



**University of
Sheffield**

Truth-Directed Testimonial Discriminations:
An Introduction and Application of the
Discrimination Condition on Knowledge

J. Herbert
University of Sheffield

This thesis is submitted for the degree of *Doctor of Philosophy*

Submitted 2023

Truth-Directed Testimonial Discriminations:
An Introduction and Application of the
Discrimination Condition on Knowledge

Jack Herbert
Department of Philosophy

This thesis is submitted for the degree of *Doctor of Philosophy*

September 2023

*Dedicated to my wonderful parents, Cath and Rich.
And to my excellent partner, Victoria.*

Acknowledgements

This PhD thesis would not have been completed without the guidance and support of many people. It has been very much a team effort to finish this work. Everyone who has extended kindness and support to me during what has been a very trying period, I thank you from the bottom of my heart.

When I started the PhD, my thesis looked quite different. Reading old drafts of that work makes me feel quite queasy now. Thankfully, I had an incredibly knowledgeable, empathetic, and supportive supervisor. I am indebted to Paul Faulkner, who has been the perfect supervisor. Paul has been enormously helpful, providing many comments on my work. Paul has also been incredibly supportive through his kindness. During the PhD, I was forced to take a medical leave of absence. He supported this and made sure that the process for taking that leave was not more difficult than need be. Thank you, Paul.

I am incredibly lucky to have had George Botterill as my secondary supervisor. George's enthusiasm for epistemology and the philosophy of science is infectious. He has helped me to appreciate the intricacies of the work I have been conducting, even when I have been finding it difficult to stay motivated. He is always very generous with his time, and I count myself very fortunate to have been under his supervision. I came away from every meeting with a new appreciation of a variety of topics. Thank you, George.

I am very grateful to both Jack Lyons, my external examiner, and Ryan Byerly, my internal examiner. Jack and Ryan asked probing questions which helped me to reflect on important points in a different way. The discussion we had during the viva was incredibly insightful and enjoyable. I thank you both very much.

I am grateful to many people for their support during my thesis: A big thank you to Felicity and Mei, who consistently provided friendship and support whilst we all navigated the difficulties of PhD life. Thank you to Jerry Viera, who supervised the module I taught with Victoria. Thank you to Eric Olson, who kindly supported

me during my mock viva. Thank you to Luca Barlassina for providing a good deal of helpful feedback on a variety of ideas. Thank you to Matt Vahdani who helped me by providing much needed support. Thank you to everyone in the Department of Philosophy. I cannot express how incredible all staff and fellow students have been. I feel very lucky to have been a student at the Department of Philosophy at the University of Sheffield.

My family have been enormously supportive throughout my studies. Thank you to my brothers Daniel and Peter. Daniel, I thank you for encouraging me to engage in philosophy. Peter, I thank you for encouraging me to pursue it. I lack the words to express the gratitude I feel towards my parents: Cath and Rich. I am so grateful for all that you have done for me. You have always shown me such love and support. Thank you.

Finally, I wish to thank my excellent partner, Victoria Kononova, and her wonderful parents, Irina and Dmitriy. Irina and Dmitriy, I am hugely grateful for the kindness and support you have shown me. Victoria, I am not able to articulate how thankful and appreciative I am for all that you have done. You have been a constant source of love and support. The final months of writing and rewriting the thesis have been very difficult. You have been incredible, and I am very lucky to have you in my life. You have managed not only to tolerate me discussing epistemology but have provided helpful comments which have led to important revisions. Thank you.

Contents

Acknowledgements	4
Abstract.....	7
1. Truth-Directed Testimonial Discriminations	8
2. The Non-Reductive Hypothesis	76
3. The Reliabilist Hypothesis	111
4. The Assurance Hypothesis	165
5. The Evidential Hypothesis	201
Bibliography.....	253

Abstract

Imagine that we are hiking together. Suppose that, while I am interested in birdwatching, I am not very good at it. In particular, I just can't tell the difference between buzzards and kites. This is unfortunate because, in the area where we are hiking, half of the birds are buzzards, and the other half are kites. Say that a buzzard flies down and lands on the trail in front of us. Let's add that I am disposed to form a belief about the sort of bird I see whenever I observe one. On this occasion, I happen to form the belief *that bird is a buzzard*. Do I *know* that the bird I see is a buzzard? Surely, the answer is no. I don't know that the bird I see is a buzzard because I can't *discriminate* between buzzards and other types of bird. I take it that cases like this tell us something interesting about the nature of knowledge, viz., to get knowledge, we need to satisfy a sort of *discrimination condition*. I apply this discrimination condition to the testimonial domain.

In this thesis, I argue that, to attain knowledge from the word of others, our testimonial discriminations must be *truth-directed*. The bulk of the thesis serves as an attempt to explain the *truth-directedness* of our testimonial discriminations. In doing so, I will examine four explanatory hypotheses generated by major theories in epistemology: i) the non-reductive hypothesis; ii) the reliabilist hypothesis; iii) the assurance hypothesis; iv) the evidential hypothesis. I defend the "externalist evidential hypothesis" *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*.

1. Truth-Directed Testimonial Discriminations

I

Imagine that we are hiking together. You know that I am interested in birdwatching. I enjoy keeping track of the different birds I see when I am hiking. The only problem is that I am not very good at it. In particular, I just can't tell the difference between buzzards and kites. This is unfortunate because, in the area where we are hiking, half of the birds are buzzards, and the other half are kites. Let's say that a buzzard flies down and lands on the trail in front of us. Let's add that I am disposed to form a belief about the sort of bird I see whenever I observe one. On this occasion, I happen to form the belief *that bird is a buzzard*. Do I know that the bird I see is a buzzard? Surely, the answer is no. I don't know that the bird I see is a buzzard because I can't discriminate between buzzards and other types of bird. I could have just as easily formed the belief *that bird is a kite*, in which case, I would have formed a *false belief*. So, the fact that I form a true belief in this case is lucky. I take it that cases like this tell us something about the nature of knowledge. They tell us that, in order to get knowledge, we need to satisfy a sort of *discrimination condition*. In this case, to know that the bird I see is a buzzard requires that I can discriminate between buzzards and other types of bird. I am interested in applying this discrimination condition to the testimonial domain.

