will also need to appeal to excuses.

Finally, excuses can arise from computational inability. Consider:

**Logic Test**
As part of a logic test Ann must decide whether a list of sentences are tautologies or not. Ann inspects a complicated tautology $\phi$. Before trying to prove or find a counter-model to $\phi$, Ann has a credence of 0.5 that $\phi$ is a tautology and suspends judgment that $\phi$ is a tautology.

**Complicated Consequence**
Betty believes that the sky is blue, and is asked to consider $\phi$, a complicated sentence which is entailed by ‘the sky is blue’. Because $\phi$ is so complicated Betty can’t see that it is a consequence of ‘the sky is blue’ and suspends judgment on $\phi$.

Tautologies and complicated consequences are supported by Ann and Betty’s evidence, and so each ought to believe tautologies and complicated consequences when asked to consider them. But they lack the computational resources to prove all of the very many tautologies and consequences they might be asked to consider. If a subject ought to assign attitudes based on what’s supported by their evidence, Ann and Betty are doing what they ought not to do. But it’s impossible to meet this standard, so each is excused for failing to do so.

*Prima facie* the normative consequences of excuses come in at least two forms. Excuses have *axiological weight*, in the sense that excused attitudes have more value – or at least less disvalue – than unexcused attitudes. And they have *deontic weight*, in the sense that excused attitudes are in some way better off with respect to what we ought to do than unexcused attitudes. I don’t want to decide how these two normative dimensions of excuses ought to be understood – that’s for an account of excuses to decide. But a good account should have something to say about them.

There might be any number of ways we can be excused. Someone who is temporarily delusional because they’ve involuntarily ingested some drug is excused for their actions. Someone overcome with emotion such that they couldn’t act rationally might have an excuse, too, at least for some actions. On the other hand, someone in a Gettier case isn’t like someone who is overcome with
emotion, or suffering from hallucinations; they’re someone who, so far as their own conduct is concerned, has behaved just as any other good epistemic subject would, but (on some externalist views) find themselves with unjustified beliefs because they have formed those beliefs in an unfortunate environment. A good account of excuses should accommodate as many of these disparate excuses as possible; perhaps on further investigation they will turn out to be fundamentally different phenomena, but a good account of excuses ought at least to illuminate the grounds of their differences.

4.1.3 Excuses and Blamelessness

Part of our intuitions here involve blameworthiness; the subjects in all of these cases appear to be normatively better-off in part because they’re blameless. But at this stage blameworthiness is as unclear as excuses themselves are. It’s for particular accounts of excuses to decide how to construe blameworthiness. There should be some intuitive sense that someone blameless is not in the wrong, has not behaved as badly as someone who is blameless, could not be fairly criticised, or in some other way is in a better position than someone who violates a norm through avoidable error. But just what blamelessness involves is too closely bound up with the nature and value of excuses to decide before examining particular accounts.

If we had an independent grip on blameworthiness perhaps we could use that to build an account of excuses. And perhaps we do have an independent grip on blameworthiness. Cameron Boult suggests that epistemic blame is made possible by inhabiting certain sorts of relationships with others: we all have “intentions and expectations towards one another simply in virtue of the fact that we are epistemic agents in an epistemic community”, and we “rely on one another in a generic way as sources and distributors of information”. We’re epistemically blameworthy when we violate one of these expectations.

This is unsatisfactory for our purposes for two reasons. First, if excuses were to be accounted for in terms of blamelessness, it wouldn’t give a clear account of the normative status of excuses. Nothing in the idea of violating the expectations of others lends itself to explaining this normative facet of excuses; unless there’s some clear link between expectations and epistemic normativity this account

---

191 For more on blame, see Rettler 2018, Daoust 2019, Piovarchy 2020, Boult forthcoming.
192 Boult forthcoming
193 Boult forthcoming p.9
doesn’t seem to do justice to our intuitions. As it happens Boult also endorses Williamson’s
derivative norms account, which I’ll discuss in the next section; my point here is only that this
account of blameworthiness on its own isn’t enough to account for the normativity of excuses.

It might be objected that this makes the unwarranted assumption that any normative notion beyond
reasonable criticisability is necessary for a good account of excuses. Our intuitive data, it might be
said, show only that there is something positive to be said for someone who is excused, but says
nothing about what that status should be. This brings me to my second argument against Boult:
Boult’s account is extensionally inadequate, precisely because our intuitions do say more than that
an excused subject has some positive status of indeterminate nature. Suppose Willie, raised in a
society like ours, is transplanted into the society considered in Right-Handed Society, in which
everyone raised to think of left-handed people as gullible. Then if Willie were to share their beliefs
about left-handed people, he wouldn’t be violating anyone’s expectations; nonetheless, Willie seems
intuitively blameworthy, because Willie should have known better than to share their beliefs.
There’s an intuitive difference between Willie and Wilson, and that difference seems to be a
normative one; Willie is doing something worse than Wilson is, even though neither violate the
expectations of their peers.

Boult could object that Willie violates our reasonable expectations of him. But that isn’t what’s
supposed to ground reasonable criticisability – that’s supposed to be the obligations we have as
members of an epistemic community, and Willie isn’t part of our epistemic community. Boult could
also object that Willie violates the reasonable expectations of the right-handed society by failing to
respect more general epistemic requirements, because Willie isn’t incorporating all his evidence –
Willie knows something his peers don’t, because he was raised in a different society, but he isn’t
incorporating that knowledge into his other beliefs. But by hypothesis, every member of Right-
Handed Society is violating such general requirements, because their belief in the gullibility of left-
handed people has been formed without evidence. If reasonable criticisability can be this broad it
doesn’t capture the intuitive notion of blamelessness.

4.1.4 The Derivative Norms Account

Williamson has suggested that excuses arise from following certain norms derived from ordinary
norms of justification without following those ordinary norms. Call the ordinary norms of
justification primary norms; then on Williamson’s account, for any set of primary norms, we also
have secondary norms: being the sort of subject who is disposed to comply with primary norms. For example, suppose the primary norm of justification is

**Knowledge Norm of Belief (KNB)** S ought to believe that p only if S’s belief would amount to knowledge.¹⁹⁴

Then someone complying with the secondary norm corresponding to KNB would be disposed to base their beliefs on sufficient evidence, taking care not to believe lottery propositions, ensure their beliefs were safe and so on.

We also have tertiary norms: acting as a subject disposed to comply with primary norms would on a given occasion. Someone complying with the tertiary norm corresponding to KNB would, on a particular occasion, believe that p only if someone disposed to comply with the secondary norm would believe p on that occasion, and so would take care that their beliefs were based on their evidence, safe and so on. Call a norm *derivative* if it is either a secondary or a tertiary norm.

On Williamson’s account, a subject is excused just when they comply with a derivative norm without complying with a primary norm. Intuitively there’s something better about complying with a derivative norm than not, so it’s plausible that someone who is excused, on this account, has a better normative status than someone not excused. Obeying a derivative norm seems to demonstrate a good faith attempt to comply with a primary norm which has been thwarted for reasons which even someone generally disposed to comply with that primary norm could not have avoided. This seems to vindicate some of our intuitions in the cases discussed above. Mavis and Martha would both trust their faculties even if they were a brain in a vat, so Mavis is complying with a tertiary norm in believing herself to have hands even though she fails to comply with the (externalistically construed) primary norm. If there is a link between obeying a derivative norm and having a preferable normative status, then Williamson can explain our intuitions about Mavis and Martha.

But I don’t think Williamson can substantiate the *(prima facie* plausible) view that compliance with derivative norms lends a subject a preferable normative status. To start, Joshua Schechter has argued that excuses so construed aren’t enough to capture the normative status of unjustified assertion because some unjustified assertions are ones we *ought* to make, not just assertions which we’re

---

blameless in making. Mavis ought to assert that she has hands (if that’s conversationally salient) even though she doesn’t know that she has hands. Derivative norm excuses can’t explain that, because what Mavis ought to do is still determined by the primary norm, and the primary norm would require Mavis not to assert that she has hands.

A good account of excuses should either meet Schechter’s challenge or explain why doing so isn’t necessary. Does Williamson’s account meet Schechter’s challenge? Williamson would say that we ought to obey derivative norms, but this ought is in some way derivative from the ought imposed by primary norms. But this isn’t yet to answer Schechter’s challenge: we don’t know in what way the derivative ought is derived from the primary, nor do we know which ought takes primacy where there are conflicts between the two. That derivative norms “need not inherit the full normative status” of their primary norms suggests that the ought of primary norms takes primacy, but we don’t know enough about the derivation to explain why or tell under which circumstances derivative norms do inherit the full normative status of primary norms.

This relates to a general problem with Williamson’s account, the normative link problem. Williamson says that anyone sympathetic with primary norms “is likely to feel some consequential sympathy for the derivative norm,” but it’s not obvious what this is supposed to amount to. It’s clear what the link between compliance with the primary norm and our normative status is – to comply with a norm of justification just is to have a justified belief; unless there’s some other status with a claim to the epistemic ought, it seems that in complying with a norm of justification we’re doing what we epistemically ought. But what is the relationship between derivative norms and epistemic normativity? What normative status do we have in virtue of obeying a derivative norm? Here are three responses Williamson could make:

(1) Williamson might want to say that derivative norms do have a clear link with a normative notion, namely, blamelessness. But, as I argued in the previous section, there isn’t a clear notion of blamelessness at play here, prior to our having a precise account of excuses. The only obvious option – reasonable criticisability – is unsatisfactory, and there is no further way to characterise blamelessness in a way that wouldn’t also be a direct account of excuses. So in defending his account of excuses, Williamson can’t appeal to a prior notion of blamelessness.
(2) Williamson might resist the normative link objection by pointing out that derivative norms are the sorts of things which could easily be confused with primary norms. It’s no objection to this account that excused subjects are intuitively normatively better off than unexcused, because that we would have such intuitions is a prediction made by the account.

But we ought to resist this move, because our intuitions here are more detailed than the account would predict. The account can’t explain why we think following derivative norms is any better than following other norms which are in some way related to primary norms. Here are some other norms which are derived from primary norms which could plausibly be confused with following a primary norm:

**Appearance Norm** Believe p if others would believe you were following a primary norm in doing so.

**Normalcy Norm** Believe p if doing so would constitute following a primary norm in normal worlds.\(^{199}\)

But our intuitions are that excuses confer a better status than compliance with just any norm derivative from a primary norm. Unless the derivative norms account can explain why tertiary norms confer a better status than appearance norms or normalcy norms we haven’t explained all of our intuitions here.

(3) Williamson could point out that derivative norms are more closely tied to successfully satisfying primary norms, in normal cases, than these alternative norms are. This is one way in which Williamson thinks the normativity of primary norms rubs off on derivative norms; if you ought to comply with a primary norm, then you generally ought to comply with derivative secondary and tertiary norms, too.\(^{200}\) I think that’s right, but note how, if the correlation between primary and derivative norms is going to be significant, we’ll have to appeal to the instrumental role of derivative norms in satisfying primary norms. If we didn’t, satisfying a derivative norm would be epiphenomenal with respect to satisfying a primary norm. Couldn’t there be other norms which

---

200 Williamson forthcoming p.8. Williamson clarifies that two sorts of oughts may not be quite the same: “the underpinning of the latter ‘ought’ is derivative from, and more complicated than, the underpinning of the former ‘ought’.”
have the same instrumental role? I’ll argue shortly that there are, and that some of these are just the
derivative norms which are central to a good account of excuses. But that would be to deviate from
the derivative norms account. It would also require us to have a good account of how instrumental
role can transmit normative status. I’ll provide such an account in the next chapter, but without that
the derivative norms account is incomplete; moreover, the account I’ll present won’t cohere well
with secondary and tertiary norms, and so isn’t available to Williamson.

4.1.5 The Virtue Theoretic Account

On Littlejohn’s view, subjects are excused when they have exercised their epistemic faculties
excellently, but have still failed to form justified beliefs. This might occur when a subject is
mistaken about a fact or norm, although they have exercised their faculties properly in trying to
determine whether that fact or norm obtains; if this mistake leads to an unjustified belief, then
Littlejohn says that the subject is rational but not justified.

How does this improve on the derivative norms account? Is there a difference between being
rational and satisfying secondary norms, or having a rational belief and satisfying a tertiary norm?
One important difference is in the available answers to the normative link problem. A subject who
has exercised their epistemic faculties excellently is epistemically virtuous, even if they don’t have
justified beliefs. The normative link between being epistemically virtuous and other notions, like
justification and the epistemic ought, might be obscure, but that isn’t a special problem for the
virtue theoretic account. This does better than the derivative norms account because Littlejohn isn’t
committed to the normative status of rational belief being derivative from that of justified belief,
and can instead appeal to whatever the link is between epistemic virtue and other normative notions.

On Littlejohn’s schema, not every blameless unjustified belief is excused. Some beliefs are exempt
in the sense that the subject can’t exercise their epistemic faculties at all. Because Littlejohn
distinguishes between excuses and exemptions, he can deny that someone whose reasoning has
been addled by a pill has an excuse; such a subject is exempt, because they aren’t able to exercise
their faculties at all. Conversely, someone like the subject in Intuitionist Society is excused, because
they have exercised their epistemic faculties excellently, although given the society in which they
were raised their excellent exercise has still resulted in a belief which is unjustified.

201 Littlejohn forthcoming b
Does the success of this response to the normative link problem rest on a change in focus from the positive status of the attitude to the positive status of the subject? When we find epistemic virtue praiseworthy it is primarily a subject which is praiseworthy, and their attitudes derivatively if at all. The problem with this is that it fails to answer Schechter’s problem. It seems that Mavis ought to assert that she has hands, in just the same way as Martha ought to assert that she has hands. The virtue theoretic account can explain why Mavis is behaving excellently in asserting that she has hands, but it can’t explain our intuition that Mavis’s act of asserting is a reasonable one, one which she ought to do; it can explain the former but not the latter because it can explain why an unjustified subject is excellent but not the normative status of an act of that subject.

The following wouldn’t be a good explanation: Mavis’s assertion is excellent in a derivative sense because it’s the product of Mavis’s excellently exercised faculties. Actions can be the product of excellent exercises of faculties while being unreasonable. If Wilson discriminates against left-handed people because he was raised in the right-handed society, his act of discrimination might be the product of excellent exercises of his faculties, but we feel no temptation to say that he was doing as he ought. So action-excellence is both not equivalent to reasonability, nor is it easily confused with reasonability. The virtue theoretic account can neither answer Schechter’s challenge, nor can it explain why it doesn’t need to be met through some sort of error theory.

4.2 To be Excused is to Satisfy a Guidance Norm

4.2.1 What is Guidance?

Both the derivative norm and virtue theoretic accounts seemed to be getting at the thought that there is a positive normative status associated with doing the best you can, given your resources, to satisfy norms of justification, even if your attempt is bound to fail. But neither could explain the normative status of the attitude which is held as a result of doing the best you can, given your resources. But there is a normative notion in this area which seems to accommodate Williamson and Littlejohn’s key insight here: the distinction between a set of norms which constitute good performance relative to some target or which establish standards for evaluating good conduct, and a set of norms which are suitable for guiding the subject to successful compliance with constitutive norms or which prescribe conduct which would be evaluated as good. To get an intuitive feel for this distinction, compare the following pairs of norms:
Truth/Justification Consider two norms of belief:

**Truth Norm of Belief (TNB)** S ought to believe p only if p is true

**Justification Norm of Belief (JNB)** S ought to believe p only if S’s belief that p would be justified.

TNB and JNB both plausibly capture obligations on all epistemic subjects. They can also come apart; it’s possible to have false justified beliefs and true unjustified beliefs, so TNB and JNB will sometimes require incompatible attitudes towards a proposition. But neither seems eliminable in favour of the other. TNB captures one sense in which correct beliefs are true beliefs; if TNB were false, it would seem that we could be indifferent to having a false belief as long as it were justified. But TNB doesn’t seem, on its own, enough to guide us in forming true beliefs or give us any prescriptions which would help us have true beliefs. JNB does, because aligning our attitudes with our evidence while respecting enkratic principles is something which is in our power even when having true beliefs isn’t. So TNB appears to be a constitutive or evaluative norm, while JNB is a guidance or prescriptive norm.

Externalist/Internalist Evidentialism Consider two evidentialist requirements:

**Externalist Evidentialist Requirement (EER)** S ought to believe p only if S has sufficient evidence for p, where S’s evidence includes factive and non-factive mental states, as well as non-mental states.

**Internalist Evidentialist Requirement (IER)** S ought to believe p only if S has sufficient evidence for p, where S’s evidence includes only S’s non-factive mental states.

EER seems to be guiding in a way that TNB isn’t, for the same reason as JNB can guide when TNB doesn’t. But it seems less guiding than IER is. Consider two cases:

**Genuine Lemon**

Shuggie sees a real lemon, and forms a belief that there is a lemon before him.

202 That the two have this relationship is suggested in Toppinen 2015.
**Fake Lemon**

Smokey hallucinates that there is a lemon before him, and forms the belief that there is a lemon before him.

Both Shuggie and Smokey will find it easy to comply with IER; both can form a belief on the basis of their non-veridical visual sensation as of seeing a lemon. But Smokey will find it hard to comply with EER; when he attempts to do so, he’ll be misled by his hallucination of a lemon and fail to form a belief based on his total evidence, including factive mental states. Because it can sometimes happen that our non-factive mental states are readily discriminable while our factive mental states aren’t, it can sometimes be easy to comply with IER when it isn’t easy to comply with EER. Sometimes this is taken as an argument against EER. But we can also take the EER/IER distinction to parallel the TNB/JNB distinction. EER might give standards which constitute good epistemic conduct, while IER gives guidance towards good epistemic conduct. Note that while TNB seemed not to be guiding at all, EER is guiding in some cases, although not all. So the distinction between constitutive and guidance norms isn’t absolute, but relative between pairs of norms; EER is guiding relative to TNB, but not relative to IER.

**Good Directions/Bad Directions** Ray is given two sets of directions to Kilburn; one involves landmarks with which Ray is familiar, and directs Ray between them using right and left turns and distances in metres. The other uses locations specified in coordinates, uses rotations in radial angles and distances in millimetres. Satisfying either will get Ray to Kilburn, but the first set of directions are much easier for Ray to follow. So satisfying either constitutes taking the best path to Kilburn, but only the former directions are guiding; we could use failure to satisfy the bad directions as a standard with which to evaluate Ray poorly, but Ray could only follow the good directions.

**Objective Tax Code/Subjective Tax Code** Stevie has to satisfy a very complicated tax code T when preparing a form. T says to enter deduction X only if C obtains. Determining whether C obtains is very hard. But the subjective tax code, ST, says to enter deduction X only if Stevie believes C to obtain. It’s much easier for Stevie to determine whether he believes C to obtain than to determine whether C obtains, and hence ST is much easier for Stevie to satisfy than T. Still, getting the form right is constituted by satisfying T, not satisfying ST.
Let a norm which is helpful in the intuitive sense picked out by these examples by a guidance norm, and a norm which determines correct conduct, whether or not it is guiding, by a constitutive norm. There are three things which we need to know about guidance norms:

(4) What is the right account of guidance?

(5) How do guidance norms relate to constitutive norms? Are all norms both guiding and constitutive? Are there are any which have one feature but not the other? Are there any which have none? Does being guiding entail being constitutive, or vice versa?

(6) What is the normative status of complying with a guidance norm?

I’ll answer (4) in this chapter and (5) and (6) in the next.

Guidance norms look like they are suitable for explaining the normative status of excuses because of two joint considerations:

(7) Abiding by a guidance norm appears to have some link with other normative notions, one which we have reason to hope would be borne out once a satisfactory account of guidance norms has been presented. It’s natural to say that you ought to abide by a guidance norm if its corresponding constitutive norms have normative force. And there is some reason to think we have a grip on the way in which guidance norms link up with other normative notions – the ‘ought’ of guidance seems to have some instrumental value in securing the ‘ought’ of constitutive norms. That can’t be the whole story – we’d still need to have an answer to the swamping problem for guidance norms, for example. But it’s a reason to think guidance norms might be able to do the job we want excuses to do.

(8) It’s implausible that following guidance norms will always be compatible with following the constitutive norms to which they correspond. TNB has different requirements to JNB, EER to IER, and T to ST. This parallels the link between excuses and justifications. Being excused can’t entail being justified, even if being excused means you have gone some way towards being justified.

If these two considerations are right – if there’s a good account of guidance norms out there which can both vindicate the intuition that guidance norms are fallible while conveying some positive normative status on an attitude formed through compliance with those norms – then we have the account of excuses we’re after. To be excused in holding an attitude A to p is to have that attitude through compliance with guidance norms without satisfying the corresponding constitutive norm. It might turn out that there is no account which can satisfy (7) and (8) simultaneously; perhaps our intuitions are confusing two subtly distinct notions of guidance. But there ought to be a presumption that an account of excuses which can answer the normative link problem is available here.

4.2.2 Accounts of Guidance

There are lots of accounts of guidance available in the literature, and the entailment relations between each form of guidance aren’t obvious. One natural thought is that norms are guiding when they take into account a subject’s first-person point of view:

**First-person Guidance** A norm N is *first-personally guiding* for S just in case, if N requires φ-ing when C obtains, whenever C obtains it makes sense for S to φ from S’s own point of view. JNB is guiding because it recommends things which make sense from a subject’s own point of view, while TNB isn’t because it sometimes recommends believing things which, from S’s perspective, seem extremely unlikely. But the idea of making sense from a subject’s point of view needs sharpening before First-person Guidance is to be clear. I’ll now look at four accounts of guidance and argue that they aren’t suitable for capturing the notion we’re after. It’s important to look at each of the accounts which are current in the literature because, if there are unexamined alternatives, it might be doubted whether the account we arrive at has the important normative links which we need to give an account of excuses. It’s plausible that some form of guidance has a solution to the normative link problem; in the next four sub-sections I’ll argue that none of the available accounts of guidance puts us in a position to provide such a solution.

4.2.2.1 Lucid Guidance

First-person Guidance could be sharpened in a number of different ways. It might be that the conditions C under which N requires φ-ing are those which S knows, or at least is in a position to know, obtain:
**Lucid Guidance** N is *lucidly guiding* for S just in case S is always in a position to know whether φ-ing would satisfy N.\(^{204}\)

Lucid guidance captures the way that good directions are guiding when bad directions aren’t. The good directions tell Ray to take a right turn when he gets to St John’s Wood. The bad directions tell Ray to rotate \(\pi/2\) radians clockwise when at coordinates \((x,y,z)\). Ray is in a position to know when he’s at St John’s Wood and knows that taking a right turn would be following the directions. But Ray isn’t in a position to know that he’s at \((x,y,z)\) and doesn’t know that rotating \(\pi/2\) radians would be following the directions. So the good directions are lucid and the bad directions aren’t.

But it doesn’t capture guidance in the other cases, because the other cases involve conditions which we aren’t always in a position to know obtain. If Williamson’s anti-luminosity argument is sound then we aren’t always in a position to know whether we’re justified in believing p, or what our evidential mental states support, or what we believe the tax code demands, because those are non-trivial conditions which can be arbitrarily difficult to discriminate from cases which don’t obtain.\(^{205}\)

Suppose N requires φ-ing when in C, but not in ~C. Then for S to be in a position to know whether φ-ing would satisfy N, S must be in a position to know whether C obtains. Assume that knowledge requires a modal condition like sensitivity which ensures that S knows that C only if in very close situations where ~C S doesn’t believe that C. Now consider the case where C is sufficiently borderline that S would be unable to detect whether ~C. If N is lucid, S is in a position to know C. But then S is in a position not to believe ~C if ~C obtains in very close situations. But this seems to attribute to S implausible – even indefinite – capacities to detect whether or not C obtains. This is a contradiction with otherwise plausible premises, so we ought to reject the assumption that N is lucid.