Let me start by noting that, clearly, we know a great deal from the testimony of others. We come to know facts about geography, history, current events, our own birth, and much more through accepting what speakers tell us. My knowing, say, that Russia is the largest country on Earth, that Henry VIII succeeded Henry VII, that king cobras are venomous, and that I was a rather heavy baby are all the result of accepting the word of others. But in coming to know things through testimony, we rely on speakers being honest when they deliver their reports. This presents an issue when we recognize that testimony is not always delivered with the intention to

transmit knowledge; speakers sometimes lie to further their own ends, for example. And even when the speaker is sincere, their competence is not guaranteed. In short, sometimes speakers deliver *false testimony*. This is no secret. As a result, we don't just accept whatever we are told. Sometimes we *accept* what we're told and sometimes we *reject* what we're told. In other words, we *discriminate* between bits of testimony. The hope is that we accept *veridical* testimony and reject *false* testimony. This is because we want to attain *knowledge* and we can only know that which is true. Moreover, we wish to avoid having *false beliefs*, where possible.

In order to know some proposition, *p*, from your report that *p*, I must first *accept* what you tell me. A lot of work in the epistemology of testimony has focused on this part of the testimonial exchange, attempting to detail the conditions under which an audience is justified in accepting a speaker's report. Perhaps this focus is understandable given the generally accepted position that "knowledge would not be possible without belief justification (or something very much like it)"¹ (Audi, 2003: 3). But we don't just accept whatever we are told; we *discriminate* between bits of testimony. Arguably, what makes testimonial acceptance rational is that it rests on our capacity to make testimonial discriminations in a particular way. That is, a capacity to distinguish some property from some other property.

In his 1976 paper "Discrimination and Perceptual Knowledge", Goldman remarks that a particular sense of the verb "know" has been "neglected by philosophers" (1976: 772). This sense of "know" is that of "knowing from" as in: "I don't know you from Adam", "I know good from bad", "I know my left from my right", and so on (ibid.). I do not think that such phrases give any weight to the argument that "knowing from" is an instance of knowing, distinct from knowing-that or knowing-how. Rather, I take it that *knowing involves a discriminatory capacity*.

¹ Though not mentioned explicitly by Audi, I take it to be a generally accepted position that testimonial knowledge requires *justified* or *rational acceptance* (Coady, 1992), (Burge, 1993), (McDowell, 1998), (Graham, 2000), (Lackey, 2008), (Goldberg, 2010), (Sperber, Clément, Heintz, Mascaro, Mercier, Origg, & Wilson, 2010), (Faulkner, 2011).

In focusing on *testimonial discriminations*, I am focusing on what I take to be a neglected area within the epistemology of testimony. Historically, focus has been on testimonial *acceptance* or testimonial *belief*. Of course, questions regarding justified testimonial acceptance and justified testimonial belief are of great epistemological importance. However, by focusing on acceptance and belief, we are neglecting the *route* to belief. This is not to suggest that questions concerning our testimonial discriminations are entirely separate from the questions concerning testimonial acceptance. It is instead to say that a full account of how we attain testimonial knowledge should discuss testimonial discriminations as well as testimonial acceptance. Testimonial discriminations and testimonial acceptance can, I take it, be regarded as two sides of the same coin. And this is because however it is we regard, or explain, our testimonial discriminations will have some effect on how we regard, or explain, testimonial acceptance. We demonstrate our capacity to discriminate testimony in accepting the testimonies we do.

Consider an agent, *A*, who discriminates between bits of testimony on the basis of socio-economic class. If a speaker, *S*, belongs to, what *A* determines, is a higher-than-average socio-economic class, *A* tends to accept what *S* says. If not, *A* tends to reject what *S* says. Here, we see that an instance of testimonial acceptance is explained by that which explains the manner in which *A* discriminates between bits of testimony. In this sense, a given testimonial acceptance is impacted by *A*'s manner of making testimonial discriminations. Having a particular approach to testimony, in the form of discriminating between bits of testimony in a certain way, will restrict the sort of moves one can make with respect to testimonial acceptance. A better understanding of our testimonial discriminations should aid our understanding of testimonial acceptance then, given that testimonial acceptance is influenced by testimonial discriminations. My focus in this thesis is with the question of how we discriminate between bits of testimony in such a way as to allow for the attainment of justified belief and knowledge.

In this chapter, I will argue that, to attain knowledge, our testimonial discriminations must be *truth-directed*. That is, the way in which we discriminate between bits of testimony must be directed towards the truth if we are to attain knowledge via testimony. In doing so, I forward two arguments to the same conclusion. In my *argument from doxastic responsibility*, I propose that if an agent's testimonial acceptance is based on a discrimination which is not truth-directed, then that agent behaves in a doxastically irresponsible manner. Given that doxastically irresponsible conduct prohibits knowledge acquisition, a given instance of testimonial acceptance based on a discrimination which is not truth-directed thereby fails to allow for beliefs acquired in that manner to amount to knowledge.

In my *argument from epistemic luck*, I propose that if an agent's testimonial acceptance is based on a discrimination which is not truth-directed, then it would be a matter of veritic epistemic luck that that agent attained a true belief. Given that veritic epistemic luck is incompatible with knowledge, a given instance of testimonial acceptance based on a discrimination which is not truth-directed thereby fails to allow for beliefs acquired in that manner to amount to knowledge.

When we add to these arguments the premise that, by and large, we acquire knowledge from testimony, we can reach the conclusion that, by and large, our testimonial discriminations must be truth-directed. Since, if they were *not* truth-directed, by and large, we would fail to acquire knowledge from testimony, by and large. But how are we to understand these discriminations? More specifically, what explains their truth-directedness? Is a reliable process in operation? Are we being sensitive to evidence? Does trust have a role to play here? This thesis serves as an analysis of a variety of positions attempting to explain the truth-directedness of our testimonial discriminations. The aim of the current chapter is to establish the truth of the claim that a necessary condition on the acquisition of testimonial knowledge and justified testimonial belief is that our testimonial discriminations are truth-

directed. That we know a great deal from the testimony of others entails that the truth-directed testimonial discriminations condition is satisfied, by and large.

II

It is clear that we discriminate between bits of testimony. What is not clear, however, is the basis on which those discriminations are made and how this allows for justified belief and knowledge. Undoubtedly, we acquire a great deal of knowledge from the testimony of others. But we don't accept everything we are told; we discriminate. So, we must ask *how do we discriminate between bits of testimony in a way that allows for the attainment of justified belief and knowledge?* In this chapter, I will argue that our testimonial discriminations must be *truth-directed*, thereby introducing the truth-directed testimonial discriminations condition on testimonial knowledge.

Before presenting my arguments, however, it will be necessary to clarify what exactly I mean by the term "testimony". In his discussion of testimony, Coady distinguishes between a "natural" and "formal" kind (1992: 26). The latter occurs in the courtroom. The former occurs in, what Coady calls, "everyday circumstances" (ibid. 38). This would include a wide range of reports, such as "giving someone directions to the post office, reporting what happened in an accident, saying that, yes, you have seen a child answering to that description, telling someone the result of the last race or the latest cricket score" (ibid.). By contrast, "Formal testimony" is treated as evidence and is subject to restrictions. For example, courtroom testimony is generally considered only when it is a first-hand account (expert witnesses are an exception), and some testimony can be ruled as inadmissible². In my discussions of testimony, I allow for both "natural" and "formal" kinds.

² In English courts, 'hearsay evidence' is inadmissible, for example. Though I recognize there are some exceptions to this rule.

However, I wish to rule out some utterances as cases of testimony. In doing so, I agree with Faulkner, who writes that “not all intelligible utterances are testimony. Statements made in jest or uttered as part of a theatrical performance are not. This is because these statements do not present their contents as true. For the same reason a question is not testimony, though we can learn from what a question presupposes” (2011: 16). If you ask me whether Tanya has landed in Paris yet, I can infer *that Tanya took a plane to Paris*. In such a case, I am presented with an opportunity to gain knowledge, but this is not because you *report* that Tanya took a plane to Paris, instead you ask a question and I make an inference on the basis of this. I can come to know what is presupposed by your question, but given that questions are not *propositional*, and therefore, fail to have truth-values, they cannot present their contents as true.

To avoid ambiguity, I will state explicitly that, unless otherwise stated, my discussions of testimony are restricted to involving only those statements which present their contents as true.

III

I present the following argument in support of the claim that our testimonial discriminations must be truth-directed:

Social Fact (SF): We discriminate between bits of testimony.

Epistemic Principle (EP): If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience’s resulting belief cannot amount to knowledge.

Epistemic Fact (EF): We acquire knowledge from testimony (by and large).

Epistemic Consequent (EC): Our testimonial discriminations must be truth-directed (by and large).

Call this *the argument for truth-directed testimonial discriminations*.

SF requires little defence. Clearly, we neither accept nor reject everything that people tell us. If we accepted everything that we were told, then we would have a huge number of inconsistent beliefs. We would accept, for example, that climate change is a fact and that it is a fiction. We would also have to accept testimony we know to be false just because it is offered as testimony. If, on the other hand, we rejected everything that we were told, we would not form any beliefs from testimony. Here, one might respond that we can and do form beliefs in cases of rejecting a speaker's testimony, at least sometimes. I might, for example, form the belief that you are telling me a lie, or simply, that you want me to believe whatever it is you are telling me. Note, however, that in such cases I do not form a belief *from your testimony*. Instead, my belief is the result of an inference I make as a result of observing your behaviour. The reasoning might take the form of something like "this person told me that p . But I know that p is false. Maybe they are trying to deceive me". At any rate, what is important here is that, at least in our world, in the *actual world*, we discriminate between bits of testimony. The truth of SF should be uncontroversial.

Neither does EF require much in the way of a defence. Consider all that we know via testimony. My knowledge that Russia is the largest country on Earth comes from testimony, rather than traversing the entirety of the country with a ruler in hand. I am also yet to measure the chemical composition of air, but despite this, I know that it is comprised mostly of nitrogen. Again, I know this from testimony. And whilst there are of course cases in which we fail to acquire knowledge through the insincerity or incompetence of speakers, it is undoubtedly the case that we know a great deal from testimony. Should we subtract that body of knowledge classed as "testimonial knowledge" from our catalogue of knowledge, we would be devastated by the amount that we lost (Hume, 1748/1999: 77), (Reid, 1764/1993: 190), (Coady, 1992: 13). Having now discussed SF and EF, I can move to the focus of this chapter: my defence of EP.

EP holds that *if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge*. As such, EP makes it a necessary condition of attaining testimonial knowledge that we make truth-directed testimonial discriminations. In defence of EP, I forward my *argument from doxastic responsibility*:

1. If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience behaves in a doxastically irresponsible manner.
2. If an audience behaves in a doxastically irresponsible manner, then the audience's resulting belief cannot amount to knowledge.
3. (EP) If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge.

Put otherwise, if an agent, *A*, comes to believe that *p* from a speaker's, *S*'s, testimony that *p*, but behaves in a doxastically irresponsible manner in doing so, then *A*'s belief that *p* cannot amount to knowledge. This is because *doxastic irresponsibility is incompatible with knowledge*. One cannot be doxastically irresponsible in the course of coming to believe that *p* and acquire knowledge that *p*.

In defending my argument from doxastic responsibility, I focus on Aesop's fable of *The Boy Who Cried Wolf*:

A shepherd boy is tasked with guarding the sheep. He becomes bored and decides to play a trick on the local villagers. The boy shouts "Wolf!" very loudly. As a result, the villagers rush to the boy's aid, fearing that a wolf is in the flock. When they arrive, they see no sign of a wolf and so they leave. The next day, the boy plays the same trick. Again, the villagers rush to his aid, but see no sign of any wolf. The following day, a wolf comes along and attacks the sheep. The boy shouts "Wolf!", but no villagers arrive. They hear him, but they no longer believe him.

In this fable, the villagers believe the boy on his first two cries of "Wolf!". That is, the villagers believe there to be a wolf in the flock, given the boy's first two reports³. On the third occasion, however, they do not believe the boy. And it seems like they have good reason to adopt the doxastic attitude of disbelief on that third occasion as they had been tricked twice by the boy. In fact, if they continued accepting the boy's reports, one would regard them as gullible. After all, the moral of the story here is not to tell lies because if you do, people will no longer believe what you say.