While the anti-luminosity argument has been criticised\(^{206}\), it seems preferable to avoid requiring lucidity, all else being equal. If the anti-luminosity argument is successful in general, it’s successful with respect to JNB, IER and ST. The anti-luminosity argument seems to apply to all non-trivial conditions. As long as any condition C admits of cases where ~C but which are indiscriminable from C, then the anti-luminosity argument will go through. That seems to be the case with having sufficient justification, having internalist evidential support, and believing that p. Justification and

\(^{204}\) See Srinivasan 2015.
\(^{205}\) See Williamson 2000, 2008.
\(^{206}\) See Vanrie 2020

124
internalist evidential support can be very finely balanced; our evidence can be arbitrarily close to any given threshold of sufficient evidential support. Since we don’t have arbitrarily fine-grained powers of discrimination with respect to our exactly which credence our evidence supports, there will be cases where our evidence supports φ-ing but there are close indiscriminable counterfactuals where our evidence doesn’t support φ-ing. Similarly, we often hold a belief weakly enough that there are close indiscriminable counterfactuals where we don’t hold that belief, so we aren’t always in a position to know that we have a belief.

If the anti-luminosity argument is right, no non-trivial norms are lucid with respect to all triggering conditions C and action-types φ. But some norms are more lucid than others. Some norms are lucid with respect to more pairs of action types and triggering conditions. This motivates a notion of relative guidance:

**Local Relative Lucid Guidance** N is guiding relative to N* with respect to subject S, action type φ and triggering conditions C just in case S isn’t in a position to know whether φ-ing would satisfy N* in C and S is in a position to know whether φ-ing would satisfy N in C.

If a norm is lucid in more cases than another, then it’s better at guiding:

**Global Relative Guidance** N is guiding relative to N* (for S) just in case N is locally guiding relative to N* (for S) with respect to more pairs of action types and triggering conditions.

The JNB, IER and ST would be guiding relative to the TNB, EER and T respectively if they are lucid with respect to more triggering conditions and action types.

Global relative guidance ensures a total ordering on relative guidance, but it sometimes gives counter-intuitive results. Consider:

**Chess Guidance**

Irma is considering two sets of directions for winning at chess, Alpha and Deep. Alpha gives directions which are lucidly guiding in games which start with the most popular openers, but isn’t lucidly guiding with any others. (Maybe Alpha always presupposes some background knowledge

---

207 The threat of such indiscriminability might be thought to motivate the denial that evidence supports a unique, very fine-grained credence; perhaps instead it supports an interval of credences. But the same problem will recur at the endpoints of such intervals, since those endpoints can also be arbitrarily fine-grained.
and Irma only understands that background for games with popular openers.) Deep gives directions which are lucidly guiding for many games not starting with the most popular openers, but not for those which do. (Perhaps Deep has been constructed using situations where players only use obscure openers.) Let’s stipulate that Deep is lucidly guiding for more games than Alpha overall. Then Deep is guiding relative to Alpha. But that seems wrong; Irma will prefer to follow Alpha’s directions because they’re more helpful for actual games, since actual games tend to start with popular openers.

So N can be globally guiding relative to N* without N being more helpful than N*. N* might be locally guiding relative to N in just those contexts where we care about getting things right. Or N and N* might each sometimes be the more locally guiding norm relative to one another in different contexts all of which we consider important. Lucid guidance doesn’t tell us when a set of norms ought to be preferred, even if the only grounds for preferring a set of norms is how guiding they are. And as Srinivasan points out, guidance may be just one value among many; if N isn’t guiding relative to N* but does a better job at securing some important good (true belief, say) then N might be preferable on that basis. So lucid guidance doesn’t seem helpful for understanding excuses, because there seems no reason to excuse someone for following a relatively lucid norm at the expense of following a less lucid norm; in many such cases our intuitions will be that they ought to have complied with the less lucid norm, and they could easily have done so.

4.2.2.2 Subjective Guidance

There are other ways of sharpening First-person Guidance. Sometimes guidance is understood in terms of the contrast between objective reasons and subjective reasons. When N requires S to φ when C, C is an objective reason to φ. C may involve other reasons; it may be that C is a complex of (possibly competing) pro tanto and/or prima facie and undermining reasons for and against φ-ing. Then if R is a member of that complex which speaks in favour of φ-ing, R is also an objective reason to φ, and if R is a member of that complex speaking against φ-ing, R is an objective reason not to φ.

It’s hard, perhaps impossible, to comply with some norms whose triggering conditions are objective reasons because subjects who don’t possess those reasons are unable to comply with those norms. Because they’re unable to comply with those norms, it’s often much harder to satisfy those norms. If you don’t possess some of the reasons comprising C, then it’s impossible to φ because C. So there

126
are reasons out there which you might not possess – objective reasons – which are no help in complying with norms, and there are reasons which you do possess, and which therefore can help you comply with norms. Call reasons which you do possess your subjective reasons.

Subjective reasons to φ might be objective reasons to φ plus some restriction, or they might be something different from the objective reasons to φ.

If the former, the most plausible restriction is some epistemic relation towards an objective reason. If I lack a reason to φ, it’s because I’m not aware of C, or not aware of some R which forms part of C. That epistemic relation might be belief, or justified belief, or knowledge, or something else.

I’ll follow Schroeder’s account of the latter understanding of subjective reasons. For Schroeder, subjective reasons are (potentially non-factive) reasons which, if true, would be objective reasons. So if I falsely believe C obtains, and N requires φ-ing when C, then my reason for φ-ing is C. The thought is that if subjective reasons aren’t a subset of the objective reasons to φ, they must be non-factive reasons which bear some suitable relationship with the objective reasons to φ; I’ll take Schroeder’s account of this as a good representative of the general view.

Whatever our account of subjective reasons, we can now see the motivation for:

**Subjective Guidance** N is subjectively guiding for S just in case, for every triggering condition C in N, C is a subjective reason for S.

Subjective guidance seems important because complying with a norm seems important, and we can seemingly only do that by responding to our subjective reasons. But there are three problems for subjective guidance.

First, we seem perfectly able to respond to reasons which aren’t subjective. I can take a drink from my glass because it contains water, not just because it seems to contain water, or because I believe it contains water. What I’m responding to is that there’s water in my glass; that’s the natural way to explain what’s going on, anyway. This isn’t a problem for a view like Schroeder’s, because Schroeder admits non-factive reasons.

---

Schroeder 2008

127
Second, subjective reasons don’t in general seem sufficient for compliance. We can be mistaken about our beliefs, justified beliefs and knowledge. So even relativising what we ought to do to the recommendations of our subjective reasons doesn’t guarantee compliance with the resulting norm. At most it might make it easier to comply with the norm, or raise the probability of our so complying, but in this respect subjective reasons don’t have any special relationship with compliance – there are lots of ways to adjust norms such that the result is easier to comply with, and if subjective reasons are a particularly effective way to do this it is at most a difference of degree rather than kind. A similar problem applies to Schroederian non-factive reasons, because Schroederian reasons are also unsuitable for explaining the differences between the easy and hard norms in the four cases considered above. The difficulty with complying with the bad directions over the good directions isn’t that the good directions’ reasons are non-factive, or even that they’re indifferent to being non-factive. There’s more to guidance than being guided by subjective reasons in Schroeder’s sense.

Third, it remains unclear what relevance subjective reasons have for what we ought to do. However we understand subjective reasons this is a problem. Either subjective reasons are objective reasons plus a restriction, or they’re something else. If the former, then if our subjective reasons recommend φ-ing when the objective reasons require not φ-ing, our subjective reasons recommend failing to comply with a norm. If the latter, and they’re something like what Schroeder has suggested, then subjective reasons have at best some hypothetical relationship with what we ought to do – if our subjective reasons were true, then we ought to do what they recommend. But this has no bearing on what we actually ought to do.

We might want to say that subjective reasons are essential for compliance, and anything essential for complying with a norm inherits the normative status of that norm. But we can’t appeal to that here, if subjective reasons are neither necessary nor sufficient for compliance with a norm. The first two problems I’ve raised suggest that subjective reasons at best make compliance with a norm easier, but that doesn’t mean we must follow our subjective reasons.

Someone sympathetic to subjective reasons might maintain that ‘ought’ is ambiguous. There’s an objective sense of ‘ought’ fixed by the totality of objective reasons, and a subjective sense fixed by the recommendations of our subjective reasons. If ‘ought’ is baldly ambiguous, this risks changing the question – we’re interested in whether we ought to believe our faculties to be truth-conducive, and if it turns out that, although objectively we ought not, there is an unrelated subjective sense in
which we ought to, it’s not clear what bearing that unrelated sense has on our original question. It’s more plausible if ‘ought’ is polysemous, possessing multiple distinct but related meanings. But this is plausible only if there is some important connection between the objective and subjective senses of ‘ought’, and that connection would have to be spelled out for this to be a satisfactory response. But spelling out that connection requires finding some substantive connection between objective and subjective oughts, and without supplying that the ambiguity theory is only offering us a promissory note.

4.2.2.3 Deliberative Guidance

Subjective Guidance doesn’t seem plausible as an account of excuses. But it might be getting at something more plausible; subjective guidance is motivated by consideration of the sort of reasons which we can respond to, and perhaps that consideration can give us a better account. Such a consideration is appealed to when philosophers place a requirement on reasons that they be guiding in deliberation:

Deliberative Guidance A reason R for φ-ing is deliberatively guiding just in case S can incorporate R in their deliberation about whether to φ. 209

Deliberative Guidance is a plausible account of guidance because it’s motivated by consideration of what it takes to comply with, rather than merely satisfy, a norm. It’s plausible that complying with a norm requires satisfying a norm because you’re sensitive to what that norm requires of you, and it’s plausible that that sensitivity is captured by the idea of deliberating from a triggering condition to an action-type.

Deliberative guidance avoids some of the problems with subjective guidance because it permits non-epistemic and non-mental facts to be guiding. We can deliberate about whether to drink from the glass from the premise that it contains water, not just that we believe that it contains water. That avoids the first problem raised for subjective guidance.

But deliberative guidance doesn’t seem to be sufficient for guidance in the sense targeted here, because it doesn’t explain the differences found in some of the four cases found above. Perhaps it’s impossible to be deliberatively guided by the truth norm. But it’s not impossible to be deliberatively

guided by the objective tax code – good accountants do so all the time. Similarly, we can be guided by our externalist evidence. We’re not infallible in doing so – when we’re in inhospitable environments our deliberations will go awry. But deliberative guidance would be implausible if it demanded that we be able to incorporate our reasons into infallibly successful deliberation. So deliberative guidance is too weak to get at the concept picked out by the four cases considered above.

Perhaps we should read Deliberative Guidance as referring to what a particular subject is able to do. So perhaps a trained accountant can incorporate all of the triggering conditions found in the objective tax code into their deliberations, but I can’t do so, and hence the objective tax code isn’t deliberatively guiding for me. The problem is that for some bits of the objective tax code I do seem to be able to incorporate its triggering conditions into my deliberations, but can’t do so well; even when the objective tax code tells me to φ under conditions C comprising reasons R₁...Rₙ, and I start off with premises R₁...Rₙ, I get muddled and wind up not φ-ing. Whether I’m able to incorporate R₁...Rₙ into my deliberations seems less important than whether I’m able to do so successfully for distinguishing the objective tax code from the subjective tax code. Similarly, I can incorporate some premises including coordinates and radians into my deliberations, but I’m worse than when I include landmarks and bodily rotations in following directions.

We could try to make Deliberative Guidance more demanding by requiring that we be able to incorporate reasons into successful deliberation, or reliably successful deliberation, or something like that, but doing so seems to depart from the original motivation for deliberative guidance. That original motivation is that incorporation into our deliberations is necessary for φ-ing because it satisfies a norm. Being deliberatively guiding is supposed to give a subject a fair chance at complying with a norm – norms which weren’t deliberatively guiding would be impossible to comply with if compliance requires being able to deliberate about the triggering conditions of norms. Deliberative guidance gets close to what I’ll argue is the right account, but falls down at explaining why guidance ought to extend to inabilities rather than solely enabling us to incorporate reasons into our deliberations.

4.2.2.4 Advice Guidance and Directive Guidance

Finally, First-person Guidance might be getting at the difference between norms applying to yourself and to others. Those applying to you need to direct you. Those applying to others need only
evaluate them. That makes guidance sound like advice. We can think of advice like this: an utterance is good advice if it’s an utterance which will make it sufficiently likely for S to comply with N, and make S likely in the ‘right way’\textsuperscript{210}. Analogously, the axioms of ZFC fix the valid transitions of axiomatic set theory, but the sort of utterances which might help someone to comply with those axioms might not be a simple enumeration of those axioms.

But prescriptive norms in the sense of the best advice you can give to someone seems much too particular to count as norms. Take N as our evaluative norm. Then it won’t generally be the case, for agents A, B, C, D, that the most helpful advice for A to give B for complying with N will be the same as the most helpful advice for A to give D; nor is the most helpful advice for A to give B necessarily the same as for C to give B. So helpful advice will be very variable not only between subjects but between pairs of subjects. Our difficulties with complying with evaluative norms might take all sorts of forms – does everyone have the same difficulties with the axioms of ZFC? Often the advice might be quite different – suggestions for helpful sets to use as part of the axiom of separation, parsing the axiom of replacement into mathematical English, replacing the axiom of choice with an equivalent but more familiar theorem such as the well-ordering principle. If anything has deontic consequences here – and surely there are things we ought and ought not to do in set theory – it can’t be prescriptive norms in this sense.

But the thought might be that good advice is approximating some maximally simple set of instructions; there are some very basic transitions all subjects can make, and where S can’t follow N, advice breaks N’s requirements down into more basic transitions which they can make. Directive guidance would be the limiting case of good advice:

**Directive Guidance** N is directive guiding just in case, for every triggering condition C with required action-type φ in N, C is a basic input and φ is a basic action.

Although good advice can vary, directive guidance is plausibly unique. However we understand basic inputs and basic actions, they’re supposed to be the fundamental building blocks of complex inputs and actions; if there were more than one sort of directive guidance, there’s be more than one way of breaking down complex inputs and actions. However that turns out, directive guidance doesn’t seem to explain the four cases considered above, because complying with directive guidance

\textsuperscript{210} The ‘right way’ condition is necessary avoid deviant causal chains; presumably it will involve requirements that an utterance is successful in virtue of its informational content about N and N’s consequences rather than in virtue of constituting a threat, promise or some other non-communicative speech act.
guidance seems very hard for persons, if not for sub-personal processes. Directive guiding norms would be more like bad directions than good directions and more like the objective than the subjective tax code. So directive guidance isn’t what we’re looking for.

4.2.2.5 Compensation Guidance

4.2.2.5.1 Statement of the Account

For any norm N, there are some abilities which do and some which don’t have some bearing on S’s success in complying with N, or in acquiring whatever good N aims at, if any. When S lacks those abilities, S has N-relevant failings. When N is the bad directions to Kilburn, a subject’s N-relevant failings include their ignorance of global coordinates, because lacking such knowledge prevents them from complying with N. When N is ST, a subject’s N-relevant failings include their inability to process their tax-related knowledge in a way which will ensure they pay the legally-required amount of tax. That’s because, although our failings don’t prevent us from complying with N, N aims at some further good which isn’t constituted by compliance with N: paying the legally-required amount of tax. So our failings here are still N-relevant in the sense stipulated.

There’s a task for philosophical psychology of explaining how background abilities and attitudes enable a subject to comply with a norm, and hence how N-relevant failings prevent a subject from complying with N. But explaining how we’re able to comply with a norm is a different task to determining whether a norm guides. To be guiding, I claim, is just to have sufficient compensation for any N-relevant failings which could prevent a subject from complying with a norm:

**Compensation Guidance** N is guiding for S (with respect to φ-ing when C obtains) just in case S is able to comply with N (by φ-ing when C obtains), in spite of S’s N-relevant failings.

The thought is that guidance norms, in the sense picked out by the four cases discussed above, are norms of compensation. For any subject S and any two sets of norms N, N*, N is a norm of compensation for N* just in case S has N*-relevant failings which prevent them from complying with N*, and complying with N is the best way of achieving the value aimed at by N* which is possible given S’s N*-relevant failings.\(^{211}\) For example, suppose that the ideal route around a track would have Daisy take a corner at 70 mph, but that in fact attempting to do so would cause Daisy to

---

\(^{211}\) Norms of compensation are discussed further in DiPaolo 2019a, 2019b.
panic and slow down, or else steer off course; then a norm of compensation would recommend that Daisy take the corner at a slower speed because doing so would form part of the ideal route once we compensate for Daisy’s limitations.

The idea behind compensation guidance is that there might be many ways in which we can fail to be able to comply with a norm. Just as there may be many ways of failing to know that \( p \), there may be many ways of failing to comply with \( N \) without there being anything which unifies the varieties of failed compliance. Compensation guidance can acknowledge the motivations which lie behind other forms of guidance. Sometimes the flaw which prevents a subject from complying with a norm is lacking knowledge about what a norm requires. Sometimes it might be that we don’t possess the reasons which would get us to \( \varphi \) when we ought to \( \varphi \); this, too, is an inability that can be overcome by compensation norms. But there need be nothing which unifies this collection of inabilities. There are lots of ways to fail to comply with a norm, and they don’t seem to form a natural kind.

It might be thought that compensation guidance is too binary; guidance norms seem to be easier for \( S \) to comply with than constitutive norms, but in both cases \( S \) can comply with the norm. \( S \) can comply with \( TNB \) – \( S \) can believe \( p \) because \( p \) is true. It’s just that it’s easier for \( S \) to comply with \( JNB \).

This isn’t a problem on one way of understanding inabilities. On that understanding, \( S \) is unable to comply with a norm when there is some block on their capacity to comply with that norm. By default, any subject can comply with a norm. If \( S \) can’t, it’s because they’re defective in some way. Now, when we say that it’s easier to comply with \( N \) than \( N^* \), that’s because we can have more defects and comply with \( N \) than we can have and comply with \( N^* \). Stevie can comply with \( ST \) while having lots of defects – poor attention span, innumeracy and so on – while Stevie couldn’t comply with \( T \) while having those same defects. So easiness of compliance seems to reduce to the capacity to accommodate defects.

Another picture of easiness is about risk of failure. When something’s easy, I put a high chance on my being able to do it. When it’s hard, I put a low chance. I’d expect to be able to comply with a hard norm some of the time, but not very often. This doesn’t fit nicely into the binary understanding of easiness where it’s a matter of whether I’m able to comply with a norm at all; it’s more that my success rate goes down continuously the harder a norm is to comply with.
But presumably there is some difference between the cases where we fail and where we succeed, even if we can’t make it out. Failure and success aren’t uncaused; there’ll have been some antecedents which have led to failure to comply or successful compliance. Then what’s going on here is you’re able to comply with norms in fine-grained circumstances but not in others, so you can comply with a norm with respect to φ-ing in conditions C, but you can’t comply with a norm with respect to φ-ing in conditions C*. Being guiding is intended to be a fuzzy property constituted by being guiding with respect to enough pairs of action-types and triggering conditions, where ‘enough’ is to be sharpened according to contextually-specified standards.

4.2.2.5.2 Relationship with Other Accounts

Compensation guidance has various entailment relations with the other notions of guidance already considered. It doesn’t entail advice guidance because compensation norms need have no pedagogical role; they give the best possible way of compensating for an inability, but they need not be effective in helping someone compensate for an inability when communicated in an utterance. Similarly, compensation norms need have nothing to do with the sort of instructions which are found in directive guidance; norms of compensation need not spell out behaviour in basic inputs and action-types to provide a best possible course of conduct.

We can often comply with norms without being able to know what a norm requires, as the anti-luminosity argument shows, so there is no need for norms of compensation to be lucid. The converse entailment also doesn’t hold; we can fail to comply with a norm through inabilities beside not knowing what a norm requires, as when we lack the ability to φ even when we know a norm requires us to φ.

If it’s not possible to comply with a norm without subjective reasons, then compensation guidance is always subjective guidance. If so, compensation guidance is stronger than subjective guidance, because there are more ways to lack the ability to comply with a norm than lacking reasons. I’ll argue in Chapter Five that compensation guidance is able to explain the link between guidance and normativity in ways unavailable to subjective guidance alone.

Compensation guidance would entail deliberative guidance if being able to take a reason into our deliberation were required for complying with a norm. There are reasons to doubt that that’s the
case, so whether compensation guidance entails deliberative guidance should be controversial.\textsuperscript{212} But the converse relation doesn’t hold, because there are more ways to fail to comply with a norm than not being able to deliberate about its triggering conditions. We can deliberate about the triggering conditions of the bad directions or the objective tax code without being able to comply with those norms. Compensation guidance is therefore able to explain relative guidance in cases which deliberative guidance couldn’t; compensation guidance is stronger than deliberative guidance and so it’s easier to fail to be compensation guiding than it is to be deliberatively guiding.

If guidance norms just are norms of compensation, then we might think that guidance norms are just whatever is recommended under bounded rationality.\textsuperscript{213} Bounded rationality looks like it’s concerned with compensating for our limitations and giving ideal strategies conditional on our being non-ideal subjects. But I don’t think bounded rationality captures the sorts of norms of compensation which are in play when compensating for failure to abide by constitutive norms. Bounded rationality is primarily motivated by the thought that we have limited computational resources, and choosing to aim at higher standards of accuracy imposes costs – it imposes opportunity costs, because aiming at higher accuracy about p in general requires trading off higher accuracy about q, and it imposes practical costs, because if we intend to have higher accuracy about p this will in general come at the expense of achieving competing goals.

This can’t be the notion we’re interested in. Bounded rationality is concerned with diachronic norms of inquiry – should we put resources into investigating p, given the costs of doing so? But when we’re recommending epistemic conduct, we’re concerned with synchronic norms which tell you what you ought to believe given the considerations already available to you. We’re also interested in doing so in such a way that you can be as good an epistemic agent as possible. That means ignoring trade-offs, both practical and epistemic. But bounded rationality is inherently concerned with both sorts of trade-offs. So bounded rationality isn’t what we’re concerned with here. For example, introducing a satisficing rather than an optimising requirement for norms of justification would be one way of incorporating bounded rationality. Perhaps this would work like this: rather than violating a norm by failing to believe based just on sufficient evidence, an agent

\textsuperscript{212} Some such premise is assumed in Hughes 2018, forthcoming. But compare compliance with a norm with inferring according to a rule. Boghossian 2001, 2003, 2014 argues that inferring requires more than believing in accordance with a rule, but believing in accordance with a rule because the rule licenses the belief. There are views which accommodate this without requiring that we take what a rule licenses as a reason to believe (see, eg., Valaris 2016, Broome 2014, Hlobil 2014, 2019). If any of those views are plausible, then we ought to be sceptical of requiring deliberative guidance for compliance.

\textsuperscript{213} For links between bounded rationality and reasoning, see Lewis, Howes and Singh 2014, Karlan 2020.
violates satisficing norms of justification only if they hold beliefs which deviate ‘too much’ from their evidence, where ‘too much’ is sharpened by giving a metric of evidence-deviation and a threshold for how much deviation will count as too much.

However this turns out, it’s clear that this doesn’t capture the constitutive/compensation contrast – there seems nothing in principle impossible about having optimally-evidentialist beliefs once it’s granted that we can have satisficingly-evidentialist beliefs. Once we’re sensitive to evidence at all, it is at most a matter of costs (perhaps quite extreme) to ensure that we have beliefs which are optimally sensitive to the evidence. The motivation for satisficing norms would seem to come then from a concern with costs and trade-offs, not with the sort of epistemic agent who can’t make trade-offs.

4.2.2.5.3 Application to Motivating Cases

Norms of compensation capture the sort of guidance at issue in the four cases considered above:

Truth/Justification Subjects like us are defective in the sense that we can have misleading evidence – we can sometimes be misled about what’s true. But we also need evidence as a means to form true beliefs, and so can be misled into having false beliefs. This is a very general defect, and it’s not worth blaming ourselves for it. But it’s not a totally universal defect. An omniscient being like God doesn’t have this defect, because that p is itself a reason for God to believe p. Nonetheless, we can comply with JNB even with this defect. It’s plausible that complying with JNB is the best available way to satisfy TNB, given this defect. If that’s right, then JNB is a norm of compensation with respect to the truth norm.