But some clarifications must be made before proceeding. An impoverished villager cannot stand to lose a sheep to a wolf. So, one might expect the villagers to visit the flock despite the boy's trickery. However, it is not clear that in such cases the villagers should be regarded as believing the boy to be delivering veridical testimony. In such cases, the villagers might act *as if* the boy's report is true, but they need not believe his report. In this scenario, it is due to *practical considerations* that the villagers act as if the boy's report is true. A villager might think "that boy is probably playing his old tricks again, but I had better check on my sheep just to be safe". It is important then to make this distinction between cases. It is possible to think of cases in which the villagers visit the flock given the boy's third cry of wolf *because they believe him*, and cases in which the villagers visit the flock given the boy's third report *just to be safe and not because they believe him*. Perhaps we can imagine a third case, in which the villagers are unsure whether to accept the boy's report, and so they suspend belief, deciding to visit the flock just to be safe. My interest is not with the act of the villagers visiting the flock. Rather, my interest is in the doxastic states adopted by the villagers in response to the boy's testimony. That is, my interest is with the villagers *believing* or *disbelieving* the boy's testimony.

In Aesop's *Boy Who Cried Wolf*, the villagers do not believe the boy's third cry; they take him to be deceiving them again. And the intuition here is that the

³ I am here taking it that the boy's cry "Wolf!" can also be read as a report with propositional content along the lines of *there is a wolf in the flock*.

villagers are right to do so. It would be foolish to continue believing the boy after he has been caught lying on this very matter. But one might disagree here. Perhaps only two deceptions do not allow us to charge the villagers with gullibility. We can imagine a case wherein the boy has always been a trustworthy testifier and so, his conduct here presents a contrast to that of his normal behaviour. In response, I suggest that we amend the case, thereby putting concerns like this to the side.

Imagine a different version in which the villagers accept *49 deceptive cries of wolf*. The 50th cry of wolf comes from the boy seeing a wolf and sincerely trying to alert the villagers of the presence of a wolf in the flock. Let's say that the villagers accept this 50th cry. Call this version *Continued Acceptance*. Do the villagers in *Continued Acceptance* know that there is a wolf in the flock *on the basis of the boy's 50th report*? Of course, when they visit the flock and see the wolf, they know that there is a wolf in the flock. But my interest is with the question of whether they acquire knowledge that there is a wolf on the flock *from the boy's testimony*. I argue that they do not. And this is because, in accepting the boy's 50th report, the villagers act in a *doxastically irresponsible manner*. Given that doxastic irresponsibility is incompatible with knowledge, it follows that the villagers cannot know that there is a wolf in the flock on the basis of the boy's report.

Before discussing the concept of doxastic responsibility in detail, it will be beneficial to reflect more on the conduct of the villagers in *Continued Acceptance*. Intuitively, the villagers in *Continued Acceptance* are doing something other than they should. To continue accepting the boy's testimony after having been deceived so many times appears gullible or reckless. The intuition is that they should keep in mind the boy's awful track-record of reporting truths and reject his testimony. The issue is that, unlike the villagers in the original case, they don't do this. So, we can say that there is a sort of epistemic failing here. They *believe* other than they should insofar as they adopt the doxastic state of *belief* when they should adopt the doxastic state of *disbelief* towards the boy's testimony. One might add further that

they *inquire* other than they should insofar as they fail to respond to available considerations which support the claim that the boy is not reporting truthfully. So, to believe the boy's testimony is to do other than they should, and to neglect important data is to do other than they should as well. Let us explore this further.

In continuing to accept the boy's testimony, the villagers eventually come to have true beliefs *that there is a wolf in the flock*. This is because, even though they accept 49 false reports, they also accept the boy's 50th report *which is veridical*. Nevertheless, it seems mistaken to ascribe knowledge to the villagers here. Surely, they don't *know* that there is a wolf in the flock. One suggestion as to why that is the case is that the villagers are *inattentive* or *careless* or *gullible*⁴. Underlying this thought is the idea that, to attain knowledge that *p*, one must come to believe that *p* in the right way; through *epistemic virtues* (Zagzebski, 1996) (Sosa, 2007) (Sosa, 2015). In *Continued Acceptance*, however, the villagers come to have true beliefs through exhibiting defective qualities, or *epistemic vices* (Greco, 1993) (Baehr, 2011) (Battaly, 2014). It is through gullibility that they accept the boy's testimony, not through attentiveness.

Epistemic vices, writes Battaly, "are qualities that make us defective people" (2014: 52). She continues, "a person can be defective in a variety of ways: for instance, she can be defective insofar as she had bad vision; or insofar as she has bad logical skills; or insofar as she is dogmatic, unjust, or cruel" (ibid.). If you have bad vision, then this will negatively impact your pursuit of knowledge on the basis of your eyesight. A person with, say, 20/200 vision is going to have a very difficult time gaining knowledge on the basis of their vision (ibid.). The person with 20/20 vision, meanwhile, will have a much easier time. As such, the person with 20/20 vision does not suffer an epistemic vice in the way that the person with 20/200 vision does. But, as Battaly notes, one might be "defective" in a different way. One might

⁴ As we will see during my discussion of epistemic luck, one might also argue that the villagers are lucky to form true beliefs. In being lucky in this manner, so the argument goes, their beliefs cannot amount to knowledge.

be extraordinarily gullible, for example. Through exhibiting such gullibility, one might fail to consider important information which could be advanced in favour of the truth, or falsity, of one's belief. A gullible person, like one of villagers in *Continued Acceptance*, for example, will be negatively impacted by that gullibility insofar as it can dispose them to accept falsehoods.

Epistemic *virtues*, by contrast, are those qualities which aid, rather than hinder, our pursuit of acquiring justified belief and knowledge. If the person with 20/200 vision has an epistemic vice insofar as they have some quality which negatively affects them epistemically, the person with 20/20 vision has an epistemic virtue insofar as they have some quality which positively affects them epistemically. Likewise, the gullible person is negatively affected by that gullibility, whilst the attentive person is positively affected by that attentiveness. But what is it about vices that negatively affect us? And what is it about virtues that positively affect us? The answer one gives to this will depend on the kind of virtue epistemology one supports.