Externalist/Internalist Evidentialism We can be misled about the nature of our factive evidence in cases where we aren’t misled about our non-factive evidence. We can be misled because our powers of discrimination aren’t infallible; Smokey can’t tell the difference between a veridical appearance of a lemon and a hallucination of a lemon. This infallibility isn’t a necessary feature of hallucinations; we can sometimes tell when we’re hallucinating, and with sufficient background knowledge or finer powers of discrimination, Smokey would be able to tell that he was only hallucinating a lemon. So it’s our flawed powers of discrimination which mean we’re sometimes unable to comply with EER. But those flaws don’t prevent us from complying with IER. And IER is
generally the best way of satisfying EER if we’re unable to comply with EER. So IER is a norm of compensation for EER.

*Good Directions/Bad Directions* Ray is defective because he can’t match his sensory inputs with global coordinates, and can’t match rotations in radians with bodily movements; the bad directions can’t be complied with, given Ray’s defects. Note that, for very different subjects, the bad directions might be guiding and the good directions not. A mathematician unfamiliar with London who can deduce their current coordinates by sextant and compass will find the bad directions easier to follow than the good, which presuppose knowledge of local landmarks. And the verdict of the compensation guidance account seems to track our intuitions about the particular case – there aren’t many mathematicians with good navigational knowledge around so our initial intuition is that the bad directions really are bad, but when the possibility of such a mathematician is raised it seems plausible that the bad directions would be more guiding for them than the good directions. The mathematician-lack the normal failings with respect to the bad directions, and so the good directions aren’t norms of compensation for the bad directions with respect to the mathematician.

Note also that Good Directions/Bad Directions is a case where the norm of compensation is just as good as the norm which is compensated for: Ray is able to get to Kilburn just as quickly following the good directions as the bad. It’s not always the case that compensating for a failing requires accepting a worse outcome. Compensation norms give the best possible outcome given a subject’s failings, but sometimes the best possible outcome conditional on having certain failings is identical with the best possible outcome *simpliciter*.

*Objective Tax Code/Subjective Tax Code* Stevie lacks the ability to process his tax-related knowledge correctly. The best possible way for Stevie to pay the taxes he owes, given his inability, is to fill out his form as best he can given what he believes about his tax affairs, assuming that his beliefs are formed to the best of his abilities. Doing so will get Stevie closer to paying the right amount of tax than guessing or copying his friend’s form.214

It might be thought that the best possible way for Stevie to pay his taxes is to hire an accountant. If that’s right then ST isn’t a norm of compensation for T, because it isn’t the best way of achieving

---

214 At least, it would be expected to be. It might be that in Stevie’s case, just by chance his friend owes exactly what Stevie owes, and copying his friend’s form would get Stevie closer to paying what he owes. This is a problem which I’ll address in Chapter Five.
the value T aims at. In general, it seems that norms can be relatively guiding without one norm being the best possible way of compensating for a subject’s failings. Sometimes, as may be the case here, it’s only necessary that a norm of compensation be good enough. Perhaps what is best is hiring an accountant, but it’s satisficing to comply with the subjective norm.

So it might be too strong to require that norms of compensation be the best possible way of compensating for failings. In many cases our intuitions will track norms which are good enough, but not ideal. In this norms of compensation are similar to many other norms. Our intuitions about justification can come apart from ideal justification, as happens with logical omniscience. This doesn’t mean that justification proper isn’t ideal justification, but that certain possibilities are excluded in certain contexts. Similarly, the requirement that norms of compensation be ideal is often loosened in practical contexts. That doesn’t show that norms of compensation simpliciter aren’t the best possible norms of compensation available. It might be that there is no best norm of compensation – perhaps for some N, N’s norms of compensation all approximate N better and better but there is no best norm of compensation. In such cases we’ll only be able to fall back on the satisficing understanding of norms of compensation. But where there is a best norm of compensation, we ought to comply with that norm.

Nonetheless, I don’t think that even the best norm of compensation would recommend that Stevie hire an accountant. Norms of compensation seem to come in two forms. Most of the time a norm of compensation requires that behaviour which would be best (for a contextually-specified understanding of ‘best’) conditional on possessing inabilities. Sometimes a norm of compensation will require that a subject do their best to lose their inability. This latter can happen when there are no norms which a subject can comply with which would achieve the aim of the constitutive norm. In such cases there is no way to compensate for our inability, and norms of compensation can only guide us by requiring us to lose our inability. But that isn’t the case with Stevie. There are norms available which wouldn’t require him to lose his inability, namely, ST. Hiring an accountant would be a way of Stevie losing his inability – he’d be making use of his accountant’s ability to reach his goal of paying the right amount of tax. So in this case norms of compensation won’t recommend that Stevie hire an accountant.

Note that it’s important that the subjects can’t comply with the constitutive norm in each of these cases, even if they can satisfy them. Perhaps it’s always possible to satisfy a constitutive norm, if only by chance. But in general the relevant defect is such that they’re unable to satisfy the norm in
any way that isn’t lucky, and so these subjects can’t satisfy a norm because it’s the norm. Nothing stops Smokey from believing on externalist evidence, Ray from walking to Kilburn or Stevie from filling in the right numbers on a form, but they can’t comply with EER, the bad directions or T respectively.

4.3 Application to Norms of Justification

Where a constitutive norm can’t be complied with by non-ideal agents, it sometimes admits of norms of compensation which can be complied with; these are guidance norms. This doesn’t give us a recipe for coming up with norms of compensation. Nor does it mean that there is a unique norm of compensation.

But the points considered so far suggest a natural way of finding norms of compensation. In general norms of compensation rely on pre-existing constitutive norms, and piggyback off of those norms to generate non-ideal norms. For example, suppose being justified in travelling to a destination means following a distance-minimising route; distance-minimisation is independent of starting location, because we can derive a distance-minimising route from any two points. That means travelling-justification admits of norms of compensation which will guide a flawed traveller along the ideal route, conditional on their making mistakes.

Let $R(x,y)$ be the distance-minimising route from $x$ to $y$, and $R(x,y) \& R(y,z)$ be the route $R(x,y)$ followed by the route $R(y,z)$. Then suppose $S$ starts at $a$, wanting to arrive at $b$ and takes a wrong turn, ending up at $c$. Then $S$’s route now ought to be $R(c,b)$. $S$ will also have taken some route to $c$; call that route $R^*(a,c)$. Then although $S$ has taken $R^*(a,c) \& R(c,b)$, they would have been distance minimising conditional on taking a wrong turn to $c$ had they taken $R(a,c) \& R(c,b)$. As long as $R(x,y)$ is well-defined for any $x,y$ we can generate non-ideal distance-minimising routes which compensate for mistaken diversions.

Alternatively, let $|R(x,y)|$ be the distance of a route $R(x,y)$. Then consider an ordering relation $<_D$ such that $R(x,y) _<_D R(v,w)$ just in case $|R(x,y)| < |R(v,w)|$. Consider the set $R[x,y]$ of routes from $x$ to $y$. Then $R(x,y) _<_D R^*(x,y)$ for each $R^*(x,y)$ in $R[x,y]$ not identical to $R(x,y)$. Then let $R_t[x,y]$ be the set of routes in $R[x,y]$ which $S$ can’t follow. $R[x,y] \setminus R_t[x,y]$ need have no smallest member with respect to $<_D$. Nonetheless, $R[x,y] \setminus R_t[x,y]$ still has an ordering, and we can make sense of a route in $R[x,y] \setminus R_t[x,y]$ being above or below a certain distance threshold; hence we can make sense of a
route in $\mathbf{R}[x,y] \cup \mathbf{R}_I[x,y]$ being satisficing, in the sense of being below a certain distance threshold. This gives us a recipe for norms of compensation: $\mathbf{R}^*(x,y)$ is a norm of compensation if it is in $\mathbf{R}[x,y] \cup \mathbf{R}_I[x,y]$ and below a certain threshold. That threshold might be fixed further by context – if it's crucial that we get from $x$ to $y$ under a certain distance, then that distance will fix the threshold. Nonetheless, this doesn't guarantee uniqueness for norms of compensation, because there might be more than one route beneath the threshold.

What's crucial for this toy example to work is that we have some good which we want our norms to achieve – distance-minimisation, in this case – and some measure of how far norms deviate from one another in achieving that good, such that we can compare pairs of norms. As long as our measure is linear, we’ll be able to determine satisficing norms for a given threshold.

Norms of justification also ought to admit recipes for generating norms of compensation, because we can also compare different ways of responding to evidence and satisfying structural constraints on justification. In particular we can compare them with respect to how well they compare to the ideal case, where our attitudes are entirely supported by the evidence and we satisfy all structural constraints. A norm of compensation would be one which generates responses to evidence which compare sufficiently well to other responses to meet a certain threshold.

We can only get a notion of doing sufficiently well on the lines of the toy example considered above if we have a linear ordering between the responses generated by a norm of compensation. Whether we have such a linear ordering may depend on the account of justification we endorse. One variety of account which would generate such a linear ordering would be one on which responding correctly to evidence is having a set of credences which abide by probabilistic axioms, perhaps in addition to other constraints on credences such as satisfying the Principal Principle. I don’t want to endorse this account, but assume it for the purposes of generating an account of epistemic excuses. If this is what justification is like, then we have a clear account of excuses in the epistemic case. That this account of justification does lend itself to generating norms of compensation speaks in its favour, but isn’t decisive, and the work that’s been done on non-ideal credence assignments seems prima facie generalisable to other accounts of justification.

Here’s how we can generate norms of compensation for norms of justification, probabilistically construed. I’ll follow Julia Staffel’s work to do this.\textsuperscript{215} First, as in the toy example above, we need to

\textsuperscript{215} Staffel 2019
define a distance norm between credence assignments, such that it’s determinate when one credence assignment is closer to probabilistic coherence than another; we define an incoherence measure for a set of credences by using one of a number of distance measures from the credences which would be assigned by norms of justification.\textsuperscript{216} Let $J(X)$ be the ideal set of attitude assignments, taking sets of evidence as arguments. Then let $J[X]$ be the set of sets of attitude assignments taking $X$ as an argument, and $J_I[X]$ be the set of sets of attitude assignments in $J[X]$ which $S$ can’t possess. Then as before, there need be no minimal element of $J[X] \setminus J_I[X]$. But there may be some satisficing elements with respect to the distance measure fixed for attitude assignments. To comply with a norm of compensation would then be to comply with a norm which generated credence assignments in $J[X] \setminus J_I[X]$ which were satisficing with respect to the distance measure fixed for attitude assignments.

Staffel’s work gives distance measures for credences. It’s not obvious how it should be generalised to attitude assignments, because it’s not obvious how to measure distance between sets of beliefs.\textsuperscript{217} There are three natural ways to do so. First, we could identify beliefs with credences over a certain threshold. Then we could apply Staffel’s method straightforwardly to sets of attitude assignments, because attitude assignments would just be credence assignments. Alternatively, we could remain neutral about whether beliefs are to be identified with credences, and say that the distances between attitude assignments are nonetheless determined solely by their credence assignments; since beliefs are supposed to be determined by credences in a context, any distinct attitude assignments must contain distinct credence assignments, so this will always generate a sensible distance between attitude assignments. Finally, we could treat outright belief as a credence 1 and outright disbelief as credence 0 in a proposition, and apply Staffel’s method again. This wouldn’t involve commitment to identifying beliefs with credences, but rather treating belief and disbelief as 1 and 0 in an attitude assignment; such attitude assignments may behave like credence assignments for the purposes of determining distance, but need involve no commitment to belief and disbelief being credences. Whichever method we pick will depend on our views about the relationship between belief and credence, but as long as there is some such relationship there will be an available method for determining norms of compensation.

4.4 Application to Cases

\textsuperscript{216} Staffel 2019 pp.76-85 discusses a number of available distance measures. The most salient are the Brier score and Euclidean distance measures. The Brier score gives the sum of the squared difference between actual credence and an indicator function outputting 1 or 0 if the proposition is true or false respectively, while the Euclidean distance measure gives the square root of the sum of the squared difference between actual and ideal credences.\textsuperscript{217} Staffel 2019 pp.171-98 discusses this further.
When N has normative force for S, S is *excused* when they fail to satisfy N but successfully comply with a norm of compensation which is satisficing with respect to the good which N aims at. Here’s how this can accommodate some of the examples of excuses with which we started.

**Right-Handed Society**

Wilson is, like the rest of us, dependent on the testimony of his society to form true, justified beliefs. Wilson can’t comply with norms of justification at all without being receptive to testimony, or he’d be ignoring evidence, and he lacks the ability to find defeating evidence for every bit of misleading testimony which he comes across. So the crucial inability is an inability to acquire and process the immense quantities of evidence which would be required in order to refrain from relying on testimony. So any norm of compensation for Wilson will assign credences in line with his having very many undefeated bits of testimonial evidence. Since Wilson does believe in accordance with what such a norm of compensation recommends, Wilson is excused for believing left-handed people are gullible.

**Misleading Library**

Lee can’t comply with norms which require him not to believe p when his belief that p would be true only by luck. The best that Lee can do is form beliefs in accordance with JNB; most of the time, the attitude assignments which are required by JNB will be equivalent to the attitude assignments which are required by KNB. So, if we think that subjects in Gettier cases have impermissible beliefs because they lack knowledge, we can still say that they’re excused because they’ve complied with JNB. Lee is excused for believing himself to be in the philosophy section.

**Distorting Pill**

Otis is unable to respond to his evidence properly. He is also unable to adjust his attitudes in light of the fact that he can’t respond to his evidence properly, because he’s unaware that he’s taken the distorting pill. But Otis can be responsive in the light of what his evidence seems to support. Such responsiveness is better than not being responsive to his evidence at all – for most attitude assignments in $J[X]\setminus J_{[X]}$ the two will be equivalent, because what Otis’s evidence seems to support generally is what his evidence supports – so the norm of compensation will require Otis to believe what his evidence seems to support. So Otis is excused for believing $\sim p$.

**Logic Test**

142
Ann ought to believe that \( \varphi \). But Ann can’t comply with any norm which would require believing that \( \varphi \), because Ann doesn’t have the computational resources to deduce every proposition. The least incoherent assignments in \( J[X], J_0[X] \) will be those which assign credence 0.5 to complicated logical expressions about which Ann has no evidence. So Ann is excused for suspending judgment on \( \varphi \).

If this account is right, not all mistakes are excuses. Constitutive norms might write mistakes into the conditions for justification. For example, mistaken eating during a fast isn’t excused in Islamic law, it’s justified: mistaken eating is considered not to be in violation of a fast, and hence a subject who eats mistakenly complies with the fasting norm to exactly the same extent, although not in the same way, as a subject who eats nothing. But we shouldn’t expect that constitutive norms will, in general, write mistakes into their conditions. When they don’t, we can be excused for making those mistakes.

4.5 Objections

4.5.1 Incorrect Extension

On this account JNB is also a guidance norm relative to TNB. Being excused means complying with a guidance norm but not its constitutive norm, so if we comply with JNB but not TNB we’re excused. That means that when we have a false justified belief we’re excused. Doesn’t that make excuses much too broad? We want to explain the difference between Right-Handed Society and

Justified Society

Winston is raised in a society which falsely teaches that left-handed people are slightly more likely to be younger siblings. This is widely believed on the basis of a number of statistically significant studies and there is no contradictory evidence available. In fact these studies happen to be misleading, and the perceived correlation has occurred by chance.

Winston has a justified false belief, but there is an important normative difference between Winston and Wilson. This account seems to put both subjects in the same category, as both are declared excused. Hence the account gets the extension of excuses wrong.

---

218 Hallaq 1997 p.108

143
While there is an important normative difference between Winston and Wilson, that is already explained by the account – Winston is closer to the ideal epistemic subject than Wilson is. What still needs to be explained is whether Winston is in any way normatively worse off than a subject who complied with TNB. But Winston does seem to be worse off – Winston ought to have true beliefs. Winston also ought to try to have true beliefs, so Winston ought to be justified. But intuitively Winston ought to try to have true beliefs because Winston ought to have true beliefs. So Winston is normatively worse off than if he had complied with TNB. Whether we call Winston excused is a terminological matter; what matters is his normative status. Guidance norms can account for that normative status.

4.5.2 Excused without Inabilities

Intuitively a subject in an inhospitable environment isn’t lacking any ability. The subjects in Misleading Drink, Misleading Library and New Evil Demon don’t lack any ability, because it’s their environments which are causing the problem.

Abilities are always abilities relative to environments. Let’s say that S can₁ comply with N if they can comply with N in a hospitable environment, and can₂ comply with N in their actual environment. Then norms of compensation are intended to compensate for those cases where S can’t₂ comply with N, not cases where S can’t₁ comply with N; that is, we need compensation when our environment is inhospitable to norm compliance as well as where we lack an ability₁ to comply with a norm.

4.6 Conclusion

I’ve argued that excuses arise in a number of areas, so we have good reason to be committed to them. Existing accounts don’t seem satisfactory because they don’t solve the normative link problem. I’ve argued that norms of compensation can give an extensionally correct account of excuses, which is readily applicable to justification and which gives the right results for the cases which originally motivated the introduction of excuses. Norms of compensation are plausible because they’re necessary for giving an account of guidance, an account we would need even if we weren’t committed to excuses. Such an account looks like it might have a solution to the normative link problem, because guidance norms look like they have normative force; guidance norms are means to satisfy constitutive norms, so they might inherit the normative properties of constitutive
norms in virtue of being instrumental for satisfying those norms. If this account of excuses is to
solve the normative link problem, it needs to make the case that guidance norms do inherit those
normative properties. I’ll make that argument in the next chapter.
Chapter Five
Excuses, Guidance and Normativity

5.1 Must all norms be guiding?

I’ve argued that norms are guiding when a subject can comply with them in spite of their failings, and that a subject is excused when they comply with a norm which is guiding relative to another norm which has normative force for that subject but which they fail to satisfy. So far this tells us nothing about the normative status which attaches to excuses. Supposed we defined schmuidance norms as those norms which are described in some book in the nearest library, and a schmexcuse as compliance with a schmuidance norm but not some other norm which has normative force. There would be nothing normatively interesting about schmexcuses. In this chapter I want to explain how excuses differ from schmexcuses.

Norms have normative force in virtue of the consequences that they have. Norms have deontic consequences and axiological consequences. I understand the deontic consequences of a norm N to be those particular propositions of the form ‘S is in C and so S ought to φ’, ‘S is in C and so S is permitted to φ’ or ‘S is in C and so S is forbidden to φ’ which are entailed by N. A norm has deontic force when it has deontic consequences. I don’t assume that S must exist for N to have these consequences; norms have deontic consequences even if no subjects exist.219 Similarly, norms have axiological consequences when they have consequences of the following forms: ‘S’s φ-ing in C is good’, ‘S’s φ-ing in C is best’, ‘S’s φ-ing in C is bad’ and so on. Goodness and badness here relate to what is valuable about S’s φ-ing, not whether S has done what they ought to do.

Prima facie it’s not obvious whether deontic force and axiological force can come apart. Perhaps all norms have both; perhaps some norms have just one. But to have normative force at all a norm has to have consequences which aren’t purely descriptive, so they must have at least one of axiological or deontic force.

RST2 is supposed to be about deontic force – it’s supposed to be about why self-trust is the sort of thing we can do while doing as we ought. So the crucial question is the relationship between

---

219 If this sounds wrong, thinking of deontic consequences as propositions like ‘If S exists and is in C, S ought to φ’ captures everything we want about deontic consequences.
excuses and deontic force. To get at that I’ll start by exploring the relationship between guidance and deontic force.

A good explanation of that relationship needs to answer Schechter’s challenge. It needs to explain why a subject, excused for believing p, very often ought to assert that p and act on p. And they ought to do so without any caveats or reasons weighing against their so doing; Wilson, the excused subject in Right-handed Society, should assert that left-handed people are gullible when asked.

I’ll argue that ‘ought’ is polysemous, and that we ought to comply with guidance norms in one sense while we ought to comply with constitutive norms in another. I’ll do so as follows. First, I’ll argue that any norm with deontic force admits of a guidance norm (sections 5.1.1-5.1.4). I’ll then argue that guidance norms are necessary for compliance with constitutive norms and therefore have instrumental value (5.2), responding to problems that arise with instrumental normativity.

5.1.1 Some Reasons to Think All Deontic Norms Are Guiding

Some philosophers endorse

**Strong Guidance Normativity (SGN)** Every norm with deontic force is guiding.\(^\text{220}\)

Whether SGN is true depends on the account of guidance we supply. We saw that SGN fails for lucid guidance, because no non-trivial norm is always lucidly guiding; since some non-trivial norms have deontic force, lucid guidance isn’t necessary for deontic force. SGN seems implausible for subjective guidance – the norm ‘If S is an innocent, don’t kill S’ has deontic force if anything does, but it isn’t subjectively guiding, because the triggering condition need not be part of a subject’s reasons.

On the other hand, before we supply an account of guidance, SGN seems *prima facie* plausible, for reasons I’ll discuss in the next three sub-sections.

5.1.1.1 Coincidence of Form

\(^{220}\) Gluer and Wikforss 2009 and Shah 2006 are both sympathetic to this view.
Here are some platitudes about norms with deontic force: their requirements are things that we should do, and we sometimes do do them; their triggering conditions are things that we are generally in a position to be receptive towards, and we are sometimes receptive towards them; requirements are about what you, a particular individual, ought to do, or perhaps sometimes about what groups of which you’re a part ought to do. There may be some exceptions or caveats needed here – perhaps if you’re responsible for someone else’s conduct, there is a norm which has force for you but its requirements include things which someone else does. Perhaps there are some triggering conditions which we can never be receptive to. But there is a striking clustering around a norm’s form with antecedent conditions which we’re receptive to and consequences which are largely our behaviour, broadly construed.

These are platitudes but they still require an explanation. We need to explain why norms with deontic force don’t have things like ‘Caesar conquers Gaul’, ‘The Moon continues to orbit the Earth’, ‘One of your sub-personal processes outputs signal s’ or ‘You come to feel pain through no action of your own’ as requirements. This is best explained by the guidance requirement. We can only be guided by a norm if its requirements are things that we do, not things which other people do, or which are beyond our control, or things which our sub-personal processes do. We can only be guided by a norm if we’re receptive to its triggering conditions.

Any account which denies SGN has to have an alternative explanation of these platitudes. It’s not obvious what such an alternative explanation would be. It wouldn’t be a good alternative to reiterate that norms with deontic force just are about subjects like us, our actions and the things we’re receptive to. This wouldn’t explain why deontic force only attaches to norms with this form.

It might be denied that norms with deontic force always have this form. Don’t we also say that knives ought to be sharp, kidneys ought to purify blood and so on? These functional norms seem to have deontic force without applying to subjects like us, or requiring actions from us. But these norms don’t have deontic force; they’re teleological, because they’re about the sorts of functions which certain objects have. The teleological ought doesn’t seem to have any bearing on deontic oughts. Biological weapons ought to maximise damage to innocents if they’re to be effective and achieve their function, but this has no deontic force. The existence of such teleological norms cuts off another alternative explanation of the form of norms with deontic force. We can’t explain that coincidence in terms of general properties of all norms, because not all norms do have that form;

221 This is considered in Jarvis 2012, Gibbons 2013, Simion, Kelp and Ghijsen 2016.
teleological norms don’t. So it’s something about deontic force in particular which means norms with deontic force have this particular form.

5.1.1.2 Explaining deontic force for a subject

There’s a difference between S taking N to have deontic force for them and being acquainted with N without attaching any normative force to N. This is the difference between being acquainted with Levitical dietary law and satisfying Levitical dietary law, but not satisfying Levitical dietary law because you feel any obligation to satisfy that law. A vegetarian can satisfy Levitical dietary law quite easily by coincidence, but Levitical dietary law wouldn’t have deontic force for them; on the other hand, moral norms around animal ethics would have deontic force for them, because they satisfy those norms in part because they take those norms to have deontic force for them.