Virtue epistemologists roughly divide into two major camps: *virtue-reliabilism* and *virtue-responsibilism*. The two sides disagree most notably about the *nature* of virtues and vices. Common to both sides, however, is the idea that epistemic virtues are desirable whilst epistemic vices are not. Further, there is an agreement that one must believe as a result of epistemic virtue if one is to attain knowledge. That is, *the acquisition of knowledge requires belief through epistemic virtue*. Through appealing to virtue epistemology, we can make sense of the intuition that the villagers in *Continued Acceptance* are doing something other than they ought to do when they believe the boy's 50th cry, viz., they believe through an epistemic vice.

Let us look at both positions, starting with virtue-reliabilism. Battaly writes that, on this view:

"[V]irtues get their positive value from the intrinsically good ends they produce. Correspondingly, vices will get their negative value by producing intrinsically bad ends or effects. In short, vices will be qualities that reliably produce the bad" (ibid. 56-57).

Poor vision tends to produce bad effects in the sense that it will tend towards *false* belief. This is what explains why it is an epistemic vice. Conversely, good vision would count as a virtue on the virtue-reliabilist position, given its tendency to produce good effects in the form of *true* beliefs. Gullibility can be said to work similarly.

Our world is one in which people sometimes tell the truth, but sometimes fail to do so. Exhibiting gullibility in this world will tend to produce bad effects insofar as it can dispose one to accept falsehoods. This is not to say that gullibility is a vice *necessarily*. We can imagine possible worlds in which speakers report only truths. In such worlds, exhibiting gullibility will not dispose one to accept falsehoods. So, it is important to note that the tendency of a given virtue to produce good effects or bad effects is contingent on facts about that world. As such, when discussing virtues and vices in a virtue-reliabilist sense, the reader should note that the status of these virtues and vices are indexed to *this world*. So, in *this world*, gullibility is a vice in that it produces bad effects. Exhibiting attentiveness, by contrast, disposes one to form true beliefs, and is consequently, a virtue. For the virtue-reliabilist, virtues and vices are therefore *instrumentally* valuable in that they produce good or bad effects, rather than being good or bad in and of themselves (ibid.).

In the literature, it is common to see virtue-reliabilists talk of virtues in a way analogous to a "skill" or an "ability". For example, both Greco (1993) (2003) and Pritchard (2010) talk of "abilities" rather than "virtues". And Sosa opts for "skill" and "competence" (2007) (2015). Typically, we understand such terms as dispositions to bring about some state of affairs. In arguing for his virtue-reliabilist account, for instance, Sosa includes the example of the archer who manifests skill in reliably

hitting her target (ibid. 22-23). Greco includes the example of a baseball player manifesting skill in hitting baseballs (1993: 414). In these examples, the agents in question achieve their goals *through* their abilities which aim towards some end. For virtue-reliabilists, acquiring knowledge is realised in a similar respect.

Consider the following passage from Greco, concerning knowledge attribution:

“When we say that S knows *p*, we imply that it is not just an accident that S believes the truth with respect to *p*. On the contrary, we mean to say that S gets things right with respect to *p* because S has reasoned in an appropriate way, or perceived things accurately, or remembered things well, etc. We mean to say that getting it right can be put down to S’s own abilities, rather than to dumb luck, or blind chance, or something else” (Greco, 2003: 116).

Here, Greco implicitly states that, to know that *p* requires that we reach the belief that *p* in the right sort of way. Consider the archer once more. She might hit her target by accident. Perhaps a strong wind interferes with the trajectory of the departed arrow yet directs it toward the target. In such a case, if the arrow still hits the target, we cannot credit the archer’s success in hitting the target to her abilities. Instead, it is good fortune. The same can be said of true belief and knowledge (Sosa, 2007). Knowledge requires that we believe *because of* an epistemic virtue(s) whilst true belief does not.

But this is not the only way to think of epistemic virtues. The likes of Montmarquet (1992), Zagzebski (1996), Hookway (2003), and Baehr (2011), to name but a few, follow in Aristotle’s footsteps, maintaining that epistemic virtues are *character-traits* (Zagzebski, 1996: 79-85), (Hookway, 2003: 184). Examples include qualities such as inquisitiveness, attentiveness, honesty, and impartiality (Baehr, 2011: 21). Virtue-responsibilists deem such traits to be virtues not because of any tendency they might have to produce true beliefs, but because exhibiting them

makes a person excellent⁵ (Battaly, 2014: 54), or because they aid activities such as *inquiry* (Hookway, 2003: 188). Montmarquet, for example, suggests that traits such as impartiality, intellectual courage and conscientiousness are epistemic virtues because they demonstrate an underlying desire to reach the truth and avoid falsehood, irrespective of their ability to reliably lead to the truth (1992: 336). Baehr, on the other hand, writes that, whilst epistemic virtues can be 'grouped' into different categories, concerning initial motivation, integrity, consistency in evaluation, and so on, "[t]he basis of the groupings is the role that intellectual virtues play in the context of *inquiry*, by which I mean an active and intentional search for the truth about some question" (2011: 18).

At any rate, virtue-responsibilism presents us with a very different way of regarding epistemic virtue. It isn't just about getting to the truth; it's about getting to the truth *in the right way*. Consider an analogy with moral conduct here. Say that you are moving house and, without prompting, your friend promises to help you and makes good on her promise. How are we to assess her conduct? We might focus on her acts, claiming that the acts she performed are morally praiseworthy. But what if your friend only helped you because she was hoping to receive some gift in exchange? It seems that your friend's motivations for helping you affect how we assess her moral conduct. One might judge, as Aristotle does, that "actions are not enough" (*NE* IV.2: 1105a). Looking at your friend's acts alone misses out something important when it comes to the assessment of moral conduct, viz., it misses out the *character* of your friend (*ibid.*). To assess her conduct, we must ask: was she acting out of *generosity*? This suggests that there is a distinction between merely performing some act and performing that same act *in the right way*. Virtue-

⁵ There is disagreement, however, as to whether it is necessary that such virtues reliably get us to the truth or not (Battaly, 2008: 645). Zagzebski, for example, argues that epistemic virtues must reliably get us to believe truly (1996), whilst Montmarquet denies this (1992).

responsibilism can be seen as a way of capturing this intuition in our epistemic practices.