Norms with deontic force are the sorts of things which can apply to us. No norm with deontic force could be essentially like Levitical dietary law is for the vegetarian, or it wouldn’t have deontic force at all. This also requires an explanation. SGN provides such an explanation; being guided by a norm entails taking that norm to have deontic force for you, because it entails complying with that norm. So if a deontic norm must be guiding, it must also be able to establish the distinction between satisfying a norm and taking that norm to have deontic force for you.

We can be right or wrong about whether a norm has deontic force for us. Plausibly, we sometimes wrongly take gender norms and norms arising from institutional roles to have deontic force which they don’t have. On the other hand, we seem right to take moral, practical and epistemic norms to have deontic force for us. A norm can start and stop having deontic force for someone; when someone becomes a parent they rightly take parental norms, norms governing their special duties to their children, to have force for them, and these norms can cease to have deontic force for the parent as their children age or if their children die. A good account of norms ought to be able to explain how a norm can have deontic force for someone. It’s hard to see how they could do that without appealing to the idea of a subject rightly taking a norm to have force for them. But to rightly take a norm to have deontic force for themselves would require it to be possible to take a norm to have deontic force, and that’s only possible when a norm is guiding. So not only is guidance essential for explaining the idea of taking a norm to have force for someone, it may also be essential for explaining the idea of a norm’s having force for someone at all.
5.1.1.3 Ought-implies-can

The following principle looks plausible:

**Ought-implies-can (OIC)** If N has deontic force for S and requires φ-ing when C, S can φ when C.\(^{222}\)

OIC looks plausible because there are many things which are good but which we’re nonetheless not required to do, and a plausible explanation of that is that while they are good they aren’t possible. Here are some things which would be good but probably aren’t possible: winning the lottery (so as to give the money to charity); preventing bad events which have already happened; reversing entropy; fixing things by wishing that they didn’t happen; fixing things by doing little rituals; writing down every moral truth; checking what to do based on the book containing all the moral truths. It’s striking that none of these good things seem to be obligatory, although some are much more valuable than many other things we take to be obligatory. Perhaps we could explain the non-existence of these obligations piecemeal, but it’s simpler to appeal to OIC to rule them out.

Perhaps we could deny that we lack these obligations, and maintain that we’re merely blameless, excused or have some other status which we’re mistaking for being permitted to flout these alleged obligations. But that seems strained, and systematically cuts across our intuitions about what we ought to do. I don’t have any intuition that someone who has failed to write down every moral truth has in any way failed to do as they ought.

SGN entails OIC, and explains why OIC obtains. SGN entails OIC because, if SGN obtains, every norm is guiding, so every norm can be complied with. If a norm can be complied with, it can be satisfied, so every time a norm requires φ-ing, it’s possible to φ. This explains OIC nicely, because it appeals to the entailment between compliance and satisfaction. If we wanted to deny SGN and maintain OIC, we’d have to maintain that every norm is possible to satisfy, at least through luck, but there are some norms which can’t be deliberately satisfied or complied with. That sounds strange; why would norms be such that we can always get lucky and satisfy them, but there’s no guarantee

that we can comply with them? If that sounds strange, then it’s plausible that SGN serves as a good explanation of OIC.

### 5.1.2 Reasons to Think Not All Deontic Norms Are Guiding

Even if these three considerations aren’t decisive, SGN seems to be in the right area. Any rejection of SGN ought to take the form of a modification to SGN which preserves as many of the explanatory virtues of SGN as possible.

Here are some reasons to reject SGN:

#### 5.1.2.1 Belief can’t be guided by TNB

The truth norm plausibly has deontic force – we ought to have true beliefs, and this isn’t just a teleological ought. But complying with TNB requires that you first have to work out whether TNB requires believing \( p \). But to do that you have to form a belief about whether \( p \) is true. This induces a vicious regress, so it’s not possible to comply with TNB. This is a problem with complying with TNB, not satisfying TNB. But since guidance is supposed to be possible compliance accounting for inabilities, not possible satisfaction, this is a problem for guidance.

This might be an argument against SGN for deliberative guidance. But it doesn’t seem decisive against SGN for compensation guidance. It’s not obvious that compliance requires explicit deliberation about whether \( N \) requires \( \phi \)-ing. Compliance may be a matter of having certain dispositions, or having a non-doxastic attitude of taking, or attaching deontic force to \( <C,\phi> \), none of which seem to require explicit deliberation.

#### 5.1.2.2 Failure of OIC

---

224 Hlobil 2015, Tracy 2020 and Jarvis 2012 take this as a reason to deny SGN, while Gluer and Wikforss take it as a reason to deny that TNB has deontic force.  
225 Broome 2014  
226 Boghossian 2003, 2014  
227 I adapt this from Hlobil 2019, which suggests that we infer by attaching inferential force to a pair of premises and conclusion; we might similarly think that we comply with a norm by attaching deontic force to a pair of triggering conditions and action-type.
Since SGN entails OIC, if OIC is false, SGN is false. There’s some reason to think that OIC is false. We can fail to believe as we ought if a belief is produced in us by a brain lesion, by unconscious bias, by computational inability, or by insuperable mistakes about our evidence. Smokey can’t help but believe that there is a lemon before him because he can’t help being mistaken about the nature of his evidence. Still, Smokey seems to have done something he ought not to do, because he hasn’t believed what his evidence supports. Similarly, someone like Ann, who is unable to perform complicated calculations about a given tautology, is failing to believe what her evidence supports. Someone like Wilson might have a deep-rooted and permanent bias against left-handed people, but Wilson ought not to have such a bias. And if Alvin is convinced that he will shortly be president as a result of a brain lesion, Alvin is failing to believe as he ought.\textsuperscript{228}

While the data supporting OIC still need explaining, this suggests that we can have obligations which we can’t meet. These obligations seem to be closer to things which we can meet, or which other members of groups to which we belong can already meet. Lots of people manage not to be biased against left-handed people or fail to believe they’ll shortly be president, and Smokey and Ann very often believe things which are supported by their evidence. Writing down all moral truths in a book, or reversing entropy, or any of the other things motivating OIC in 5.1.1.3 seem much further from what we or anyone like us are able to do. This isn’t yet to explain all of these data, only to point out that something more nuanced than OIC is required. But then OIC must be false, and since SGN entails OIC, SGN must be false.

5.1.2.3 The Anti-luminosity Argument

As discussed in the last chapter, no norms with non-trivial requirements can have luminous triggering conditions. If guidance is lucid guidance, then no non-trivial norms can be guiding. But this could extend to all forms of guidance which involve compliance if complying with a norm requires being in a position to know that the norm requires φ-ing. Hughes\textsuperscript{229} argues that this is the case, because that N requires you to φ has to be a reason for you to φ. If you think that your reasons are what you know, then complying with N requires being in a position to know that N requires you to φ.

\textsuperscript{228} This example is taken from Plantinga 1993.
\textsuperscript{229} Hughes forthcoming p.12
There seem a range of equally plausible options which deny that ‘N requires φ-ing’ must be a reason for which you φ (even if it is also a reason to φ, it’s not a reason on which you base your φ-ing). If any of these views are right, we can comply with N without being in a position to know that N requires φ-ing. What’s crucial is that we can explain the relevant Moorean phenomenon:

**Compliance Moorean Phenomenon (CMP)** It’s never reasonable to comply with N in φ-ing, and disbelieve that N requires φ-ing.

CMP can be accounted for, on each of these views, without assuming that we must be in a position to know that N requires φ-ing.

But the anti-luminosity argument can still cause trouble for SGN as long as it’s pointed out that sometimes lacking knowledge prevents us from complying with a norm. Consider TNB. True belief can be Gettiered and so can be true by luck. But if we only satisfy TNB through luck, have we really complied with TNB, rather than merely satisfying it? In what sense have we φ-ed because TNB requires φ-ing if we would have φ-ed if TNB didn’t require φ-ing?

But as Lasonen-Aarnio argues, this problem of non-accidentality also extends to knowledge. Total non-accidentality seems impossible to avoid, no matter the way of explaining the Inferential Moorean Phenomenon. We ought to be content with any account of compliance which explains CMP and not hope for an account which can give us non-accidentality.

### 5.1.2.4 Outweighed Reasons

You can’t φ for p if p is massively outweighed by other reasons. For example, you can’t move to Wisconsin because it has your favourite flavour of ice cream if it’s massively outweighed by much more important considerations, like your family and work life. This tells against SGN for deliberative guidance, but isn’t directly a problem for SGN for compensation guidance, because compensation guidance doesn’t require that you be able to φ for reasons but rather φ because a norm requires you to φ. This isn’t directly caused any difficulties by the existence of outweighed

---

230 See fn. 206 above.

231 This is adapted from the Inferential Moorean Phenomenon discussed in Hlobil 2014, 2019.

232 Lasonen-Aarnio 2019

233 This example is adapted from Schroeder 2007; see also Way and Whiting 2016.
reasons because a norm will require you to φ only if the total reasons speak in favour of φ-ing; there’s no requirement that component reasons could be acted upon.

Where it does cause problems is for the epistemic project approach to epistemic normativity. Suppose a norm N is part of our norms of justification because it recommends the best way to form true beliefs about p, and suppose your evidence supports p, so N requires believing p. But your total reasons might massively outweigh that evidence; perhaps p is that your partner is unfaithful, and you have very strong practical reasons not to believe this. Then we might not be able to believe p on the basis of our evidence because our evidence is massively outweighed by our practical reasons.

But I don’t think this parallels the ice cream case. The ice cream case is outweighed because the deliciousness of ice cream is too trivial for us to be guided by when we have outweighing reasons. But believing the truth about a partner’s trustworthiness isn’t trivial — it wouldn’t be absurd for someone to believe what their evidence suggested rather than following their practical reasons. Perhaps they would be unreasonable in doing so, but it seems perfectly possible to do so. That’s because satisfying curiosity is a value which is independent of other values and which is a plausible competitor with those other values. Acquiring delicious ice cream isn’t a plausible competitor with the reasons which outweigh the deliciousness of the ice cream, so the cases don’t parallel one another. Note that if we made the ice cream extremely delicious, the original argument becomes less plausible — if ice cream has the same appeal as heroin, it’s more plausible that someone could be guided by the deliciousness of ice cream to the exclusion of other values, even if those other values ought to outweigh acquiring ice cream.

5.1.2.5 Self-effacing Reasons

R can be a reason to φ even if φ-ing for R would spoil the reason. That your surprise party is at 5pm is a reason to go home on time, but it’s not a reason you can act for. Way and Whiting discuss a similar case: if you have a psychological delusion, for example, that you’re Napoleon, you have a reason to visit a psychiatrist, but not one you can act on. This causes problems for SGN even on the compensation guidance reading. If practical norms require you to visit a psychiatrist when you believe yourself to be Napoleon, it’s impossible to comply with those norms, because if you were in a position to comply with those norms in visiting a psychiatrist you wouldn’t be required to visit a

---

234 This example is also taken from Schroeder 2007 and discussed in Way and Whiting 2016; Kiesewetter 2016 denies that these are reasons at all because he’s committed to the view that reasons must be deliberatively guiding.
psychiatrist. Similarly, we can’t comply with a norm of attending surprise birthday parties, because we could never attend a surprise birthday party because it’s a surprise birthday party.

5.1.3 Arguments for WGN

Three of the above five arguments, 5.1.2.2, 5.1.2.3 and 5.1.2.5, seemed to cause real problems for SGN. It’s tempting to retreat to a weaker alternative to SGN. The thought is that, even if we can’t comply with a norm N, there is a norm in the vicinity of N, N*, which is guiding and which achieves some of what we can achieve through complying with N.

**Weak Guidance Normativity (WGN)** Every norm N with deontic force is guiding for an ideal subject, and for non-ideal subjects there exists a norm of compensation for N which is guiding.

WGN seems plausible for the following reasons:

5.1.3.1 Compensation Schema

In the last chapter I presented a way of generating compensation norms, one which could be applied to norms of justification. That there is such a schema suggests that we always will be able to find a satisficing norm of compensation for any norm which we are unable to comply with, at least as long as we’re able to identify some fundamental good which that norm aims at, and a measure of distance between norms in achieving that good. So WGN isn’t making an implausible claim that compensation norms always exist.

5.1.3.2 Everyday Practice

Srinivasan points out that we seem to shift to lucid norms when norms are non-lucid. Suppose we’re laying the table for a seder; we ought to lay one more plate than the number of guests. We often can comply with this norm, because we often know how many guests are coming to our seder. But where we don’t know how many guests are coming, we typically comply with a different norm aimed at achieving the same end but which we can comply with; we’ll form our best guess of the number of guests coming, and lay one more place than that. Of course, we might be wrong to do so; if the number of guests who come is different to the number we guessed would come, then we’ll have laid the wrong number of plates. But that we generally do make this switch to a relatively
guiding norm suggests that WGN is true because it suggests that there always is such a compensation norm which we can advert to when necessary.

5.1.3.3 Explains Derivative Norms

Norms of justification seem to have some relationship with the truth. They seem to be aimed at achieving true beliefs in a way which compensates for our inability to believe p just when p is true. A nice way to explain that is to point out that they’re necessary for guidance. Both norms of justification and the truth norm seem to make competing demands on what we ought to believe. A natural way to explain this is by giving the sort of structure found in WGN. The truth norm tells us what an ideal subject can do, and norms of justification tell us what we ought to do given our inability to comply with the truth norm.

5.1.3.4 Intuitions about Ideal Standards

Smithies\(^\text{235}\) points out that we are happy to consider individuals to be bound by standards which they can’t meet – someone with Capgras delusion is unable to respond correctly to their evidence, but intuitively they are wrong to have the beliefs they do. We might say that deluded subjects are bound by norms which apply to groups to which they belong. But why should individuals be bound by standards they can’t meet but not groups? Moreover, as Smithies again points out, human beings as a whole seem unable to meet ideal standards of chess or morality, but we still consider ourselves bound by these standards; there seems no reason to think epistemic normativity is different. WGN can explain this. We’re bound by ideal norms because there are norms which have deontic force for us but which we’re not able to comply with.

Moreover, denying that non-ideal subjects can be bound by the same norms as ideal subjects would entail the denial of closure under conjunction for obligation. One way in which non-ideal subjects fall short of ideal subjects is because they can’t jointly satisfy many of the obligations which they could satisfy individually. With a lot of time and effort we could probably ensure that our beliefs involving a single proposition p were probabilistically coherent. So non-ideal subjects would be obliged to have probabilistically coherent beliefs involving p. But we couldn’t have probabilistically coherent beliefs involving p\(_1\), p\(_2\), p\(_3\), etc. for each sentence in our language. So our obligations aren’t closed under conjunction. Some will be happy to deny closure for epistemic

\(^{235}\) Smithies 2015
obligation for independent reasons, but for those that aren’t this is an unfortunate consequence of exempting non-ideal subjects from ideal obligations. It’s an advantage of WGN that it doesn’t entail the denial of closure under conjunction for obligation.

5.1.3.5 Retains the explanatory virtues of SGN

Constitutive norms give obligations which some ideal epistemic subject could follow. We saw from 5.1.1.1 that deontic norms are supposed to have some relationship with behaviour, even if norms can be applicable to subjects who can’t comply with them; deontic norms place requirements on the actions of particular individuals under circumstances which they’re generally responsive to. This is explained by WGN. All norms with deontic force must be guiding for some subject, if only an ideal one. That means that the explanation for SGN still goes through; if deontic norms are to guide an ideal subject, they must require behaviour which that subject can do, and so they must require things which an individual subject can do under circumstances which they’re responsive to. So deontic norms which are binding for us non-ideal subjects retain the form predicted by SGN.

WGN can explain the phenomenon noted in 5.1.1.2, the distinction between satisfying a norm with which we’re acquainted and a norm having deontic force for us. If WGN is true, whenever a norm has deontic force for us, we can either comply with that norm, or comply with a norm which aims at the same end as that norm. Either way, this can explain the difference between a norm’s having and not having deontic force for us. Suppose you don’t know that your meal contains pork, but are under the misleading impression that it doesn’t. Then the Levitical dietary law will have deontic force for you if you comply with the relatively guiding norm ‘Eat food only if you believe it is permitted by the Levitical dietary law’, and won’t have deontic force for you if you comply with the relatively guiding norm ‘Eat food only if you believe it contains no meat’. So we can still explain the distinction between having and lacking deontic force using WGN.

WGN can explain the phenomena noted in 5.1.1.3. There we found that we’re generally not required to do anything impossible. WGN can explain this; deontic norms which are guiding for ideal subjects can’t require doing anything impossible, because not even ideal subjects can do what’s impossible. Still, as noted in 5.1.3.5 below, this doesn’t entail OIC for non-ideal subjects. The examples considered in 5.1.3.3 are those where we violate a deontic norm which ideal but not non-ideal subjects comply with. That WGN can accommodate the same phenomena as SGN without entailing OIC is an advantage of WGN.
5.1.3.6 WGN avoids the arguments against SGN

I argued that the argument of 5.1.2.1 (inability to comply with the truth norm) and 5.1.2.4 (we can’t be guided by outweighed reasons) weren’t convincing, but that the remaining three arguments did cause problems for SGN. Here’s how WGN can answer them.

5.1.2.2 WGN doesn’t entail OIC, and so the argument of 5.1.2.2 isn’t a problem for WGN.

5.1.2.3 WGN doesn’t require that all norms be lucidly guiding, nor does it require that all norms be non-accidentally guiding under all circumstances. It does require that in particular circumstances there be a relatively guiding norm, and that there should be no luminosity failures which prevent a subject from complying with that norm. This is compatible with the anti-luminosity argument, which shows that no non-trivial norm is always lucid, but doesn’t show that for all circumstances, there is no lucid norm which applies to a subject in that circumstance.

5.1.2.5 WGN can answer the problem of self-effacing reasons. When you’re unable to comply with a norm like ‘attend surprise birthday parties’, the relevant compensation norm will recommend norms which you can comply with – that might include things like ‘trust your friends to do good, unexpected things for you’, or ‘go about your normal routine unless you have reason not to’. This will satisfy your aim of attending surprise parties without your being able to comply with a norm of attending surprise parties, which is what we want guidance norms to do. Similarly, when a norm mandates attending a psychiatrist when you’re delusional, the relevant compensation norm which achieves the same end might require you to attend a psychiatrist when your friends urge you to do so (perhaps this will feel to you as though you’re humouring your doubting friends). Again, this compensation norm does what the constitutive norm is aiming at while being in a form which you’re able to comply with.

5.1.4 Arguments against WGN

5.1.4.1 Misinterpreting the evaluative/prescriptive distinction

It’s not obvious that the distinction between evaluative and prescriptive norms is best captured by the distinction between ideal and non-ideal subjects. But here the only alternative distinction seems
to be that between axiological and deontic consequences. Certainly some norms can only tell us what would be good to have, without saying anything about what we ought to do. But norms with purely axiological consequences are irrelevant. It would be best for the avalanche to go to the left without it being the case that the avalanche ought to go to the left, and perhaps we can say that the avalanche ought to go left, if that just means that it would be best if the avalanche went left. 236 But when we ask what we ought to believe, we aren’t asking about what would be good, or what we deserve; saying that a brain in a vat ought not to believe that they have hands isn’t like saying that an avalanche ought to go left. So this axiological understanding of evaluative norms seems to fail. The only other competitor that I can see would be the teleological reading, which I’ve already dismissed. 237 That leaves the ideal/non-ideal interpretation of the evaluative/prescriptive distinction. Since that reading seems to capture what we want out of the distinction, I see no reason to reject it.

### 5.1.4.2 There aren’t always guiding norms available

Hughes objects that derived norms aren’t that good at being guiding. Norms of justification aren’t that good at being guiding 238, because we can be biased about our own motivations, reasons etc. And our reasoning is shot through with such biases – it’s not as though we can say that justification is guiding more often than the truth norm, because every case of attempting to comply with norms of justification will suffer from ignorance of our own reasons. So even leaving aside the normative status of derived norms, why think that there even are any such things which do a better job of guiding than constitutive norms?

But all that’s necessary is that norms of justification be relatively guiding, and that doesn’t mean that norms of justification be guiding in any more occasions than the truth norm. What that requires is just that norms of justification eliminate some of our inabilities which prevent us from complying with the truth norm. That might not yet be enough to comply with norms of justification, because we might have further inabilities which prevent us from doing so. But we can continue finding norms of compensation until we have accounted for all such inabilities; we might be fallible subjects, but we don’t have infinite inabilities, and at some point we’ll arrive at norms of compensation which we can comply with. The importance of norms of justification as an example is

236 This example is taken from Gibbons 2013. This feels more natural when we say that England ought to be two goals up when England are 1-0 down; this isn’t a claim about what England should have done, but about what England deserve, about what the best result would be given standards of footballing fairness. This example was suggested by discussion with Danny Lev.

237 See 5.1.1.1.

238 Hughes forthcoming pp.16-9
that they illustrate the structure of being relatively guiding, but that doesn’t mean we must be committed to norms of justification ever being guiding themselves.

5.2 Why guidance norms have deontic force

5.2.1 Guidance norms are means to satisfying constitutive norms

I’ve so far argued that guidance norms exist when another norm fails to be guiding for a non-ideal subject. Satisfying a norm always requires a guidance norm, either because that norm is itself guiding, or because satisfying the norm can best be attempted via compliance with another norm. Complying with JNB is the best way to satisfy TNB, complying with IER is the best way to satisfy EER, and so on, given our various inabilities.

Guidance norms have deontic force because they’re the best available means to the end of satisfying a norm. Instrumental oughts are already the sorts of things we need to be committed to, so this isn’t positing anything ad hoc. If moving the knight to E5 is the best way of achieving checkmate, then it’s natural to say that you ought to move your knight to E5, given that you ought to achieve checkmate. The instrumental ought here doesn’t say anything about which norms you’re complying with, but rather which actions you ought to take. So the instrumental ought isn’t something which presupposes a prior understanding of the normativity of guidance norms, but is something we already understand before considering guidance norms and which we can therefore appeal to in giving an account of the normativity of guidance norms.

Moreover, guidance norms are intrinsically instrumentally valuable; there’s no way to comply with a guidance norm that isn’t the best possible way of satisfying constitutive norms. Guidance norms are specified by reference to a specific set of constitutive norms. That makes them different to habits which are helpful for complying with constitutive norms but which are unnecessary for so complying and which can be specified independently of those norms. Thinking about a proposition for five minutes before coming to a conclusion may be a helpful way of forming justified beliefs, but it isn’t necessary, and thinking about a proposition for five minutes is an activity understandable without any reference to norms of justification. But guidance norms are constitutively related to

those norms which they are intended to implement; it’s part of the nature of guidance norms that they are helpful for securing compliance with such norms.

Complying with a guidance norm is the best way available of attempting to satisfy a constitutive norm, because it’s the best way available of achieving the good a constitutive norm aims at, given our inabilities. The best way of satisfying TNB is by achieving the good TNB aims at, having true beliefs. The best way of doing that is by complying with JNB. So JNB is the best available way of satisfying TNB.

Does this mean ‘ought’ is ambiguous? It at least seems to be polysemous. We saw that there were problems with some ways of understanding the polysemy of ‘ought’. If that meant that ‘ought’ might refer to the objective ought or the subjective ought, there were problems. There seemed to be no way to give multiple readings to ‘ought’ which doesn’t assign the normative force to just one of them. You can say that you subjectively ought to do something if you have subjective reasons, but subjective reasons aren’t anything like objective reasons, and have no normative force – at best they have some hypothetical relationship with objective reasons and so would have normative force under different circumstances. But this doesn’t have the same problem. Guidance norms don’t have a hypothetical relationship with constitutive norms, as subjective reasons have towards objective reasons. It’s actually the case that guidance norms are the best way of satisfying constitutive norms. So the earlier argument against polysemy doesn’t go through.

This gives an answer to Schechter’s challenge. Schechter thinks that any account of excuses has to explain not only why an excused subject is blameless but also (a) often ought to do as they have done, and (b) does so in a way which fits with “a general account of norm-governed behaviour”.240

This account answers (a) because when guidance norms require φ-ing, we (instrumentally) ought to φ. That’s because we can’t opt out of attempting to satisfy constitutive norms, and the best way to attempt to satisfy constitutive norms is by attempting to comply with guidance norms. This also satisfies our intuitions that someone excused may be required to φ to just the same extent, or in just the same way, without any mitigation, as someone who’s justified; our intuitions will often be that someone excused and someone justified have both done as they ought to just the same extent and in just the same way because both will have complied with guidance norms, even if only one has

240 Schechter 2017 p.155
complied with the relevant constitutive norm. That’s because the guidance norm may be the most salient norm for evaluating both and judging whether both have done as they ought.

It meets (b) because guidance norms are understood in terms of constitutive norms and the normative status of complying with a guidance norm is understandable in terms of the status of satisfying a constitutive norm together with a general understanding of instrumental value.

5.2.2 Swamping Problems

We ought to comply with guidance norms insofar as they help us to satisfy constitutive norms. That suggests that if they are unsuccessful in so doing then we no longer ought to comply with them. Similarly, you ought to heat the oven if you want to bake a cake. But if you have none of the required ingredients and are unable to bake your cake, it’s not the case that you ought to heat the oven. There’s no obvious reason why guidance norms should be any different. But this gives the wrong result in lots of cases. It would mean that we ought not to have justified false beliefs. But in many cases we ought to have justified false beliefs. So this gives the wrong account of the deontic force of guidance norms.

Here’s my answer to this problem. Forming true, justified belief isn’t like baking a cake because you can opt out of baking a cake on a particular occasion if you aren’t able to bake a cake. But you can’t opt out of having some attitude towards a proposition which you’re committed to through relying on that proposition in action or to come to other beliefs. Since we have to have attitudes towards such propositions, the relevant constitutive norms apply. In particular, norms of justification always apply to any attitude you take towards a proposition.

Perhaps we could opt out of an epistemic project for specific propositions. But that would leave the epistemic project incoherent; if you’re relying on certain propositions to justify your commitment to other propositions, then you need to have some probability assignment which makes your reliance coherent. Otherwise, your attitudes to those propositions which you are content to hold within the epistemic project will also be incoherent. So we can’t individually opt out of an epistemic project for particular propositions, because doing so would have a knock-on effect for the justificatory status of other propositions.
That’s because it’s plausible that all propositions form one justificatory cluster, where a justificatory cluster is a set of propositions such that the removal of any proposition from the set would change the justificatory status of all of the others. If all salient propositions form a justificatory cluster, you can’t refuse to apply norms of justification to just one, because doing so will cut across all of your other attitudes. It’s not as though we have specific domains – beliefs about history, food, sports or whatever, which are isolated from one another; you couldn’t choose to opt out of believing anything about football without that having knock-on effects for your justification for beliefs in other areas.

Suppose you did try to opt out of an epistemic project for p. Then your credence in the truth-conduciveness of the faculty which outputted p ought to be lowered accordingly. That ought to lower your credence in the other propositions outputted by that faculty. That ought to lower your credence in the faculties which have also outputted those propositions. That ought to lower your credence in the faculties which outputted those propositions. It seems unlikely that we’ll be able to cut this process off before it covers all of our propositions – our faculties are too tightly interwoven for that, because every faculty has ties to other faculties which have ties to yet more.

For example, suppose you try to opt out of an epistemic project satisfying your curiosity about football for just one proposition, that Arsenal lost last Saturday. Suppose you’ve received evidence through testimony that Arsenal lost last Saturday. Then you would either have to arbitrarily ignore this evidence in just this case, or downgrade evidence from testimony, whether testimonial evidence in general or about football in particular. But if you downgrade such evidence in general, your attitudes about all sorts of other propositions will also have to be altered.

Moreover, even if there are some independent faculties with no connections with any other cluster, the fact that our faculties are all supposed to be responding to the world means that they are all vulnerable to interaction with other faculties, if that’s how future inquiry turns out. So opting out of norms of justification for one belief will lead to systematically unjustified credences across the whole of your attitude assignments.

The point generalises. When a norm has deontic force for us there can be no opting out of attempting to satisfy that norm. Doing so would lead to incoherence with our other actions and
attitudes. So when a norm has deontic force for us, we always ought to attempt to satisfy that norm in the best way available to us, so we always ought to comply with guidance norms. Even if we don’t satisfy the constitutive norm on a particular occasion, consistent compliance with guidance norms is necessary for compliance with guidance norms on those occasions which do lead to satisfying a constitutive norm.

This might make it sound like the reason we ought to have unsatisfied guidance now is because it will help us have some other instance of satisfied guidance. In the epistemic case that threatens epistemic trade-offs. That isn’t what I want to say. Instead I’m suggesting that instrumental rationality requires us to not only take the best available means towards an end but to obey consistency constraints in so doing. Instrumental rationality requires that we not be committed to something like ‘I ought to follow N when in C and not follow N when in C*, but there is no relevant difference between C and C*. If that principle of instrumental rationality is plausible, then it’s plausible that we instrumentally-ought to have unsatisfied guidance.

Steglich-Petersen worries that instrumentalist accounts of justification can fail to account for the fact that truth is an *internal* aim of justification.241 An internal aim of justification would be one which constitutes justification. On some accounts of instrumentalism, it turns out that justification is only extrinsically related to securing truth. For example, if processes are justified if they *tend to produce* true belief, then truth isn’t an internal aim of justification, because it’s possible to follow a justified process without having a true belief. It’s *tending to be true*, not *being true*, which is an internal aim of the process. But these two characterisations of internal aims can come apart. The aim of securing true belief plays a constitutive role in norms of compensation, because epistemic norms of compensation just are the best norms available for satisfying TNB. But we can fail to have true beliefs while complying with a compensation norm.

Steglich-Petersen might say that if we can *successfully* comply with a norm without having a true belief, then that norm can’t aim at truth, because successfully doing anything that aims at truth must guarantee truth, or else truth is merely contingently related to that thing. But while this seems right for processes which tend in the aggregate to produce true beliefs, it isn’t true for norms which are the best way of forming a true belief about this particular proposition. They seem constitutively

241 Steglich-Petersen 2013a

164
related to truth if anything is; how much more constitutively aimed at truth can something be than being the best possible way of satisfying TNB?

The worry with some instrumentalist accounts is that they seem to be trading off between helpfulness and satisfying a constitutive norm. But compensation norms aren’t like that. Because they’re the best available norms with which you can comply, there’s no trade off involved – norms of compensation are the best means you have to satisfy the constitutive norm, and there aren’t any alternative norms with which you could comply, whether less helpful or not, which would do better at satisfying the constitutive norm.

That explains why we ought sometimes to comply with a guidance norm without satisfying its constitutive norm. But we also need to explain the possibility of satisfying a constitutive norm without complying with a guidance norm. Instrumentalist explanations seemingly can’t explain how we can have unjustified true belief, because, if justification is supposed to just be the means by which we get true belief, then in a particular situation we have a true belief, we have employed some means by which we have acquired that true belief, and so we ought to count as justified, on the instrumentalist understanding of being justified. But intuitively it is possible to have a true unjustified belief, so the instrumentalist explanation is false.

There are two questions here. First, there is the question of how true unjustified belief is possible; in general, how can we satisfy a constitutive norm without satisfying a guidance norm? Call such a state of affairs a case of unguided satisfaction, and the case where we satisfy a constitutive norm while complying with a guidance norm a case of guided satisfaction; the question is how unguided satisfaction is possible. Second, there is the question of the value of true unjustified belief; intuitively it is inherently less valuable than true justified belief, but if the value of justification is exhausted by its instrumentalist value, how is this possible? In general, how is guided satisfaction more valuable than unguided satisfaction?

In answer to the first, guidance norms are those norms which compensate for an inability. That means they give the best possible way of satisfying a constitutive norm given our inabilities. They can only do that if they give general rules. When someone has a true unjustified belief, it’s because they’ve believed without sufficient evidence but have acquired a true belief as a matter of luck. But there’s no way for guidance norms to recommend believing as a matter of luck while still being
believing. ‘Believe p if doing so would satisfy the truth norm as a matter of luck’ isn’t a norm anyone can follow, and it’s no way of compensating for our inability to comply with the truth norm directly. So in cases where we do believe p truly as a matter of luck, we won’t have complied with a guidance norm, but will have satisfied the truth norm, and hence will have unguided satisfaction.

There have been a number of responses suggested to the swamping problem which could answer the second question. Some of these won’t work in the general case of guided satisfaction even if they can explain the relative value of justified true belief. I’ll first discuss those which won’t work and explain why before looking at one which will. Goldman and Olsson suggest two: first, that guided satisfaction raises the probability that future, related behaviour will also satisfy the constitutive norm, and raises it more than unguided satisfaction would; second, that guidance is valuable because the value of a token guidance satisfaction isn’t just a matter of the instrumental value of that token, but the instrumental value of the token of which it is a type.\textsuperscript{242} On this latter – call it type instrumentalism in contrast to token instrumentalism – guidance norms are instrumentally valuable for satisfying constitutive norms, but an instance of satisfying a guidance norm isn’t (or isn’t just) instrumentally valuable for satisfying a constitutive norm.

The first won’t work for our purposes, because it risks allowing epistemic trade-offs – if part of the value of having a belief is raising the probability of having other true beliefs, couldn’t we have beliefs which are valuable because their conduciveness towards forming other true beliefs outweighs the disvalue of believing without evidence? I’ve been thinking of epistemic projects as establishing norms for satisfying curiosity about particular propositions, and you can’t satisfy your curiosity if you’re engaging in these trade-offs.

The second doesn’t work either – aspirin is type-instrumentally valuable, but taking aspirin and curing a headache isn’t more valuable than curing a headache spontaneously. Goldman and Olsson think it only works when a value which is initially type-instrumentally valuable acquires autonomous status. Values can have autonomous status when they are initially type-instrumentally valuable but later come to be valued in their own right, separate from their instrumental role; the process of coming to be valued in their own right is the process of autonomisation. But it doesn’t look like guidance norms have always undergone such a process of autonomisation – it looks like many compensation norms can correct for idiosyncratic inabilities, and these aren’t widespread or long-term enough to achieve autonomisation. Consider a norm, N, which compensates for having

\textsuperscript{242} Goldman and Olsson 2009
swallowed a distorting pill. Suppose N recommends believing p, and p is true. Intuitively, if S complies with N in believing p then this is more valuable than if S merely satisfies N in believing p. But norms around distorting pills don’t seem to have undergone any autonomisation process.

This also applies to solving the swamping problem via epistemic value pluralism. On this approach we’d say that there are lots of different fundamental epistemic values, like knowledge, truth, justification and others, each with their own independent value; that would explain how justified true belief is more valuable than unjustified true belief, because justification would have independent value above and beyond its instrumental value. But for all sorts of guidance norms, satisfying them doesn’t seem to be a fundamental value, because guidance norms are often ad hoc and particular. So this doesn’t work in the general case.

But there is a solution to the swamping problem for justification which does generalise to all unguided satisfaction. Carter and Jarvis suggest that justification has instrumental epistemic value in maintaining a true belief. A true unjustified belief would have to be discarded because we would have no justification to maintain it. A true justified belief wouldn’t have to be discarded. So true justified belief has greater value than a true unjustified belief, and the value of justification is instrumental – justification is a means to maintaining true beliefs, rather than forming true beliefs. This extends to unguided satisfaction. If satisfying a norm requires guidance norms, then it seems plausible that continuing to satisfy that norm requires guidance norms. If that’s right, then guidance norms will continue to be instrumentally valuable because they’re required to maintain satisfaction of a constitutive norm, and hence guided satisfaction is more valuable than unguided satisfaction.

Dutant objects that once we disambiguate between temporal indices this maintenance solution fails. Consider the instrumental value of maintaining a house for having a pleasant house. That means that maintaining a house at time t is a means to having a pleasant house at t*. If t=t*, then the maintenance isn’t merely instrumentally valuable for having a pleasant house. If t* is later than t, then the maintenance doesn’t avoid the swamping problem – if our house were caused to be pleasant in some other way, through chance, then that house would be no less valuable than the house which we had maintained.

244 Carter and Jarvis 2012; see also Carter, Jarvis and Rubin 2013.
245 Dutant 2013
The last point has a slight asymmetry with belief, because it might be that the only way to maintain a belief is to continue complying with the norms which produced that belief. But even if that were true, we’d risk engaging in epistemic trade-offs again – believing that you will survive a serious illness is instrumentally valuable for maintaining a belief in the future, but that doesn’t make it valuable. If the maintenance solution is to work, it will have to be about maintenance when t=t*.

Carter and Jarvis would want to resist this; Carter, Jarvis and Rubin think that what’s important is that maintenance has no clear terminus, and there needs to be constant upkeep of our beliefs. They’d deny that we should think of the instrumental value of maintenance as obtaining in either of the cases where t=t* or t* is later than t, but over an indefinite period of time. Compare the instrumental value of pressing down the pedal on a piano for a period of a piece. Your pressing down the pedal at an instant both maintains the tone of the piece at an instant, and is a means to producing the right piece over the whole period. Similarly, abiding by guidance norms at an instant is a means to having the constitutive belief now and for maintaining the belief over a period of time. Having a winning lottery ticket is required to claim a boat, and maintain your claim to ownership over your boat. The lottery ticket’s constant possession is a means to constant possession of the boat, and that isn’t exhausted by being a means to possession of the boat now or at some future period, but over the whole period, including the present moment.

Dutant would probably object that the value of pressing the pedal takes two forms: first, it is non-instrumentally valuable for producing the right note at t; second, it is instrumentally valuable for producing the right piece of music. But the latter could be swamped (if the same note had been produced by chance) and the former isn’t instrumental and so isn’t a counter-example to the swamping problem.

The worry would be that this is still just the value of having a true belief now and the value of having a true belief later, and that there can’t be any extra value just from having both of these if neither has value on its own. But it isn’t generally the case that two things which are individually not valuable are collectively not valuable. You might dislike gin and tonic water individually, but enjoy a gin and tonic. Similarly, we value the successful maintenance of a process over and above the value of its component parts. It’s maintaining your production of a piece of music or the possession of a true belief which is primarily valuable, not maintaining that production or possession over sub-intervals. The former is plausible because what we care about is playing a piece

246 Bondy 2018 is also sympathetic to this point.
of music, not playing a succession of notes. The latter is plausible because what we’re trying to do is satisfy our curiosity. Satisfying your curiosity means getting your attitudes aligned with your evidence over a period in which you acquire no new relevant evidence. So there is at least one plausible response to the swamping problem which generalises to guided and unguided satisfaction.

Again, it’s the shift to norms of compensation which is crucial here. The analogy with the house is problematic because there are other ways of having a tidy house than maintaining it. But there are no better ways of satisfying TNB than complying with JNB. The point is that JNB isn’t merely helpful for satisfying TNB, but the best norm-governed way of satisfying TNB available.

We might object that instrumental rationality requires taking the best available means, not the best available norm-governed means. But I don’t think that’s true. If we’re to take rationally-assessable means to an end, they will themselves have to be norm-governed. So the only alternative rationally-assessable means to satisfying TNB will themselves be norm-governed and hence will fall under some alternative norm of compensation. But then they will be worse, by those standards, than the current means.

The worry might still be that having norm-governed means to an end isn’t any better than just having that end. But again, there is no way to have that end without some norm-governed means of doing so. So what we’re being asked to consider when we compare guided satisfaction with unguided satisfaction must be comparing two norm-governed means of achieving satisfaction, one worse than the other. But if one is worse than the other, then one will do a worse job of maintaining the belief than the other. This is the crucial premise missing in Carter and Jarvis’s argument. We should deny that it’s possible to have another means which would be equally good at maintaining the belief; this would be to deny that we’ve had guided satisfaction in the first place, because it would be to deny that we’ve followed a norm of compensation in the first place.

5.3 Objections

5.3.1 Norms don’t give unique recommendations conditional on violating those norms

Clayton Littlejohn has argued that in at least some cases when we violate norms of justification, then there isn’t a further justified option we can take – we’ve put ourselves in a situation where
we’re just bound to be unjustified. If there’s nothing further we can say here – not even what the second-best option would be – then norms don’t in general give recommendations conditional on your having made a mistake. Suppose you don’t believe some rational principle – that beliefs ought to be based on evidence, for example. Then what do evidentialist norms tell you to do? What’s the second best strategy? Couldn’t our norms just not supply an answer to someone who’s gone this badly wrong? If so, there are at least some inabilities for which no norms could compensate.

So far I’ve said that norms of justification can generate norms of compensation, conditional on having certain inabilities, by forming attitudes towards other propositions which are probabilistically coherent given the attitudes we’re bound to have given our inability to do otherwise. That’s given us a recipe for coming up with norms of compensation for norms of justification. Littlejohn would point out that figuring out which are attitudes minimise incoherence when we’ve made a mistake about norms of justification themselves is not easily accomplished – why think there’s an answer to be found at all? In particular, why think there’s an answer to be found when we’re mistaken about some deep principle of rationality, such as that beliefs ought to be probabilistically coherent? Moreover, there isn’t just a probabilistic constraint on norms of justification: we plausibly have to satisfy structural constraints to be fully justified. In particular, we need to avoid having any akratic beliefs – beliefs like ‘p and I ought not to believe p’. Satisfying these structural constraints doesn’t seem to fall out of having probabilistically coherent beliefs, so it’s not obvious that norms of justification will generate norms of compensation which also satisfy these constraints.

One caveat I want to add first is this: at most this objection ought to be that there is no unique recommendation given that the corresponding norm of justification does give a unique recommendation. That’s because it’s controversial which constraints on probabilistic coherence ought to be admitted, if any, which would generate a unique attitude assignment for every proposition given a set of evidence. I’m assuming for the purposes of demonstrating that guidance norms generate such a unique recommendation that norms of justification are also able to, and therefore that there is some appropriate set of constraints, above and beyond the axioms of probability theory, which would generate such a unique recommendation. But I want to remain neutral about which constraints they might be, and focus on the special problems which guidance

247 Littlejohn 2019
248 Smithies 2015 p.2782 argues that ideal rationality gives a hierarchy of options, and that non-ideal rationality removes options which can’t be fulfilled. On this view of ideal rationality, there is no question of what the second-best option is – it’s whatever the highest ranking option is after unfeasible options are removed.
norms introduce for generating unique recommendations, without regard to the general problems which probabilistic conceptions of justification face in this area.

I want to split this objection into three cases:

5.3.1.1 Mistakes about fundamental norms of justification

Suppose Jeff has a false but blameless belief about some fundamental norm of justification, such as Evidentialism or Enkrasia $S$ ought never to be committed to: $p$ and $I$ ought not to believe $p$.

Say that Jeff has been raised in a society which treats disbelief in one of these propositions as unquestionably certain. In these cases I agree with Littlejohn that Jeff is so badly wrong that it’s hard to see how norms of justification can tell him what to do – it doesn’t seem that there’s anything that they’d recommend, or if there is, that there is a unique thing that they’d recommend. Suppose Jeff disbelieves Evidentialism. Then, on the schema I’ve given, compensation norms would recommend believing whatever is probabilistically coherent with disbelief in Evidentialism. Most directly, that ought to raise the probability of Evidentialism’s most likely competitors, particularly Pragmatism For some $S$ and some $p$, it’s possible for $S$ to justifiably believe $p$ in part on the basis of practical reasons.

Because Pragmatism seems extremely likely if Evidentialism is false, disbelief in Evidentialism probably makes belief in Pragmatism justified, all else being equal. If Jeff believes Pragmatism, then Jeff will believe that many of his other beliefs, formed through norms of compensation conditional on Evidentialism’s being false, are not automatically guaranteed to be justified, because while they’ve been formed through norms of compensation which respect Jeff’s evidence, they don’t give any consideration to practical reasons. So Jeff will either violate Enkrasia by believing many propositions of the form ‘$p$ and I lack decisive reasons to believe $p$’, or Jeff won’t have those beliefs which norms of compensation recommend he have. Either way, norms of compensation don’t supply Jeff with a unique attitude assignment which intuitively captures what Jeff ought to do, given his constraints; Jeff seems to have gone badly enough wrong that the process of forming norms of compensation also breaks down.
How might this be a problem for WGN? It’s a problem only if norms of justification have deontic force for Jeff, and there are no norms of compensation for norms of justification with which Jeff is able to comply. The argument so far has suggested that there are no norms of compensation with which Jeff can comply while maintaining a disbelief in Evidentialism. And that disbelief is the result of an inability – Jeff is unable to do otherwise because of the society in which he was raised. So WGN is false.

To answer this I need to say something about the varying quality of excuses. At any one time we will have many inabilities, because there’ll be many things blocking us from complying with norms of justification. Some of these will be very difficult to get rid of, like our inability to believe all the logical tautologies. Some will be much easier; sometimes we’re just not in the mood to work out what our evidence supports, but it only takes a nudge to get us into the mood. Which norms of compensation we think are applicable to a subject depends at least in part on how good an excuse we have – we’ll hardly ever advert to norms which compensate for not being in the mood to assess our evidence, but we’ll hardly ever fail to advert to norms which compensate for our lack of logical omniscience. When we don’t advert to a norm which compensates for not being in the mood, our norm of compensation is just: stop being lazy, and then comply with the norm which would apply to you if you weren’t lazy. That is, sometimes norms of compensation compensate for an inability and sometimes they instruct us to get rid of our inability.

I’ll get into when and how a norm of compensation takes one of these two different forms in the next chapter, but for now it’s enough to say that whether a norm of compensation takes one of these forms seems to depend in part on whether there are available norms of compensation at all if a subject doesn’t get rid of that inability. If there aren’t, then the only available norm of compensation will instruct them to get rid of their inability. And that seems to be what happens in this case: Jeff ought not to disbelieve Evidentialism, and although doing so is very difficult, there is no alternative for him but to do so. This doesn’t risk trivialising WGN, because in many cases – in particular, all of the cases we considered in Chapter Four – there are available norms of compensation. It’s only because there is no available norm of compensation of the first type that we must advert to norms of compensation of the second.

5.3.1.2 Mistakes which threaten akrasia
Consider the following case:

**Habitual Modesty** Julian has a habit of being overly modest and is frequently disposed to think his beliefs are impermissibly held, a habit which Julian is unable to overcome. Julian believes that Johnson is Prime Minister on the basis of competent assessment of evidence which really does support that proposition, but due to his habitual modesty believes that he shouldn’t have this belief.

In a case like Habitual Modesty Julian’s higher-order belief is false and unjustified, and norms of justification would preserve enkrasia by recommending dropping that belief. Norms of compensation are supposed to give a recommendation which is possible, given Julian’s inability to comply with norms of justification, so norms of compensation are supposed to recommend a set of attitudes which are compatible with Julian’s having an unjustified false belief about his first-order belief. But why think, if his higher-order belief is excused and can’t be dropped, that there is a *unique* recommendation here? Julian could either drop his first-order belief, meeting the structural obligations of justification but failing to meet the substantive, evidentialist obligations of justification, or he could retain his first-order belief and be akratic. Nothing about the schema I’ve suggested for generating compensation norms seems to decide between these two options, so there isn’t a unique set of compensation norms here.

There is a general problem of how structural and substantive constraints relate. But there’s no reason to think that guidance norms have a special problem here. Perhaps there are some general principles relating structural and substantive constraints which are still to be discovered. Call the assumption that there are such general principles the *optimistic* view. If the optimistic view is right, we can generate compensation norm by applying those general principles as constraints on compensation norms, suitably specified such that they are conditional on our having certain inabilities. Without knowing what those general principles are, there seems no reason to think there’d be any special problem with applying them in this way so long as they are applicable to norms of justification themselves.

For example, suppose the right relationship between structural and substantive constraints is such that structural constraints always trump substantive. If that’s right, then there will be a unique norm of compensation recommending that Julian drop his first-order belief that Johnson is Prime Minister. Conversely, if substantive constraints always trump structural, then on the unique norm of compensation Julian ought to maintain both of his beliefs. In both these cases norms of
compensation can take on general principles from norms of justification and apply them to generate a unique recommendation, just as I’ve suggested in the schema given above.