Given the virtue-responsibilist framework, we can see that determining whether an agent, *S*, knows that *p* requires an understanding of whether *S* reached her belief that *p* in an epistemically virtuous way. In the example above, assessing the moral conduct of your friend requires us to ask if she was acting out of generosity. Virtue-responsibilists treat cases of knowledge in a similar way. So, to assess the epistemic conduct of *S*, we must ask whether *S* believed that *p* out of epistemic virtue or epistemic vice.

Let us now return to the case of *Continued Acceptance*. By appealing to virtue-epistemology, we can better understand the claim that the villagers are doing something other than they should. The villagers are guilty of exhibiting epistemic vices. They neglect valuable data which is relevant to their inquiry and therefore fail to come to believe that there is a wolf in the flock *as a result of any epistemic virtue*. They come to believe truly, but this is not a creditable success. They are not attentive when they should be, and so, their epistemic conduct is other than it should be. We can regard them as *inattentive, gullible, and failing to be sensitive to available data*. Consequently, as they do not believe *out of* any epistemic virtue, they cannot be ascribed knowledge.

But one needn't buy into virtue epistemology to accommodate the intuition that the villagers in *Continued Acceptance* are getting things wrong in some sense. Instead, we can introduce some condition which rules out cases such as *Continued Acceptance* as cases of a subject knowing that *p*. I forward that McDowell's *Doxastic Responsibility* condition allows us to do exactly this without having to rely on contested terms, such as "virtue" or "vice".

McDowell suggests that considerations in favour of the truth, or falsity, of a proposition, *p*, have some "rational force" (1998: 429). For an agent to know that *p*, she must be responsive to this rational force. To neglect such considerations,

irrespective of whether one gets to the truth as a result, is to block the route to knowledge through being *doxastically irresponsible* (1998: 429). Otherwise put, one cannot know what one believes to be so if one fails to be responsive “to the rational force of independently available considerations” (ibid.). Consider a case in which an individual has good reason to distrust her vision, even though it is functioning exactly as it should. Perhaps she believes that she has ingested some hallucinogens but is mistaken. Imagine that, in spite of having good reason to distrust her vision, this individual ignores this consideration, believing that p on the basis of her vision. McDowell’s claim is that, in failing to respond to the considerations related to the truth of p , this individual’s “status as a knower is undermined” (ibid.).

Applied to the testimonial domain, we can hold that, for an audience, A , to know that p from a speaker’s, S ’s testimony that p , A must be sensitive to the available considerations which could be advanced in favour of the truth or falsity of p . As such, we see that “[a]cquiring knowledge by testimony is not a mindless reception of something that has nothing to do with rationality” (ibid. 434).

Accepting the boy’s 50th report is a clear example of doxastic irresponsibility. The villagers have available to them data which could be advanced against the truth of the boy’s report *but fail to respond to it*. This data takes the form of the track-record of the boy’s truthful reporting which supports the claim that the boy is deceiving them yet again. And whilst this is *not* true on the 50th occasion, the fact that the villagers fail to respond appropriately to that data entails that they are acting in a doxastically irresponsible manner. That is, they are neglecting something relevant that they should consider.

Consider another example. Imagine that I present you with a jar filled with 100 marbles. I tell you that it contains only blue and red marbles, but as the jar is opaque, you are unable to determine the colours without observing the marbles directly. I ask you to estimate the ratio of blue to red by removing one marble at a

time and placing it on the side. You do this ten times and then provide your estimation. Let's say that there are 80 blue marbles and 20 red marbles. Let's add that you remove five red marbles and five blue marbles. Given this, I expect you to estimate that the jar contains approximately 50% red marbles and 50% blue marbles. Let's say, however, that you tell me you think there are 80 blue marbles and 20 red marbles. In this case, you are correct, but the data available to you doesn't support your belief at all. You fail to be responsive to the rational force of the available considerations and so, you are *doxastically irresponsible* even though your belief is true. Being doxastically responsible is a matter of responding appropriately to the considerations that could be advanced in favour of the truth, or falsity, of some proposition, p . In this case, you fail to do that. Your belief is instead the result of good fortune.

For McDowell, doxastic responsibility is a necessary condition for knowledge. He writes that, "[i]f one's takings of things to be thus and so are to be cases of knowledge, they must be sensitive to the requirements of doxastic responsibility" (1998: 429). It follows then that knowledge excludes doxastic *irresponsibility*. This is because, amongst other things, "following the dictates of doxastic responsibility is obviously an exercise in rationality" (ibid.). Doxastic responsibility can therefore be seen then as a condition ruling out gullibility, carelessness, stubbornness, and other such vices. If one is gullible, careless, stubborn, etc., one thereby fails to be sensitive to considerations that could be advanced in favour of the truth, or falsity, of p . The gullible person accepts testimony without due diligence. The careless person is unmotivated to attend to considerations which have a bearing on the truth of p . And the stubborn person refuses to be moved, even by considerations which have a bearing on the truth of p . These qualities are not compatible with doxastically responsible conduct. The result is that, through being doxastically irresponsible in the course of coming to

believe that p , an agent thereby blocks their route to knowing that p . But note that we can go on to make an even stronger claim here.

Consider again the agent who is doxastically irresponsible by way of their gullibility. This agent accepts everything they are told without any sensitivity to considerations that could be advanced in favour of the truth, or falsity, of a speaker's testimony. Contrast that agent with the doxastically responsible agent who demonstrates a sensitivity to relevant considerations whenever they engage with a speaker's testimony. Suppose that both the gullible and the attentive agent are told that p and accept it. However, the gullible agent accepts that p because of their gullibility while the attentive agent accepts that p because of good reasons. As it turns out, p is false. Now, despite both agents failing to know that p , given that p is false, it seems like the two agents' epistemic positions are not identical. It seems that the attentive agent attains some kind of positive epistemic standing. We can understand this by suggesting that the attentive agent can attain a *justified* belief while the gullible agent cannot. But since the only significant difference between the two agents is that one is doxastically responsible and the other is not, we might explain the apparent difference in justificatory status by appealing to the difference in doxastic responsibility.