The right relationship between structural and substantive constraints might be more nuanced than this. It might be that in some circumstances – call them C – structural constraints trump substantive, and in ~C substantive trump structural. A good account of the relationship between substantive and structural constraints on justification would give us some independent account of C. But whatever C turns out to be, norms of compensation can piggyback off of that account framed with respect to justification. If Julian is in C, then norms of compensation will take on the principle from norms of justification that structural constraints ought to trump substantive, and uniquely recommend dropping his first-order belief; conversely, if Julian is in ~C, norms of compensation will recommend maintaining both attitudes. Clearly this is incomplete unless we have some independent account of C, but this isn’t a special problem for norms of compensation, because we also need such an account to correctly characterise norms of justification.

It might be that there are no such principles relating structural and substantive constraints on justification. Call this the pessimistic view. The pessimistic view splits into two cases. On the dilemma view this means norms of justification can sometimes put us in epistemic dilemmas, cases where there’s no way – even for an ideal subject able to comply with norms of justification in all respects – to be justified.\textsuperscript{249} This doesn’t cause any special problem for norms of compensation; if the dilemma view is plausible for norms of justification, then it ought to be plausible for norms of compensation.

It might be objected that if epistemic dilemmas are admitted, then norms of compensation can’t be guiding.\textsuperscript{250} But this need not be the case. There are lots of ways of accounting for dilemmas. Perhaps all norms have the form ‘When C, φ, unless there obtains some further condition D such that C&D places you in a dilemma’. Then all norms would be guiding within just those areas where there are no dilemmas, but need say nothing about what happens when dilemmas obtain. Perhaps this would be enough for epistemic dilemmas to be compatible with WGN. But it’s incumbent on the dilemma view to come up with an account which would be compatible with WGN, because if it can’t, it can’t explain the motivations for WGN, particularly those which it inherits from SGN (5.1.1.1-5.1.1.3). Unless it can account for those motivations equally well, it’s more plausible that

\textsuperscript{249} Christensen 2007, 2010 argues for this view.
\textsuperscript{250} For more on epistemic dilemmas, see Leonard 2018, Hughes 2019, Littlejohn forthcoming a, Leonard and Cariani forthcoming.
dilemmas are cases where there are incompatible values available than cases with incompatible obligations. And as already argued (5.1.4.1), axiological force isn’t relevant here.

The intuitions behind the dilemma view also seem to be equally well-accommodated by the second strand of pessimistic view, which I’ll call the distinct norms view. On the distinct norms view there is no unique set of norms of justification. There is structural justification, with its own set of norms – presumably requiring enkratic attitudes – and there is substantive justification, with its own, presumably requiring believing whatever the evidence supports. Once we’ve got clear about these two kinds of justification there’s no further dilemma – structural and substantive justification don’t generate dilemmas any more than competing legal and moral norms generate dilemmas. Both sets of norms are therefore guiding for ideal subjects, and admit of guidance norms for non-ideal subjects. The existence of both sets of norms also explains why we’re intuitively drawn towards satisfying both structural and substantive constraints, and why we can’t satisfy both at the same time. The distinct norms view can explain this without denying that norms are always guiding for ideal subjects, so it seems preferable to the dilemma view. But if the distinct norms view is true then there isn’t a special problem with generating guidance norms out of norms of justification which arises from the tension between structural and substantive constraints, because there is no one set of norms of justification which admit of both constraints.

So on any plausible way of working out the relationship between structural and substantive constraints on norms of justification, we’ll always be able to generate guidance norms for non-ideal subjects according to the schema offered above.

5.3.1.3 Mistaken first-order beliefs

Suppose Jeremy is psychologically bound to believe p despite lacking sufficient evidence for p, where p is neither a proposition about norms of justification nor about the normative status of one of Jeremy’s other beliefs. Then norms of compensation will require holding those attitudes towards other propositions which are justified given that Jeremy is unable not to believe p. If q is highly likely if and only if p, then norms of compensation will recommend believing q; if q is highly unlikely if and only if p, then norms of compensation will recommend disbelieving q; if q is neither likely nor unlikely given p, norms of compensation will recommend suspending judgment on q; and if q is independent of p, norms of compensation will recommend assigning the same attitude to

251 Worsnip 2018 argues in favour of this view.
q as would have been assigned by norms of justification. By ‘q is highly likely/unlikely/neither likely nor unlikely given p’ I mean that q is highly likely/unlikely/neither likely nor unlikely given Jeremy’s background beliefs, conditional upon p. Since probabilistic coherence guarantees a unique assignment to q, given Jeremy’s background beliefs and belief in p, norms of compensation will always require exactly one attitude towards q.

What if q is extremely likely on p but extremely unlikely on ~p, and Jeremy’s evidence supports ~p, although he can’t help but believe p? Let p be the proposition that left-handed people are highly likely to be gullible, and q the proposition that some particular subject, Scott, is gullible. Given p, q is highly likely to be true. So the schema I’ve offered would recommend that Jeremy believe that Scott is gullible to preserve probabilistic coherence. But if we’re interested in justification because it guides us towards truth, wouldn’t it be better for Jeremy not to believe that Scott is gullible? Jeremy would sacrifice probabilistic coherence to improve accuracy, but that seems a trade worth making. In general, why should norms of compensation recommend preserving probabilistic coherence over maximising truth? As Staffel points out, there are no incoherence and inaccuracy measures such that if c is less incoherent than c*, c weakly dominates c* with respect to some accuracy measure. So the accuracy argument in favour of ideal probabilistic coherence doesn’t go through when suitably adjusted for non-ideal probabilistic coherence. That’s true even though it is also the case that whenever c weakly accuracy dominates c* according to the Brier score measure, c is less incoherent than c*, according to the squared Euclidean distance measure of incoherence.

But declining to follow coherence-maximising norms of compensation wouldn’t be guiding, for the same reason that TNB isn’t guiding except, perhaps, for omniscient beings. As ordinary, fallible, finite beings we aren’t able to comply with a truth norm, and so need JNB to guide us. Similarly, we can’t comply with norms of compensation which require believing p where p is true and it’s possible to believe p. We need some set of norms which are relatively guiding. If the move towards JNB is plausible in general on the grounds that believing p where p is sufficiently likely to be true is the only way to implement the truth norm, then the parallel move towards norms of compensation which recommend maximising probabilistic coherence ought to be equally plausible.

Staffel has one result which makes this move plausible. When a credence is incoherent, moving towards the closest coherent credence always accuracy-dominates. So if c is incoherent, c’
coherent and $c^*$ formed through a weighted average of each (i.e., for each $p$, $c^*(p) = nc'(p) + (1-n)c(p)$, for some non-zero $n$) then $c^*$ is more accurate than $c$. So norms of compensation which recommend minimising probabilistic incoherence will be expected to maximise accuracy, even if we’re unable to have the credences required by norms of justification. This can be true even though greater coherence among non-ideal credences doesn’t in general entail greater accuracy. So if norms of compensation recommend moving towards the nearest coherent credence, then they’ll be expected to improve accuracy. That they will recommend this isn’t guaranteed, because it might be that the nearest coherent credence isn’t one which you’re able to have. But in such cases you’d have no other options which would be guiding. So norms of compensation here would still be the best available way of satisfying both JNB and TNB.

It might be objected that there isn’t a good analogy between the move from TNB to JNB, and the move from norms of justification to norms of compensation, because norms of justification are necessary if we’re to incorporate our evidence into our beliefs at all. But the relevant alternative norm of compensation which is being suggested against my own view is one where we retain our initial evidence but incorporate it in a different way to what I’ve suggested. The suggestion would be that we block any evidence we have which is excused. The proposition $p$, that left-handed people are likely to be gullible, would be blocked as evidence because Jeremy’s belief in that proposition is merely excused, while his specific evidence relating to Scott would be justified and hence would be admitted as evidence. This would generate probabilistically incoherent recommendations which do better at recommending true beliefs than the norms of compensation which I’ve suggested, while also being guiding.

Call these two competing accounts of norms of compensation the blocked excuses and admitted excuses accounts. On the blocked excuses view, Jeremy shouldn’t believe that Scott is gullible and should thereby maximise true belief, while on the admitted excuses view Jeremy should believe Scott is gullible and should thereby maximise probabilistic coherence. Both norms appear to be guiding. What can decide between them?

I think the blocked excuses view is wrong because it doesn’t get the nature of sincere commitment right, and therefore doesn’t get the nature of belief right. It’s hard to see what would qualify Jeremy’s attitude towards $p$ as being belief, or sincere commitment of any sort, if Jeremy weren’t prepared to treat $p$ as a premise in at least some instances of reasoning. But if Jeremy is prepared to treat $p$ as a premise in this way on at least some occasions, then he’d violate Cross-Domain and
hence be insincere if he weren’t to consistently treat \( p \) as a premise. But this is just what the blocked excuses view recommends. Call a context where \( q \) is relevant a \( q \)-relevant context, and similarly where \( q \) is irrelevant a \( q \)-irrelevant context. Then the problem for the blocked excuses view is that, if Jeremy is to believe \( p \) at all, he must have the same inferential dispositions with respect to \( p \) in both \( q \)-relevant and \( q \)-irrelevant contexts, on pain of violating the sincerity constraint, but doing so is incompatible with what the blocked excuses view requires.

The blocked excuses view could claim that Jeremy ought to have a fragmented belief that \( p \). But that would be to violate the terms of the case – it was stipulated that Jeremy has a belief that \( p \) and can’t help but have that belief. Moreover, even on the fragmented belief approach Jeremy would still have an attitude towards \( p \) which would count as belief in some contexts. The blocked excuses view would have to explain, in those contexts, which attitudes Jeremy ought to have towards \( q \). Since we can’t make the same retreat to fragmented belief again, we have no way of stopping the same problems recurring at this, fragmented, level: either Jeremy should satisfy probabilistic coherence, and hence locally satisfy Cross-Domain, or Jeremy should not do so, in which case we still need to explain the sense in which Jeremy believes \( p \).

Perhaps the blocked excuses view should be seen as primarily requiring blocking excused beliefs where a fragmented excused belief isn’t active. But that isn’t in question, since the admitted excuses view ought also to take this position, because including inactive fragmented beliefs in Jeremy’s reasoning processes is impossible – that’s part of what being an inactive fragmented belief is.

The blocked excuses view could claim that there are reasons in \( q \)-relevant contexts which don’t obtain in \( q \)-irrelevant contexts. Cross-Domain only constrains Jeremy to have the same dispositions with respect to \( p \) across all contexts which share the same relevant reasons. But there seems nothing about \( q \)-relevant contexts which would raise the threshold for belief, which was the original motivation for the same-reasons restriction, and we’re assuming that Jeremy’s belief that \( p \) isn’t responsive to evidence, so gaining new evidence won’t make a difference to the case. There don’t seem to be any other reasons which could vary from \( q \)-relevant contexts to \( q \)-irrelevant that would plausibly make a difference to what Jeremy ought to believe, so this response doesn’t seem successful.

Finally, the blocked excuses view could deny that sincerity is an important enough constraint to trump maximising truth. Perhaps if we’re trading off between various desiderata – coherence, truth-
maximisation, and so on – then the value of sincerity also ought to be traded off against other values. The problem with this is that sincerity is necessary for an epistemically valuable attitude at all, and so maximising truth isn’t valuable unless it’s maximising the truth of sincere attitudes.

The blocked excuses view might say that this isn’t a problem for the belief that p, because the belief that p doesn’t appear to be valuable anyway – it’s false and unjustified, so why worry if it’s sincere if the only value of sincerity is as a necessary condition for these further epistemic values? The problem is that the belief that ~q also risks being insincere. To be sincere is to consistently engage in the dispositions associated with belief regardless of context, and if Jeremy believes ~q but fails to treat ~q as a premise that would diminish the likelihood of p, then Jeremy’s belief that ~q fails to satisfy Cross-Domain. That seems the right intuitive verdict in the above case – if Jeremy believes that left-handed people are gullible, and is unmoved by their supposed belief that Scott is left-handed yet not gullible, then in what sense is Jeremy not merely acting as if Scott is left-handed yet not gullible, rather than sincerely believing it? There’s intuitively something lacking about Jeremy’s belief that ~q here, a lack which Scott could pick up on and feel rightly aggrieved about. It would be natural for Scott to object that he is entitled to recognition as an individual free of false group stereotypes, and although Jeremy’s treating Scott as not gullible has some practical benefits for Scott it doesn’t secure that sort of recognition. The best way of explaining that lack of recognition is identifying it with a lack of sincere commitment, so this case seems to bear out the verdict of the Cross-Domain account that Jeremy’s belief that ~q would be insincere.

5.3.1.1-5.3.1.3 exhaust the cases, because any case not covered by 5.3.1.1 or 5.3.1.2 will be one in which nothing obstructs the normal operation of norms of justification generating a unique assignment of attitudes according to substantive constraints. The only way norms of justification could fail to generate such a unique assignment would be if structural constraints were relevant, as in 5.3.1.2, or if probabilistic coherence didn’t generate a unique recommendation and hence substantive norms of justification weren’t enough on their own to generate a unique recommendation. But the only way this could happen would be if those probabilistic norms were brought into question; so long as probabilistic coherence is in place, a unique recommendation will always be generated. So it can only be in 5.3.1.1 that substantive constraints fail to generate a unique recommendation. So we’ve exhausted the cases and shown that norms of justification always admit of a unique guidance norm on the compensation norm account.
5.3.2 The re-weighting problem

I’ve suggested that norms of justification involve abiding by evidentialist and structural constraints in ways which we sometimes can’t achieve. I’ve then suggested that there are corresponding norms which we can abide by. I’ve labelled the status induced by the former ‘justification’ and the latter ‘excuse’. That suggests a certain weighting of importance; it suggests that justification is more important than excuses, and that we should treat being excused as at best in second place to being justified. But why take this weighting seriously? We might think that ‘justification’, as it’s ordinarily used, applies to what I’m calling being excused. Similarly, we might think that ‘ought’, as it’s ordinarily used, picks out the ought of guidance norms rather than the ought of constitutive norms. If that’s true, that would make it seem like our weightings are wrong; the natural view prior to philosophical investigation is that excuses are more important than justifications with respect to what we ought to do. Such a re-weighting is plausible because guidance norms are just those norms which we’re able to comply with, and so seem closer to our ordinary talk of obligation and justification than constitutive norms. Call this the re-weighting problem.

Here are two responses to the re-weighting problem. First, this would have the central normative notion be guidance norms. But the central normative notion can’t be guidance norms because those are parasitic on other norms, and are ways of compensating for our inability to comply with those other norms. The structure of guidance norms is such that they can only hold interest if we’re already committed to the normative centrality of some other set of norms.

Second, this doesn’t require being revisionary about our ordinary normative talk, nor about our affective responses to ordinary talk and ordinary judgments about what we ought to do. In many ordinary situations all participants share similar inabilities, and where they don’t, we are disinclined to compensate for those inabilities. So it’s natural to use ‘ought’ to refer to the instrumental ought given by the norm of compensation which would compensate for the salient inabilities of all members of a situation. This would be true however we weight justifications and excuses. Similarly, our affective responses to failures to comply with guidance norms could be more extreme than to failures to comply with constitutive norms because failure to comply with a guidance norm often merits blame, whereas fail to comply with a constitutive norm which we couldn’t have complied with never merits blame. Again, nothing about how we ought to weight justifications against excuses follows from this.
5.3.3 Re-labelling Entitlement

Is this just relabelling *entitlement*, in particular the sort of entitlement that Wright seems to endorse? I’ve said that excuses are necessary to implement norms of justification at all for fallible beings like us; how is this different to entitlement, which is the status attaching to attitudes which are necessarily held to engage in epistemic projects? Both seem to be reasonable because they enable justified belief without being justified themselves. Are excuses just re-labelled entitlements?

At most this shows that entitlement is a species of excuse, because we can be excused for doxastic attitudes which are mandated by guidance norms even when those beliefs aren’t necessary for engaging in a practice at all. For example, we can be excused for not believing a particular tautology even though not believing that tautology isn’t necessary for engaging in a doxastic practice. This matters because it gets at the differing explanations for the normative statuses conferred by an entitlement as against an excuse.

That leads to two advantages for excuses. First, excuses are less *ad hoc*; they have the normative status they do because of a general link between the requirement that norms be guiding and our inability to comply with certain norms directly, and hence have the sort of deontic force which any sub-optimal but instrumentally necessary conduct carries. In focusing on the particular deontic force which obtains when a means partially constitutes an end, entitlements are needlessly specific; it’s better to see the deontic force attaching to entitlements in the context of the deontic force attaching to instrumental oughts in general. Second, excuses can explain why unjustified attitudes are reasonable even when they’re sincere; we saw that entitlement had to appeal to a special attitude of assumption to avoid this problem, but that approach ran into problems with sincerity. Excuses don’t require special attitudes and don’t have any problem with sincerity. Since excuses have theoretical advantages over entitlements, the difference between excuses and entitlements can’t be a matter of re-labelling.

5.3.4 Norms aren’t necessary for guidance

It might be admitted that constitutive norms must be guiding, if only indirectly, if they’re to have the normative force which norms require. But it’s a further claim that they must guide via generating further sets of guidance norms. Could constitutive norms not guide in some other way?

---

254 See section 0.5.
Perhaps we could observe those routinely satisfying constitutive norms and model ourselves on them. Such a view is plausible if some of the intuitions motivating virtue epistemology are plausible. Since this view circumvents guidance norms, it’s not clear what backs up the further claim that the guidance role of constitutive norms must function via sets of guidance norms.

But we would still need some account of what makes attempted compliance with constitutive norms reasonable – why would we be doing as we ought, in any sense, if we did so? That account would have to identify which ways a subject is bound to fail to comply with a constitutive norm and recommend second-best alternatives, and so would have to have the same sort of structure as norms of compensation. So simple maxims like ‘Follow the example of an excellent epistemic subject’ or ‘Do the best you can, given your limitations’ won’t be enough, because they leave ‘excellence’ and ‘best’ unanalysed. But it’s the job of an account of guidance for non-ideal agents to explain what excellence and best conduct is, given our limitations. Similarly, it’s the job of an account of guidance to explain why following a norm like ‘Do what someone who routinely satisfies a constitutive norm is disposed to do’ has any connection with what you ought to do.

5.3.5 Compensation norms introduce the wrong sort of reasons

Way points out that the problem of accounting for the rationality of subjects in cases like Misleading Drink lies in the difficulty of identifying good reasons for those subjects to behave as they do; it’s hard to find what reasons there are for Solomon to sip his drink, since that there is gin in his glass seemingly isn’t such a reason. This account has said nothing about the sort of reasons Solomon has when he’s excused. This is pressing because it might be thought that, for a norm to have deontic consequences, it must require that S comply with that norm when they have reasons for certain conduct – in particular the conduct licensed by that norm. If there is this close link between normativity and reasons, how can norms of compensation have deontic consequences if they say nothing about reasons?

There are reasons which come with norms of compensation. Norms of compensation give triggering conditions for φ-ing, and a triggering condition for φ-ing is a reason to φ. This leaves open the priority between the normativity of norms and the normativity of reasons. Perhaps compensation norms have deontic force because of their triggering conditions, or maybe their triggering conditions have normative force because they are part of compensation norms. I don’t think I need

\[255\text{ Way 2009}\]

182
to take a view on this priority debate. But in principle we could tell a similar story for reasons as I’ve told for norms. Suppose R is a reason to φ. Suppose ψ-ing is the best available means to φ-ing. Then R means that we ought to φ, and it also means that we ought instrumentally to ψ. As long as we can appeal to instrumental deontic force, it won’t matter whether reasons or norms take priority.

There is another problem with reasons. It might be thought that compensation norms introduce practical reasons because they include non-evidential reasons to believe p, namely, that S can’t fail to believe p. Perhaps Solomon’s inability to distinguish petrol from gin is a practical reason to believe his glass contains gin, for example; it’s certainly not an evidential reason to believe his glass contains gin. But that S can’t fail to believe p isn’t one of the triggering conditions for φ-ing in a norm of compensation, so it isn’t a reason for φ-ing. That a norm of compensation which requires Solomon to believe his glass contains gin has force for Solomon is because of Solomon’s inability, but within that norm Solomon’s reasons to believe his glass contains gin don’t include his inability.

5.4 Conclusion

I’ve argued that an attitude is excused when it is held in compliance with guidance norms but not constitutive norms. Constitutive norms require guidance norms because constitutive norms have deontic consequences, but those consequences can obtain without a constitutive norm being guiding. Where a constitutive norm isn’t guiding, that’s because a subject is non-ideal in some way and so is unable to comply with that norm. Constitutive and guidance norms must have this structure because there’s no other way to explain the data surrounding the form of norms with deontic force and the variety of intuitions surrounding OIC. I’ve argued that norms of compensation have the right deontic consequences; they have deontic force because we instrumentally ought to comply with norms of compensation in order to satisfy the relevant constitutive norm.
Chapter Six
Excuses and Self-Trust

6.1 Introduction

In the previous two chapters I arrived at a general account of excuses: we’re excused when we comply with a guidance norm while failing to satisfy its corresponding constitutive norm. In this chapter I’ll apply that account to self-trust. In particular I’m going to argue that we have an excuse for self-trust which obtains as a matter of metaphysical necessity. Any candidate for epistemic evaluation necessarily possesses concept, and possessing concepts entails self-reliance; since we’re metaphysically bound to rely on our own faculties, we’re metaphysically bound to be in an epistemic dilemma the best solution to which is to engage in self-trust, because the alternative to self-trust would be akrasia.

In the last two chapters I suggested a number of ways in which a subject could be excused. They can be excused when they lack the cognitive resources to believe what they ought to believe, as in Complicated Consequence. They can be excused when they are raised in a society which gives misleading testimony, as in Right-handed Society, or if they’re prevented from complying with a norm through distortions in reasoning, as in Distorting Pill. Even if we have a grip on the structure of normativity of excuses, we still need to work out what sorts of things can count as excuses, and why. When do circumstances merit adopting a norm of compensation? Is habitual laziness an excuse? An aesthetic taste for defying norms of justification? If not, why not?

There are two points to draw from the range of excuses surveyed so far. First, excuses vary in quality. Subjects might have all sorts of psychologically ineradicable attitudes, totally impervious to countervailing evidence. For such subjects to comply with norms of justification to any extent they’ll need to comply with norms of compensation which accommodate these poor attitudes, and in that sense will be behaving reasonably by the norms which directly impose obligations on them, but that doesn’t mean that we need rate their excuse very highly. While they might not have any choice about engaging in sub-par norms of compensation, that doesn’t detract from the fact that the norms which directly apply to them are sub-par. Conversely, suspending judgment on complicated logical tautologies seems to be better excused, because there’s nothing that could have been done,
whether within the subject’s control or not, which would have enabled them to believe arbitrarily complex tautologies.

If we can’t be free of all excused attitudes, we at least want to minimise the number of excused attitudes we have. The more excused attitudes we have the further our norms of compensation get from ideal norms of justification, and the worse we are at engaging in our epistemic projects. If we minimise our excused attitudes as much as possible then we’re left with excuses which we have as a matter of metaphysical necessity.

So metaphysically necessary excuses ought to be rated higher than any other excuses, because only they form part of our best possible norms of compensation. That would suggest that laziness or aesthetic pique isn’t a very good excuse – it’s very easy not to have these qualities, if we’re setting our standard by those qualities we’re bound to have as a matter of metaphysical necessity. It might seem that this is quite a demanding standard. But it should be unsurprising if our epistemic duties are quite demanding, even when our epistemic duties are determined by norms of compensation.

Norms of compensation are still in some ways idealised and removed from our everyday practices; they fix what we ought to do, but don’t yet fix what we can be criticised for, even if failure to comply with a norm of compensation is a necessary condition for criticisability. It might be that in everyday contexts we don’t hold someone’s laziness against them. But to fix the sort of normative standards which govern us, as fallible beings, we need to consider the best standards we can meet given our failings, and that means we need to fix our standards by those excuses which obtain as a matter of metaphysical necessity.

If we’re setting our standard in such a way then it would be natural to evaluate an excuse by the closeness of the worlds in which that excuse doesn’t obtain. If we were evaluating excuses by that standard we’d judge misleading society excuses to be more impressive than habitual laziness excuses, because having a society differ in one quality places us further from the actual world than having an individual differ in one quality. But this would be less impressive than not believing every logical tautology, because we’d have to be constituted very differently to be able to believe an infinite number of propositions without an algorithm to decide whether a candidate proposition is a tautology. I doubt the resulting hierarchy is well-ordered, and it is hard to know where to place cases like Distorting Pill – is the relevant closest world one where Otis hasn’t taken a pill, or one where he’s somehow able to resist the distorting effects of the pill? Still, distance from actuality
seems to serve as an acceptable heuristic for the quality of an excuse, and metaphysically necessary excuses are those which do best by the lights of that heuristic.

Suppose Wilson is raised in a misleading society, and also has a propensity towards a genetic disease which would leave him unable to have beliefs, although through a quirk of his genotype this disease happens not to be expressed. But it very easily could have been expressed. So there are very close worlds where Wilson lacks his unjustified belief. Does this put Wilson’s excuse in the same category as Willard, who is habitually lazy and is often unjustified because of his laziness?

Intuitively, that seems wrong. We’re not interested in how far from actuality we’d have to get for Wilson not to have the faulty belief, but rather how far from reality we’d have to get for Wilson not to have the faulty belief while being open to epistemic evaluation at all. That isn’t to say we aren’t interested in worlds where Wilson, as it happens, doesn’t have any beliefs. If Wilson manages to be a consistent Pyrrhonian in a close world, and thereby manages to lack his unjustified belief about the left-handed, then that counts towards our assessment of Wilson’s excuse. But the worlds which count have to at least be those where Wilson could have beliefs. So the best sort of excuse available isn’t one which is metaphysically necessary tout court, but rather one which obtains at all worlds where Wilson is open to epistemic assessment at all. Again, that doesn’t mean that the best excuse can obtain at just those worlds where Wilson engages in some epistemic project; it has to cope with the Pyrrhonian case, too. So this isn’t a recipe for constructing excuses which exploit the fact that any world where Wilson engages in an epistemic project requires Wilson to make certain assumptions, but rather a restriction to those worlds where Wilson could engage in any epistemic project at all. For simplicity I’ll refer to any excuse which fulfils these conditions as ‘metaphysically necessary’ or simply ‘necessary’.

The second point to draw from the excuses considered so far is this: being bound to hold an unjustified attitude towards p need not take the form of direct psychological compulsion. Wilson, raised in a misleading society, might have ended up with a psychologically ineradicable belief that left-handed people are gullible. This is sometimes said of actual cases of unconscious bias, when it’s said that particular biases are impossible for a particular subject to eliminate. But Wilson might also hold his belief with the ordinary level of intransigence – he’d be unjustifiably resistant to evidence because he’d underweight countervailing evidence on the basis of an unjustified prior, but

---

256 If this wasn’t true, then the genetic disease problem would generalise to show that, for some subjects, all of their excuses are of equal standing. Moreover, any of us could give ourselves sufficient brain damage to avoid all beliefs. This would put a strange upper bound on the quality of an excuse, which doesn’t seem to track our intuitions about the relevant excuses for assigning 0.7 to a tautology versus unjustified beliefs about the left-handed.
he wouldn’t be *ineradically* resistant to evidence. If Wilson encounters someone like Scott, a left-handed person who isn’t gullible, Wilson wouldn’t be sensitive to his evidence, but people in Wilson’s position can eventually change their views once they’ve received overwhelming evidence to the contrary. Wilson still seems to have an excuse in such cases. Wilson has an excuse because he holds attitudes which are consequences of circumstances over which he has no control. So the excuse for self-trust need not be direct compulsion to believe that our faculties are truth-conducive; it might be that a belief in TC is a consequence of something beyond our control, without TC being believed as a result of psychological compulsion. Here I mean that such a belief might be a *natural consequence* of circumstances beyond our control, not an *inevitable* consequence. If you’re raised in a misleading society, it’s a natural consequence that you come to believe lots of unjustified things because, in employing your faculties in the usual way, you’ll come to believe lots of unjustified things. Similarly, you might employ your faculties in the usual way and come to believe TC; if the circumstances in which you employ your faculties are unpropitious, this might get you an unjustified belief in TC.

In the next four sections I’ll argue that this is just what our belief in TC is like. We’re bound to have TC because it’s a natural consequence of having concepts. After providing some tests for concept-possession, I’ll argue that all candidates for epistemic evaluation are metaphysically bound to possess concepts (6.2). Possessing concepts commits us to ensuring that our concepts are *sound*, in the sense that they don’t exhibit any of the problems which concepts like tonk exhibit (6.3). But ensuring that our concepts are sound entails relying on our faculties, and a natural consequence of relying on our faculties is a belief in TC (6.4). So we have a metaphysically necessary excuse for self-trust. Finally, I’ll respond to some objections (6.5).

Is such an argument required? Perhaps it’s obvious that self-trust is psychologically deeply-rooted, that consistent Pyrrhonism is impossible for creatures constituted as we are, and we therefore have a strong excuse for self-trust. But notions of psychological deep-rootedness don’t help us make the fine-grained judgments which would allow us to compare our excuse for self-trust with our excuse for logical fallibility. The purpose of the argument of this chapter is to demonstrate that our excuse for self-trust is at least as good as any other sort of excuse we could have.

**6.2 Concept-Possession and Candidacy for Epistemic Evaluation**
In this section I’ll argue that all candidates for epistemic evaluation must possess concepts. I won’t attempt to provide an account of concept-possession in this section, but will instead briefly discuss three tests which are often accepted as requirements for concept-possession; if a subject meets all three tests with respect to some candidate concept it’s likely that they possess that concept.

6.2.1 The Generality Constraint

If S possesses a concept C then S can entertain any proposition of the form ‘x is C’, where ‘x’ is filled in with referential expressions in S’s vocabulary. The generality constraint was originally proposed by Evans on the grounds that crediting S with having an attitude towards a content of the form ‘a is C’ means we must also credit S with being able to have attitudes towards contents of the form ‘b is C’, ‘c is C’ and so on. Such a capacity seems necessary to credit a subject with predicating C of a; if we couldn’t also credit them with the ability to entertain other contents of the form ‘x is C’, then we’d be crediting them with doing something other than predicating when they seemed to entertain ‘a is C’. Perhaps instead the subject would be acting in the same way that someone with the ability to predicate C of a would behave in that situation, although the subject in question were relying on some sub-personal process rather than being able to form any conscious thought that a is C.

The generality constraint also seems well-motivated by the fact that our capacity to grasp propositions containing concepts seems systematic, in the sense that we generally are able to entertain ‘a is C’ if we can entertain ‘b is C’, and this would be explained if this feature were a necessary requirement for concept-possession.

The generality constraint needs some restrictions. First, it should apply only to those substitutions for ‘x’ which wouldn’t involve S in a category error – there seems no need to require that S understand what ‘The number is 2 is red’ means if S is to possess the concept ‘red’. Similarly, Charles Travis objects that S need not be able to grasp a proposition like ‘John is red’ even if they can entertain ‘The bus is red’. This is true not because there’s a category error – there are perfectly good senses in which John can be red – but rather because ‘red’ might not have a clear application in a new context. Again, we should see the generality constraint as restricted to those substitutions where there is no such change in context.

257 Evans 1982 p.104
258 Travis sees this as an argument against the generality constraint, but it also motivates this weakening of the generality constraint without requiring us to reject it entirely.
6.2.2 Inferential Dispositions

If S possesses C then S is disposed to make certain inferences which are characteristic of C. Peacocke’s view of concept-possession would motivate such a test: for Peacocke, to possess a concept is to find certain transitions, constitutive of the concept, primitively compelling.\textsuperscript{259} To possess the concept of conjunction, for example, is just to find the introduction and elimination rules for conjunction primitively compelling. But even if inferential dispositions aren’t built into the possession-conditions of concepts, we ought to expect concept-possession to lead to such dispositions because explaining how we can have such dispositions is part of what we want an account of concept-possession to do. Concept-possession is to involve more than discriminability, where we can discriminate between Cs and Ds when one of our sub-personal systems would have different outputs when in the presence of Cs than when in the presence of Ds. We could discriminate between Cs and Ds while lacking either concept if we weren’t able to recognise the possibility that there could be Cs and Ds outside of those contexts where we can discriminate between Cs and Ds. We can discriminate between squares and triangles while lacking either concept if we can’t recognise that a square in a dark room, or one too big or too small to be discriminated, is also a square. Having inferential dispositions captures this general capacity for recognition, because if we’re to have inferential dispositions characterising squares we must be disposed to make those inferences about any square. Any good account of concept-possession will therefore be such that possessing C enables S to have inferential dispositions involving C, so long as S is otherwise a fit epistemic subject. Note that S need not be motivated to make any of their dispositions occurrent; S can possess inferential dispositions even if they’re totally incurious and don’t desire to have any beliefs.

6.2.3 Thought-enabling

If S possesses C then S is able to think about C’s as such.\textsuperscript{260} If we have the concept ‘square’, we can think about squares. Moreover, once we possess a concept, we have this ability without needing to

\textsuperscript{259} Peacocke 1992, 2008

\textsuperscript{260} The thought-enabling test has been put forward by Jerry Fodor (Fodor 1998 p.3), which he this with what he calls ‘concept pragmatism’ - those views on which possessing a concept is “epistemic, dispositional, and normative” Fodor 2004 p.30 Fodor 1998, 2004 That is, while the inferential dispositions test attributes dispositions to form beliefs, dispositions which are often expected to be reliable and justified, the thought-enabling test doesn’t require any dispositions at all, and in particular no dispositions to form beliefs which could be evaluated by norms of justification.
rely on anything other than our general capacity to form thoughts. This contrasts with our capacity to have thoughts while relying on some other system, such as occurrent perception. As long as we can discriminate rock wrens from marsh wrens, then we can think about rock wrens and marsh wrens while we’re looking at rock wrens and marsh wrens. But our ability to have such thoughts relies on our being able to pick out the things we’re looking at and think about those things, and this isn’t an ability that will persist without our being able to rely on our visual experience of the birds; when we leave the area we may no longer be able to distinguish the two in thought, although the difference is obvious when in the presence of both. Concept-possession is supposed to explain the difference in our ability to form thoughts about rock wrens and marsh wrens, and our ability to form thoughts about crows and robins. I know enough about crows and robins to think thoughts about either without relying on seeing either, but I can’t do the same for rock wrens and marsh wrens. This is explained by my possessing the concept of ‘crow’ and ‘robin’ but not ‘rock wren’ or ‘marsh wren’.

6.2.4 Candidates for epistemic evaluation must possess concepts

We saw earlier that when Wilson has an unactualised propensity towards a disease which would prevent him having beliefs, we should not count those worlds at which that propensity is actualised as relevant to the strength of Wilson’s excuse. That’s because to be a candidate for epistemic evaluation plausibly requires the capacity to form beliefs; if it didn’t, it’s unclear why inanimate objects couldn’t be candidates for epistemic evaluation. Similarly, if it’s plausible that young children and animals can’t be candidates for epistemic evaluation, that’s probably because it’s plausible that they can’t form beliefs.

But possessing concepts seems to be necessary for forming beliefs. Concepts are, on the three tests outlined above, supposed to play a crucial role in facilitating our forming thoughts whose contents have inferential relations with other contents, which permit grasping other contents of a similar form, and which can be grasped without depending on experience. When we form a (sincere) belief, we exhibit these same features; we’re in a position to make inferences, to hold a belief in a variety of contexts regardless of experience, and to form beliefs in contents of a similar

261 Or at least, I couldn’t until just now; perhaps when I learned the names of rock wrens and marsh wrens that was enough to be able to form such thoughts. But there remain many birds whose names I don’t know which I could distinguish when watching them, and I would be able to think about without having the concept of their species. For more on this point, and the broader debate on conceptualism about mental states, see Heck 2000, Brewer 2005, 2006, 2011, Byrne 2005, Bermudez 2007, Toribio 2008, Schmidt 2015.

262 This position is also defended in Byrne 2021.
form. So being in a position to form beliefs entails being in a position to pass the above three tests for concept possession. If we could have the capacity to form beliefs without concepts, these tests wouldn’t be plausible. Since they are plausible, the capacity to form beliefs entails possessing concepts.

Moreover, even if it were possible to hold some beliefs without possessing concepts, I don’t think those would be the sort of beliefs we’re interested in when we ask whether Wilson would be a candidate for epistemic evaluation at a possible world. Such beliefs would, perhaps, be attitudes such as the belief that that marsh wren is taupe, where the believer relies on their being able to point to the taupe marsh wren before them to form their belief. Call beliefs of this type phenomenally-dependent beliefs. Perhaps it’s controversial whether such beliefs require possessing concepts, but I don’t need to make that claim to maintain my position. All I need is the plausible principle that to be a candidate of epistemic evaluation requires being able to hold beliefs without relying on phenomenal appearances as, perhaps, we do when we form such phenomenally-dependent beliefs.

That principle strikes me as intuitive; if I imagine Wilson only being able to form phenomenally-dependent beliefs, I feel reluctant to credit him as a competent epistemic agent or to hold him to the standards imposed by norms of justification. Wilson seems more like an animal or a young child than the mature subject which I’ve been taking as my model of a candidate for epistemic evaluation.

Moreover, it’s plausible if my account of sincerity and the role of sincerity in complying with norms of justification is plausible. To be a candidate of epistemic evaluation requires being able to form the sort of sincere attitudes which appear in norms of justification.263 But phenomenally-dependent beliefs couldn’t be sincere, because they couldn’t be held outside of certain contexts; such cross-domain variation is, on my account, precisely what constitutes insincerity.264 So the ability to form such beliefs isn’t enough to place a subject under the standards imposed by norms of justification and be a candidate for epistemic evaluation.

263 By this I mean that we can form beliefs, disbeliefs, suspended judgments, acceptances and so on, not that we can form the very sincere attitudes mandated by norms of justification, such as the belief that a complicated tautology is true.

264 We might worry that when we leave the area in which we can see taupe marsh wrens, we lose our reason for believing that that marsh wren is taupe; hence dropping a phenomenally-dependent belief wouldn’t be insincere, because the cross-domain variation would only obtain between contexts in which we possessed different reasons. But I don’t find it plausible that we do lose reasons in such cases; we can still remember our experiences, even if we lack the ability to conceptualise those experiences without those experiences being occurrent. It seems more plausible that we retain our reasons but are unable to access them. Hence we would have at most the sort of fragmented sincerity attached to fragmented belief, which intuitively doesn’t seem enough to qualify a subject as a candidate for epistemic evaluation.
Before moving on I want to say something about the sort of subject I have in mind as a candidate for epistemic evaluation, and the sort of subjects which we can leave to one side. There’s an intuitive difference between a creature which has a series of experiences but forms no opinions on the basis of their experiences and a creature who does form some opinions. It’s the latter sort of creature that we’re interested in. That doesn’t mean that we’re only interested in creatures with beliefs. We’re also interested in the Pyrrhonian who suspends judgment on whether their appearances are factive. Consider the following case:

**Monastic Sceptic** Maureen belongs to a group which lives in seclusion and refrains from forming any beliefs on the basis of her experiences. The group’s needs are satisfied without their needing to take any actions themselves, although they continue to breathe, digest food and so on involuntarily. When Maureen has an experience of a sunset, Maureen suspends judgment on whether the sun is setting. Maureen very often forms suspended judgments in the contents of her experiences, and when she doesn’t she forms no opinion at all.

The possibility of sceptical monks like Maureen seems salient for determining the quality of our excuse. But the variant case in which Maureen merely has experiences and forms no opinions at all doesn’t seem salient, because there’s nothing to distinguish Maureen from a creature which couldn’t take a rational stance on its own experiences. What’s crucial here is that although Maureen lacks beliefs, to be in a position to form beliefs at all Maureen must possess concepts; she declines to form any beliefs with contents composed of those concepts, but if her so declining is to count as epistemically reasonable she must have the option of believing.

One sort of possibility isn’t relevant and doesn’t need to be ruled out for the purposes of deciding the strength of an excuse. That’s the possibility where a subject lacks concepts but could come to possess them. Such a subject could, in some sense, be said to have the capacity to form beliefs, because they have the capacity to acquire concepts which would then facilitate their forming beliefs. But such a subject doesn’t, as yet, have the capacity to form the beliefs, only the capacity to acquire that capacity. So we need only consider those subjects with the capacity to form beliefs and hence which currently possess concepts. I’ll now argue that any such subject must rely on their epistemic faculties.
6.3 Self-reliance is a Natural Consequence of Possessing Concepts

To be a candidate of epistemic evaluation, S must possess concepts. But not all concepts are sound, in the sense that some concepts are similar to ‘tonk’ rather than similar to ‘taupe’. Tonk is that concept whose possession-conditions are the inferential dispositions to infer A tonk B from A or B and to infer A or infer B from A tonk B. ‘Tonk’ is a faulty concept because a subject possessing ‘tonk’ can infer contradictions through otherwise valid reasoning. ‘Tonk’ is a clear example of a faulty concept, but there may be other concepts which are faulty in this way; ‘time’, ‘truth’, ‘motion’, among many others, can seemingly also generate contradictory conclusions using otherwise valid reasoning.

Similarly, pejoratives are faulty although they don’t lead to contradiction. Consider a pejorative like ‘boche’, where ‘boche’ has the introduction rule: infer ‘a is boche’ if a is German, and the elimination rule: infer ‘a is untrustworthy’ if a is boche. ‘Boche’ need not lead to any contradictions, but it will lead to many unjustified beliefs about German people. So unsoundness is more than leading to contradiction, but in general any faultiness of conceptual construction which can lead to unjustified belief.

Let a concept C be unsound for S just in case: (1) it’s either the case that the beliefs S would have if they possessed C could easily be more frequently unjustified, or the beliefs S would have if they possessed C could easily be unjustified to a greater extent than would be the case had they not possessed C, and (2) (1) obtains because of inferences which S would be disposed to make in virtue of possessing C. Let C be unsound if C would be unsound for an ideal subject S, where ‘S’ is interpreted rigidly and S need not be an ideal subject in every possible world.

Tonk is unsound in the above sense. If an ideal subject possessed tonk, they could easily have unjustified beliefs, because they would be disposed to infer using the introduction and elimination rules for tonk, and those rules easily lead subjects into inconsistent beliefs; since inconsistent beliefs are never justified, ideal subjects would easily have unjustified beliefs. Similarly, an ideal subject who possessed ‘boche’ would be disposed to form many beliefs about the untrustworthiness of

265 I take tonk from Prior 1960.
266 I take this argument from Boghossian 2003.
267 I understand being unjustified to a greater extent to mean assigning attitudes which deviate further from what an ideal subject ought to believe than a subject would otherwise do, where deviation is determined by one of the measures discussed in Chapter Four.
particular Germans which wouldn’t be supported by their evidence, and so would form many more unjustified beliefs.

The above definition is first couched in terms of being unsound for a subject. That’s because whether possessing a concept leads to unjustified belief depends on the background beliefs and dispositions of a subject. To get a subject-free notion of soundness we might consider those concepts which are sound for most people, or which are sound for people behaving normally, or are sound for all mature epistemic subjects. But there is enough variation even within these groups that many concepts will be wrongly assigned sound or unsound. It seems better to think of soundness in terms of what an ideal subject would do if they possessed a concept, because that can serve as a model for what we ought to do. Other subject-free definitions of soundness seem at most relevant to what we’d be blameless for doing.

Many ordinary concepts, like ‘truth’, may be faulty in a way similar to ‘tonk’. That’s because possessing those concepts can lead to contradictory belief. But the disposition to make the inferences which lead to contradictions involving truth don’t appear to be constitutive of possessing ‘truth’. We appear to be able to refrain from making those inferences when they arise while retaining the concept of truth. This might be problematic, but we appear to get along well enough with ‘truth’ among our conceptual repertoire, while it would be disastrous to have ‘tonk’ among our conceptual repertoire because doing so would require being disposed to make the sort of inferences that lead to contradiction. 268 That’s why (2) requires that S actually be disposed to make the inferences constituting a concept rather than some weaker requirement.

Note that unjustified beliefs result only if a subject possesses a concept, not if a subject knows about a concept. We know about tonk, but we don’t possess tonk, because we don’t pass any of the three tests with respect to tonk: we can’t in general entertain propositions containing tonk or think of tons as tons, because we don’t have a grip on what a proposition containing tonk is really saying. Nor do we have any of the inferential dispositions associated with tonk. That contrasts with the concept of conjunction – we can entertain propositions of the form ‘P and Q’ for any P and Q, think of conjunctive propositions as conjunctions, and are disposed to apply the introduction and elimination rules for conjunction. So we do possess conjunction and we don’t possess tonk. And it’s only if S possesses tonk that they would have the inferential dispositions which would lead them into contradictory beliefs.

268 This is argued for in Braun and Sider 2007.
I’ll now argue that unsound concepts create a problem which means we must rely on our faculties. When we have no reason at all to think a putative concept is sound, then there’s a very good chance that a concept is unsound. That’s because most concepts that we’ll encounter will have very many constitutive parts; since it takes only the addition of one poorly-chosen part in the construction of a concept to render it unsound, most putative concepts we meet could easily be unsound, since they could easily have had one poorly-chosen part. Let a part be *unsound* relative to a collection of conceptual parts when that collection of conceptual parts is otherwise sound and is rendered unsound by its addition. For any collection of conceptual parts, there will be very many more parts which can render it unsound than render it sound; consider how many more potential concepts are like ‘grue’ than are like ‘green’, and how easy it is to render a sound concept like ‘green’ unsound by the addition of a disjunctive clause like ‘or is blue after such-and-such a date’. There are many such disjunctive clauses, so there are many unsound conceptual parts. So unless a subject has good evidence that their concepts are sound, they ought to expect their concepts to be unsound.

Then for each state Maureen has with a content p containing C, Maureen runs the risk of being committed to the following proposition:

**Conceptual Akratic Commitment (CAC)** I entertain that p and I refrain from judging whether C is sound.

Here I intend the first conjunct to be read as an expression of Maureen’s entertaining that p, rather than a judgment that S makes about what Maureen entertains; this is supposed to parallel the way in which 

269 ‘Grue’ is taken from Goodman 1983.
270 In much of the literature accounting for the unsoundness of ‘tonk’ appeal is made to some principle of harmony; the rules determining concepts ought to be harmonious in some way as to rule out tonk without being ad hoc. Different principles of harmony have been proposed; Belnap’s conservative principle (Belnap 1962, Dummett 1991), principles of un informativeness and independence of the meanings of the expressions involved (Steinberger manuscript), stability between the strength of an introduction and the weakness of an elimination rule (Tennant 2009, Steinberger 2009), and Read’s general elimination harmony (Read 2000, 2010). These are plausible ways of accounting for tonk within an inferentialist semantics, but if they’re to be satisfactory it’s they must rule out concepts which are unsound in my sense; unsoundness in the sense I’ve given is neutral between inferentialist and non-inferentialist semantics and has independent plausibility, because it’s independently plausible that we should avoid having unjustified beliefs. Principles of harmony are therefore of interest insofar as inferentialists can rule out unsound concepts, not as an initial characterisation of unsoundness.
271 There might be any number of other attitudes instead of entertaining, but all of them will run into the problem of tonk-like putative concepts which I develop below. For example, we could imagine a subject which makes no judgments but publicly exhibits its mastery of concepts through asking questions of others. We could also imagine such a subject giving commands, making jokes, making empty small talk, signalling tribal allegiance with verbal platitudes – there are any number of ways a subject can display mastery of concepts without being committed to the truth of the propositional content of their utterance. But they would be committed to the genuineness of the concepts constituting the contents of those attitudes – or so I’ll argue below, anyway.
which an ordinary assertion that \( p \) expresses a belief that \( p \). Similarly, the second conjunct is an expression of Maureen’s refraining from judgment, not the assertion that Maureen believes herself to be refraining from judgment.