The suggestion here is that the doxastically responsible belief is justified belief in virtue of being doxastically responsible. Thus, doxastic responsibility is *necessary* but also *sufficient* for justified belief. Such a suggestion is explanatorily powerful insofar as it allows us to account for why the attentive agent can attain a positive epistemic standing while the gullible agent cannot. Moreover, it allows us to explain why doxastic responsibility is necessary for both knowledge and justified belief. Importantly, understanding justified belief in this way does not restrict one to holding a particular position. As we shall see later, the reliabilist and the evidentialist can both accommodate the idea that justified belief requires agents to demonstrate

a sensitivity to considerations which could be advanced in favour of the truth or falsity of a piece of testimony.

And there is a further dimension to doxastic responsibility which becomes clear if we amend the case of *Continued Acceptance*. Let us imagine that a new villager comes to town. This person has no information about the boy, or the villagers, and she has not heard any of the boy's cries. The new villager hears the boy's 50th cry of "Wolf!". She believes that there is a wolf in the flock and, upon arrival, sees for herself the wolf in the flock. Does this villager *know* that there is a wolf in the flock *on the basis of* the boy's testimony? I see no reason why we must conclude that she does not. After all, this villager has no reason to believe anything is awry. And, provided she is disposed such that, *were* she in possession of reasons against the truth of the boy's report, she *would* reject it, we can regard her as doxastically responsible.

Let us examine more closely this dispositional element of doxastic responsibility by considering the two possibilities available here:

- a) The new villager arrives and hears the boy's 50th cry of "Wolf!". She has no reason to suspect that he is a liar. However, even if she did have reason to suspect this, she would still accept the boy's report.
- b) The new villager arrives and hears the boy's 50th cry of "Wolf!". She has no reason to suspect that he is a liar. If she did have reason to suspect this, she would carefully consider whether he was being deceitful or not.

Perhaps one might suggest that, in (a) the new villager is not doxastically irresponsible since the considerations available to the villagers are not currently available to her. But is there not an issue with her conduct insofar as she is disposed such that, even if she was aware of the same considerations, she would neglect them, and therefore be doxastically irresponsible? If doxastic responsibility is an "exercise in rationality" as McDowell states (*ibid.*), then this agent's failure to be appropriately disposed should push us to conclude that her epistemic character is

other than it should be, since her true belief cannot be credited to her attentiveness. She *should* be disposed such that, if she was aware of such considerations, she would reflect on them seriously. As she is not, she can be regarded as *doxastically irresponsible* insofar as she is not disposed to behave in a doxastically responsible manner. What this illustrates is that being doxastically responsible is not simply a matter of being sensitive to the rational force of considerations concerning some proposition, *p*, at some time, *t*. Rather, it is about responding to relevant considerations *and being disposed to do so*. This suggests then that doxastic responsibility has a modal element to it.

Appealing once more to virtue-epistemology, we can see a similarity between doxastic responsibility and the epistemic virtues of *attentiveness* or *sensitivity to detail*. The issue with failing to be doxastically responsible is that, in doing so, one thereby believes out of epistemic vice.

By demanding that agents be doxastically responsible to attain knowledge, we capture the intuition that we must engage with the target proposition, to some extent. What this means is that, if we are concerned with the truth of the claim *that the shop closes at 5pm*, we must also be concerned with the considerations that could be advanced in favour of the truth or falsity of that claim. Further, we must be disposed such that, if we were to receive evidence which has a bearing on the truth of that claim, we would respond to it accordingly.

This serves as my defence of the second premise of my argument from doxastic responsibility: *if an audience behaves in a doxastically irresponsible manner, then the audience's resulting belief cannot amount to knowledge*. As such, I will now turn my attention to premise one: *if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience behaves in a doxastically irresponsible manner*.

IV

So far, I have not provided a clear definition of *truth-directed testimonial discriminations*. I seek to rectify this now. I understand truth-directed testimonial discriminations much in the same way McDowell understands doxastic responsibility (ibid.). As we have seen, for McDowell, considerations in favour of p or $\sim p$ have some *rational force*. To be doxastically responsible is to demonstrate a sensitivity to that rational force (ibid.). It is also, as we have seen, to be disposed such that one *would* respond to relevant considerations. Applied to the matter at hand, we can say that to make truth-directed testimonial discriminations is to discriminate between bits of testimony in such a way as to be *sensitive to considerations which could be advanced in favour of the truth, or falsity, of a piece of testimony* (ibid.) (Faulkner, 2011: 100). In discriminating between bits of testimony in such a way, one's testimonial discriminations are truth-directed insofar as they are a matter of responding to truth-related considerations.

But discriminating between bits of testimony in a truth-directed manner doesn't entail always getting things right. It might be that the available considerations supporting the truth of a piece of testimony outweigh the considerations in favour of its falsity, even though it is false. No doubt, this is apparent in many cases of successful deception. You might attempt to deceive me by reporting that p when p is false, but when the considerations available to me make it seem as though p is true. In such a case, I can accept that p whilst being doxastically responsible *but form a false belief*. In such a case, I accept a piece of testimony on the basis of a truth-directed testimonial discrimination *despite the fact that I get things wrong*. To make truth-directed testimonial discriminations then is not to accept *only* veridical testimony. This is a requirement no human could be expected to meet.

But note that we can understand *truth-directedness* in more than one way. To illustrate this, imagine the case in which you are trying to get to the cathedral in a city with which you are unfamiliar. Without any kind of map, you search for street

signs pointing to the cathedral. At last, you find a sign and follow it. You find another sign, and you follow that one too. Is your travel *cathedral-directed*? The answer, it would seem, is yes. But what if all those signs had been put there by a malicious city-dweller to confuse tourists? What if those signs lead away from the cathedral? In this case, there is one sense in which your travel is not cathedral-directed but another in which it is. It is *not* cathedral-directed in the sense that, as a matter of fact, you are not travelling towards the cathedral. But we might say that your travel is cathedral-directed in the sense that, not only are you doing your best to get to the cathedral, but any reasonable person would do the same as you. By this, I mean that you are being sensitive to considerations which purport to indicate the location of the cathedral and that this is what we would expect of any reasonable person with the same aim as you.