There would be something wrong about being committed to CAC, in the same way there would be something wrong about being committed to

**Epistemic Akratic Commitment (EAC)** I believe that \( p \) and I ought not to believe that \( p \).

Perhaps commitment to EAC is permissible in certain circumstances, but in general a subject with such a commitment has done something wrong. They ought to drop their commitment to at least one of the conjuncts of EAC. That’s true because doing so is in general more likely to deliver true beliefs, and because there something inherently wrong about commitment to EAC. Epistemic akrasia, as exhibited by commitment to EAC, seems to violate a subject’s obligation to have a coherent set of beliefs, and such inter-level incoherence is problematic both because it makes it likely that a subject has a false belief and because incoherence seems inherently bad from an epistemic point of view.

Similarly, commitment to CAC seems wrong both because it makes it very likely that a subject has or will form false beliefs, and because it’s incoherent, and such incoherence is inherently bad. It’s likely to lead to false beliefs because it’s likely, from the subject’s point of view, that their concepts are unsound, and hence will easily lead them to form unjustified beliefs, which are themselves likely to be false. And it’s incoherent because concepts are supposed to do a certain sort of work: form contents which are graspable and which can be true or false. But unsound concepts don’t do that sort of work, because they don’t seem to form determinate contents at all, nor contents which could be true or false. ‘A tonk B’ isn’t a genuine content, and there’s nothing a subject could do that would go towards grasping that content, because there’s no consistent set of beliefs and dispositions they could have which would constitute grasping that content.

Because Maureen can’t rely on her faculties to acquire evidence that \( C \) is sound, she ought to refrain from judging whether \( C \) is sound. Subjects like us have lots of evidence that our concepts are sound, but that’s because we can rely on our faculties to track our generation and acquisition of concepts and observe that their constitutive parts are generally well-chosen; concepts which don’t have well-
chosen parts tend to lead to unjustified beliefs, and we can observe that such concepts don’t tend to be transmitted to others. That doesn’t mean that we ever need to reason in this way, or explicitly check each of our concepts. But at any time, normal subjects like us possess sufficient evidence establishing the soundness of our concepts, and so we are justified in our commitment to their soundness and can entertain thoughts containing those concepts without conceptual akrasia. But Maureen can’t rely on her faculties in this way, and so lacks the evidence that we have.

It might be objected here that Maureen does have evidence, because she has all sorts of perceptual experiences. She doesn’t believe the contents of those experiences, because she doesn’t rely on her perceptual faculties. But if perceptual experiences are evidence, then Maureen still has plenty of evidence. But because Maureen doesn’t have a higher-order commitment to TC, all of that perceptual evidence is defeated for Maureen. So although Maureen has plenty of evidence for the soundness of her concepts, she has no undefeated evidence, which is what matters.

Maureen neither believes nor asserts CAC, so why does CAC cause problems for Maureen? It’s problematic for Maureen because Maureen is committed to CAC, in the sense that Maureen’s epistemic position could only be reasonable if Maureen could appeal to CAC. It’s not necessary that Maureen possess the concept ‘sound’ in this sense of commitment; even if Maureen doesn’t have that concept, her position would only be reasonable if she were to first acquire the concept of soundness and then appeal to CAC in her own defence. And if we were to explain why Maureen is behaving reasonably, we would have to appeal to CAC in Maureen’s defence.

All that’s necessary for this to be the case is that that Maureen be committed to each of the conjuncts, and it’s a consequence of Maureen’s epistemic strategy that she both entertain p (or in some other way bear some non-affirmatory attitude towards p), while suspending judgment on whether the concepts constituting p are sound. And there’s something wrong with doing both of these at the same time, because in entertaining p you’re entering into a commitment that the content p is a legitimate content, and that commitment is undercut if you suspend judgment on the legitimacy of the concepts constituting that content.

But if Maureen has no beliefs, is there any risk of a concept being unsound for Maureen? A concept is unsound for Maureen only if it could easily lead her to have contradictory beliefs. But the sceptical monks considered above seem extremely committed to not having beliefs of any sort, contradictory or not. Such subjects couldn’t easily have any beliefs; there seems nothing stopping
us from stipulating that this is the case, anyway. So what would be the harm in having concepts without having gone to the effort of gathering particular evidence for whether they’re sound?

On this view it looks like the base rate of unsound concepts for sceptical monks is extremely low – much lower than it is for us. If that’s right, then perhaps Maureen and her fellow monks have propositional justification for the belief that C is sound, for any C, a concept or implicit conception in their possession. They lack doxastic justification, because they never form the belief. In fact, they probably lack doxastic justification as a matter of necessity, since, if they were disposed to form any beliefs at all, they would lose their propositional justification. Still, couldn’t having propositional justification be enough?

But again, our monks can’t have even propositional justification here because they don’t have undefeated evidence about their own position, and so they lack even \textit{a priori} evidence about the base rate of soundness of concepts for sceptical monks.

Similarly, even if the base rate of sound concepts is in general very low, it might be the case that it’s much higher for our very basic concepts. In particular, perhaps it’s very high for the sort of concepts we must employ in using our understanding to recognise correct fillings-out of concepts. It might also be very high for other sorts of basic concepts – the sorts of simple concepts like shape, position, colour and so on which we might possess in virtue of basic perceptual experiences. I think that’s right, but that’s because I have evidence generated by my faculties which implies that these more basic concepts are more likely to be sound. Maureen doesn’t have such evidence and so can’t appeal to different base rates here.

We might still think that, even if our monks are bound to be conceptually akratic, they just shouldn’t care that much. Conceptual akrasia is problematic because it runs the risk of generating false beliefs and because incoherent beliefs are inherently bad. But there’s no risk of our monks generating false beliefs. And they won’t care that much if unsound concepts don’t do the work that concepts are supposed to do – after all, that work is forming graspable contents which can be either true or false. But the only good our sceptics might expect to come from this is that we could then \textit{believe} these contents, ideally only the true ones; since the monks aren’t intending to have any beliefs, and don’t expect they could have enough true beliefs even if they did, they might see no reason to care about the failure of unsound concepts to do this bit of work.
But the point here isn’t whether being conceptually akratic is good or bad for sceptical monks. The point is that conceptual akrasia is plausibly forbidden by norms of justification. So in trying to find subjects who are candidates for epistemic evaluation who lack self-trust, we’ve found only subjects who violate norms of justification in other ways. That sets up a dilemma for us, not for these monastic sceptics. For any subject S, S can’t possess concepts, refrain from relying on their faculties, and refrain from engaging in conceptual akrasia.

How should this dilemma be resolved? I said in the first section that when we’re evaluating excuses we’re looking not just at mitigating circumstances but of natural consequences of those mitigating circumstances. That’s because the mitigating circumstances of someone raised in a misleading society are that they were raised in a misleading society, but their excuse is that under the circumstances they can’t help believing left-handed people are gullible. To account for this we need to count natural consequences of mitigating circumstances as excuses. I glossed ‘natural consequences’ as those consequences which arise from a subject employing their faculties in normal ways; the subject in the misleading society employs their receptiveness to trust in the normal way and, as a result, forms an excused belief that left-handed belief are gullible. Since we’re rational agents who can reflect on our situations, one way of employing our faculties in natural ways is by having reflective awareness of our situations. That motivates the following principle:

**Reflective Awareness Principle (RAP)** If S is in mitigating circumstances M and it’s impossible for S to have reflective awareness of their situation while φ-ing, then S is excused in M for not φ-ing.

For example, suppose Sterling is raised in a society which gives misleading testimony that anyone in pain is suffering from a viral infection. Then Sterling need not believe that anyone is suffering from a viral infection; in particular, he need not believe that he himself ever suffers from such an infection so long as he lacks reflective awareness of his situation. But when Sterling is in pain then, if he has reflective awareness of his situation, it would be impossible for him to fail to believe that he is suffering from a viral infection, and RAP would say that Sterling is excused for his belief.

In Maureen’s case, it’s impossible for Maureen to have reflective awareness of her situation and continue to refrain from judging whether her concepts are sound. It’s impossible to be reflectively aware of one’s commitment to CAC while maintaining that commitment, because the latter conjunct makes the former impossible to maintain. Maureen herself may not be capable of this reflective
awareness, because she may lack the concepts to reflect on CAC. But CAC strikes us as akratic because once both conjuncts can be reflected upon, they can’t both be maintained. Since entertaining some propositions is necessary for candidacy for epistemic evaluation, the only way for any subject to carry on as a candidate for epistemic evaluation with sufficient reflective awareness of their circumstances would be to drop the second conjunct of CAC – that is, to form some judgment about whether their concepts are sound. But the natural way to form a judgment is by reflecting on your evidence, and so the natural way for Maureen to form a judgment about the soundness of her concepts would be by reflecting on her evidence. So with sufficient reflective awareness of her situation Maureen would have to rely on her epistemic faculties.

6.4 Self-trust is Excused

This shows that self-reliance has a metaphysically necessary excuse. But if we aren’t to be epistemically akratic, we must also be committed to HOCTC. Once a subject is engaging in self-reliance, they can’t be reflectively aware of their situation without forming a higher-order commitment to the truth-conduciveness of their faculties. So not only self-reliance but also self-trust is excused as a matter of metaphysical necessity.

How much self-trust is excused? Perhaps this shows that trust in one very narrow faculty is excused – that doxastic practice made up of those transitions which are relevant to determining whether one’s necessarily-possessed concepts are sound. But what about our usual doxastic practices – sense perception, memory and so on? Is trust in these faculties excused?

Some transitions from each of our usual doxastic practices will be necessary to determine whether a concept is sound. Transitions from introspection, memory and reason will be necessary to identify and track the concepts we possess, as well as determine the permissible transitions of beliefs involving those concepts. It seems plausible that testimony and sense perception are both essential for providing us with our most basic concepts; if that’s right, then even someone with the minimum stock of concepts will still have received evidence from testimony and sense perception as to the soundness of their concepts. In general, all of our basic faculties provide us with lots of evidence as to the soundness of our concepts, both with respect to particular concepts and via inductive evidence of the regular successful application of concepts without generating contradictions. For a sceptical monk, all that evidence is defeated. But if we’re following the natural consequences of
possessing concepts, then we will rely on these faculties in gathering evidence for the soundness of our concepts.

I argued in Chapter One that when we rely on faculties, we incur a higher-order commitment to the truth-conduciveness of the doxastic practices which correspond to those faculties. I argued also that those doxastic practices are clustered by mutual truth-conduciveness – if one transition in a practice is truth-conducive then others in that practice are also highly likely to be truth-conducive. So in relying on our faculties even for a limited inquiry, we incur a higher-order commitment to the truth-conduciveness of that faculty in general. So insofar as our basic doxastic practices are involved in giving evidence for the soundness of our concepts, we incur a general commitment to their truth-conduciveness and hence our excuse for self-trust extends to our basic doxastic practices as such, not merely in respect to a limited inquiry.

6.5 Objections

6.5.1 Epistemic Enkratic Principles aren’t Plausible

There are a number of plausible counterexamples to epistemic enkratic principles. Consider Sophie Horowitz’s case:

**Sleepy Detective** After staying up late working on a case, Sam forms a first-order belief that Lucy is the culprit, based on his competent assessment of the evidence; the evidence does support Lucy’s guilt, and Sam is justified in his belief. Sam then receives testimony from Alice that nine times out of ten when Sam stays up late working his reasoning is sloppy and he tends to misassess evidence; it’s therefore likely that Sam’s evidence doesn’t support that Lucy is the culprit.

In examples like this, Sam seems to have first-order evidence for p and higher-order evidence that he lacks sufficient evidence for p. Sam is in a position to know that if he lacks sufficient evidence for p then he’s not justified in believing p. So Sam seems to have evidence for each conjunct of:

---

273 See 1.2.
274 Horowitz 2014.
275 Note that in the case given above, as in Horowitz 2014, Sam not only has evidence that his evidence doesn’t support p but also that Sam has misassessed his evidence. Ye 2018 argues that evidence of misassessed evidence is a distinct problem from evidence of insufficient evidence, but the distinct problem of misassessed evidence won’t be relevant for this discussion.
Justification Akratic Commitment (JAC) p and I’m not justified in believing p

If Sam has evidence for each conjunct of JAC, then it seems Sam’s total evidence supports JAC. This makes plausible a ‘level-splitting’ view – a view on which Sam’s first-order and higher-order attitudes seem to come apart. But I’ve relied on the assumption that such level-splitting is always unreasonable at various points. That assumption helps motivate the unreasonableness of conceptual akrasia. And it motivates HOCTC. If that assumption is questionable, the argument for self-trust seems to fail.

Such counter-examples seem relevant because, just as Sam’s evidential state makes it highly likely that Sam lacks sufficient evidence for p, so our evidential state makes it highly likely that we lack sufficient evidence for TC. It seems that on this view, if a truth-conducive doxastic practice outputs p, many subjects will be doing as they ought in believing p and believing that they’re not justified in believing p.

I accept that this follows for justification, and I agree with the level-splitting view with respect to justification. We can sometimes rightly believe p while rightly believing we lack sufficient evidence for p; since having sufficient evidence for p is necessary to be justified in believing p, we can rightly believe p while rightly believing we are unjustified in believing p. But I haven’t relied on the denial of level-splitting with respect to justification. Instead I’ve relied on

Epistemic Enkrasia S ought never to be committed to: p and I ought not to believe p.

But it can be the case that we’re unjustified in believing p while it’s not the case that we ought not to believe p. That can happen when we’re excused for believing p. So Epistemic Enkrasia can be true while the equivalent principle for justification doesn’t. Cases like Sleepy Detective aren’t sufficient as counter-examples to Epistemic Enkrasia, because those are cases where we have good evidence that we lack sufficient evidence for p. But I deny that, when we’re excused for believing p, lacking sufficient evidence for p means we ought not to believe p. The plausibility of conceptual akrasia can be motivated by Epistemic Enkrasia just as well as its justificatory equivalent. And Epistemic Enkrasia is all that’s need to show HOCTC. So there is no harm in retreating to this weaker principle and accepting level-splitting about justification.

6.5.2 Epistemic Enkrasia without Justification Enkrasia is Implausible

202
On this view every subject like us who engages in self-trust violates the plausible enkratic principle:

**Justification Enkrasia** S ought never to be committed to: p and I’m not justified in believing p.

Some philosophers would want to press on this and emphasise the specific strangeness of violating

**Knowledge Enkrasia** S ought never to be committed to: p and I don’t know that p.

It’s not clear that our intuitions support Epistemic Enkrasia more than they support either of these two other enkratic principles. This risks the argument being unmotivated, because it seems to involve picking and choosing between enkratic principles. But if some of those principles support the argument and others don’t, we ought to conclude that the argument fails, unless we have good reason to favour some enkratic principles over others.

Justification and Knowledge Enkrasia are both plausible on our intuitions. But our intuitions also seem to support:

**Infallibility Enkrasia** S ought never to be committed to: p and p might not obtain.

Unless we’re infallibilists about knowledge, it’s possible to satisfy Knowledge Enkrasia without satisfying Infallibility Enkrasia. So those sympathetic to a knowledge norm of belief will also have to pick and choose enkratic principles, because they’ll have to deny Infallibility Enkrasia while endorsing Knowledge Enkrasia.

Knowledge Enkrasia is also incompatible with

**Positive Evidential Enkrasia** S ought never to be committed to: My evidence makes p highly likely but I don’t believe p.

We can have evidence which makes p highly likely without knowing p, so we can’t satisfy Knowledge Enkrasia while satisfying Positive Evidential Enkrasia. Although Positive Evidential Enkrasia imposes a positive obligation to believe, that doesn’t seem implausible in cases where p is salient, important and interesting. But even in these cases, the Positive Evidential Enkrasia
requirement can come apart from the Knowledge Enkrasia requirement. Again, our pre-theoretic intuitions don’t seem to favour one enkratic principle over another; we’ll always have to pick and choose.

As long as there are competing norms constituting various good states we will have to make counter-intuitive endorsements of some enkratic principles over others. It’s not clear how to resolve that puzzle, but this isn’t a special problem for my view.

Moreover, this doesn’t undercut the motivation for the view, because Epistemic Enkrasia is a principle that everyone should endorse, regardless of how they understand ‘ought’ in the second conjunct of that principle. So long as Epistemic Enkrasia is well-motivated, the argument still rests on well-motivated premises. And the denials of Knowledge Enkrasia and Justification Enkrasia are also well-motivated, because they rely on plausible premises about the connection between obligation and guidance, premises which undermine those enkratic principles without undermining Epistemic Enkrasia. So there is no unmotivated picking and choosing of enkratic principles here, and the strangeness of denying Knowledge Enkrasia and Justification Enkrasia doesn’t undermine the view I’m defending.

6.5.3 Collecting Evidence for the Soundness of our Basic Concepts is Pointless

It might be objected that we need not rely on our faculties to acquire evidence for the soundness of our concepts because relying on our faculties wouldn’t be sufficient to secure the soundness of such concepts. Any evidence we could acquire for the soundness of our concepts would itself be conceptual, and so would presuppose what is to be shown, namely that all our concepts are sound.

But it seems possible to deny that evidence for the soundness of our concepts must itself be conceptual. Perhaps some of our evidence could have non-conceptual content, as some of our perceptual evidence may do. Of course some would deny that non-conceptual contents can stand in the appropriate relations with other contents to count as evidence; I’ve said that evidence consists of probability-raising contents, and perhaps probability-raising contents must be conceptual if they’re to be more than arational impulses which can raise our credences. To the extent that it’s still controversial whether contents standing in rational relations to other contents must be conceptual, this response should be controversial. Still, it seems an available response. And it puts pressure on the conceptualist position here – it must be permissible for us to be committed to the soundness of
our concepts somehow, and the non-conceptualist has a nice answer. So I don’t think this objection goes through.

6.6 Conclusion

I’ve shown that trusting our basic faculties is necessary, because doing so is necessary when we respect plausible enkratic principles and are sufficiently reflectively aware. It’s necessary under those conditions because we necessarily possess concepts, and possessing concepts imposes enkratic requirements on us which we can fulfil only if we rely on our basic faculties; such reliance is permissible only if we also have a higher-order commitment to the truth-conduciveness of those faculties. So we have the best possible excuse for epistemic self-trust.
Conclusion

In the introduction I set out to answer RST1 and RST2:

**RST1** What reasons do we have for our commitment to the reasonability of relying on our epistemic faculties, if any?

**RST2** In what way, if any, is our commitment to the reasonability of relying on our epistemic faculties reasonable?

In Chapter One, I argued that our epistemic faculties ought to be understood as doxastic practices, and that commitment to the reasonability of relying on those practices is reasonable only if commitment to the truth-conduciveness of those practices is reasonable. In Chapter Two and Chapter Three I argued that four strategies for answering RST2 failed: we aren’t reasonable in the sense of being justified, because bootstrapping, *a priori* and practical arguments fail to render self-trust justified, and substituting belief for some alternative doxastic attitude doesn’t avoid these problems, because the same problems apply to any sincere commitment to the truth-conduciveness of our faculties.

In the subsequent three chapters I argued that we’re excused for self-trust. In Chapter Four and Chapter Five I argued that excuses fall out of normative notions to which we’re already committed; in particular, excuses obtain when we comply with a guidance norm, construed as norms of compensation, but fail to satisfy the corresponding constitutive norm. We’re already committed to guidance norms to explain a number of independent phenomena, and norms of compensation capture the right sense of guidance to explain those phenomena. Moreover, we have an independent understanding of the sort of normativity involved in guidance norms; they have instrumental deontic force, and standard problems for norms with instrumental deontic force, including swamping problems, can be avoided when guidance norms are understood as norms of compensation.

In Chapter Six I applied this account to our excuses for self-trust. We’re excused because we’re bound to rely on our faculties to avoid conceptual akrasia; once we rely on our faculties in one instance, we’d be epistemically akratic not to rely on them in general. This is as good an excuse as
we can have, because it applies to us just in virtue of being the sort of fallible epistemic subjects we are, and hence obtains at all worlds where we’re candidates for epistemic evaluation.

I can now answer RST1 and RST2. The reasons we have for reliance on our faculties are the triggering conditions in all norms of compensation which apply to fallible beings like us. When we form beliefs in our everyday life through reliance on our faculties, our reasons for those beliefs are the triggering conditions of the norms of compensation which govern our everyday life; when we form a belief that there is a purple balloon in the sky upon seeing such a balloon, our reasons for our belief are that we’ve seen such a balloon, because norms of compensation require us to believe that there is a balloon under such circumstances. Because we ought to have a higher-order commitment to self-trust under any circumstances, norms of compensation will always require us to have such higher-order commitment, and any triggering conditions at all will be reasons for higher-order commitment to the reasonability of self-reliance. So our reasons for the two components of self-trust are, in the case of self-trust, our evidence for our beliefs, and in the case of higher-order commitment, any reasons whatsoever. That gives us an answer to RST1. Such commitment is reasonable not only because we have such reasons, but because norms of compensation apply to us as a matter of necessity, and norms of compensation require us to engage in self-trust; that gives us an answer to RST2. It’s reasonable to engage in epistemic self-trust because we’re excused for doing so: norms of compensation will always require us to engage in epistemic self-trust, although in doing so we won’t be justified.

This treatment of excuses and norms of compensation is helpful for answering RST1 and RST2, but it has application beyond the question of self-trust. For example, it gives a framework for better understanding parity arguments. We might give parity arguments for religious experience or mathematical intuition. These often suffer because there are specific reasons to doubt that these practices are truth-conducive – the problem of disagreement for religious experience, the Benacerraf-Field problem for mathematical intuition.276 Alston acknowledged that disagreement did some damage to our trust in religious experience, but was somehow not enough to overcome the presumption in favour of a socially-engrained practice. But it’s not clear how we are to make that calculation. The framework I’ve offered makes that clearer: if it’s necessary for us (on pain of conceptual akrasia and epistemic akrasia) to trust such faculties, then we’re excused in trusting them. It’s plausible that certain basic mathematical concepts are among those which any candidate of epistemic evaluation must possess; if that’s right, then if mathematical intuition is necessary for

gathering evidence about the soundness of such concepts, trust in mathematical intuition is also excused. It would take further argument to show the premises here; the point is that there is a clear framework for how such parity arguments ought to be made and evaluated.

It should be unsurprising that excuses, guidance and norms of compensation are important tools for understanding a range of normative behaviour, epistemic and otherwise. We’re fallible creatures and our normative theories ought to find a place for that fallibility. If this essay has been successful it will have cleared away some of the obstacles to completing that task.
Bibliography


Barnett, David James. 2014. ‘What’s the matter with epistemic circularity?’. Philosophical Studies 171: 177-205.


Bermudez, Jose Luis. 2007. ‘What is at Stake in the Debate on Nonconceptual Content?’ *Philosophical Perspectives* 21: 55-72.


Christensen, David. 2016. ‘Disagreement, drugs, etc.: from accuracy to akrasia’. Episteme 13 (4): 397-422.


Hornsby, Jennifer. 2000. ‘Personal and sub-personal; A defence of Dennett’s early distinction’. 
*Philosophical Explorations* 3 (1): 6-24.


Kiesewetter, Benjamin. 2016. ‘You ought to φ only if you may believe that you ought to φ’. The Philosophical Quarterly 66 (265): 760-82.


Oliveira, Luis R. G. 2018. ‘Deontological evidentialism and *ought* implies *can*’. *Philosophical Studies* 175: 2567-82.


Piedrahita, Oscar A. 2021. ‘Can hinge epistemology close the door on epistemic relativism?’. Synthese. https://doi.org/10.1007/s11229-020-02995-4


Pritchrd, Duncan. 2011. ‘What is the swamping problem?’. In Reasons for Belief, edited by Andrew Reisner and Asbjørn Steglich-Petersen, pp.244-60. Cambridge: Cambridge University Press.


Sharadin, Nathaniel P. 2016b. ‘Epistemic instrumentalism and the reason to believe in accord with the evidence’. Synthese. DOI: 10.1007/s11229-016-1245-3


230


Steinberg, Florian. Manuscript. ‘Harmony and logical inferentialism’.


[https://doi.org/10.1007/s11229-018-1679-x](https://doi.org/10.1007/s11229-018-1679-x)