If truth-directed testimonial discriminations are understood by appealing to McDowell's concept of doxastic responsibility, then this suggests an understanding of truth-directedness in the second sense. It suggests that what it is to make truth-directed testimonial discriminations is to do what any reasonable person would do to get to the truth, irrespective of whether, as a matter of fact, one is travelling towards it. It suggests that to make truth-directed testimonial discriminations is to have one's testimonial discriminations dictated by one's aim of getting to the truth (and avoiding falsity). This is because, if being doxastically responsible is a matter of being appropriately responsive to the rational force of available considerations that could be advanced in favour of p or $\sim p$, then making truth-directed testimonial discriminations should be seen as a matter of being appropriately responsive to considerations that could be advanced in favour of the truth or falsity of a piece of testimony. As the cathedral example shows, however, one might be appropriately

responsive to “misleading signs”. This is unfortunate but does not alter the fact that one is doing one’s best to get to the truth⁶.

Given that I make sense of truth-directed testimonial discriminations by appealing to doxastic responsibility, one might claim that premise one - *if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience behaves in a doxastically irresponsible manner* - amounts to a tautology. If truth-directed testimonial discriminations and doxastic responsibility are both understood as a matter of responding to considerations which could be advanced in favour of the truth or falsity of a piece of testimony, it follows that a failure to make truth-directed testimonial discriminations is thereby a failure to be doxastically responsible as they are both realised in the same fashion. This is not to make a claim of synonymy, however. Doxastic responsibility is a broader term whilst truth-directed testimonial discriminations is a narrow term relating to the testimonial domain only. One can be doxastically irresponsible in the perceptual domain. As such, while failing to make truth-directed testimonial discriminations entails doxastic irresponsibility, doxastic irresponsibility does not necessarily entail a failure to make truth-directed testimonial discriminations (though it can). Truth-directed testimonial discriminations can therefore be seen as an *application* of doxastic responsibility to the testimonial domain.

If truth-directed testimonial discriminations are indeed an application of doxastic responsibility to the testimonial domain, then we should see an alignment between cases of testimonial acceptance on the basis of discriminations which are not truth-directed and cases of doxastic irresponsibility. The agent who fails to make truth-directed testimonial discriminations should *also* be the agent who is doxastically irresponsible.

⁶ In this way, making truth-directed testimonial discriminations might be understood in a way similar to how Montmarquet regards epistemic virtues. For him, epistemic virtues are such because they indicate an underlying desire for the truth (1992: 336). Following this, we might say that making truth-directed testimonial discriminations indicates an underlying desire for the truth.

To be clear, the first premise of my argument from doxastic responsibility is that if one fails to make truth-directed testimonial discriminations, then one is doxastically irresponsible. I do not claim that if one is doxastically irresponsible then one thereby fails to make truth-directed testimonial discriminations. The truth of the first premise follows from what making truth-directed testimonial discriminations entails, viz., doxastically responsible conduct. To make the claim that doxastically irresponsible conduct entails a failure to make truth-directed testimonial discriminations would be mistaken.

To combat any concerns as to the truth of this premise, however, let us continue the defence. To show that premise one is false requires the truth of the antecedent - *that if an agent's testimonial acceptance is based on a discrimination which is not truth-directed* - and the falsity of the consequent - *that the subject does not behave in a doxastically irresponsible manner*. Let us suppose then *that if an agent's testimonial acceptance is based on a discrimination which is not truth-directed, the subject does not behave in a doxastically irresponsible manner*. My aim is to demonstrate that this leads to a contradiction.

Insofar as making truth-directed testimonial discriminations involves doxastic responsibility, it is the case that *if A's acceptance of S's testimony that p is on the basis of a discrimination which is truth directed, then A behaves doxastically responsible with respect to p*. This is true simply in virtue of how truth-directed testimonial discriminations are defined. This means that where we have the antecedent, we will also have the consequent. But we need to establish the truth of the antecedent *and* the falsity of the consequent of premise one. The problem is that the manner in which truth-directed testimonial discriminations have been defined prevents that. It is not possible to establish the truth of the antecedent and the falsity of the consequent. To allow that is to allow contradiction. As such, premise one of my argument from doxastic responsibility must be true.

But we can better make sense of this with an example. Call the following case, *Flattering Testimony*:

Jeff is a petty boss who looks fondly only on those who flatter him. As a result, the best way to succeed in the company is to flatter him. Jeff also has a particularly strange approach to responding to testimony: he accepts what his employees say *only when it flatters him*, otherwise he rejects it. Jeff is told by an employee that he is great at tennis. As a result, Jeff forms the belief that he is great at tennis.

In *Flattering Testimony*, we have an agent who makes testimonial discriminations which are not truth-directed. Rather than accepting and rejecting on the basis of discriminations which are directed towards the truth, Jeff accepts and rejects on the basis of discriminations which are directed towards *flattery*. Jeff receives his employee's testimony and accepts it *because it flatters him*. So, we have the truth of the antecedent of premise one: *that a given piece of testimonial acceptance is based on a discrimination which is not truth-directed*. The question now is whether Jeff behaves in a doxastically irresponsible manner. If he does not, then the consequent is false and premise one of my argument from doxastic responsibility fails.

For the consequent - *that the subject behaves in a doxastically irresponsible manner* - to be false, it must be the case that, in spite of his flattery-directed testimonial discriminations, Jeff is not doxastically irresponsible when he accepts his employee's testimony that he is great at tennis. But to determine this, we must determine whether Jeff is sensitive to the considerations that could be advanced in favour of the truth or falsity of the proposition *that Jeff is great at tennis*. And it seems that Jeff fails to demonstrate such a sensitivity. If Jeff *had* been sensitive to those considerations, then his acceptance would have been on the basis of a discrimination which was truth-directed, rather than flattery-directed. But this would

entail the falsity of the antecedent⁷. If, on the other hand, he had *not* been sensitive, then whilst his acceptance would have been on the basis of a flattery-directed discrimination, his failure to demonstrate a sensitivity to considerations concerning the truth of his employee's testimony would entail his doxastic irresponsibility. And this is because, what it is to neglect, or be *disposed to neglect*, considerations which could be advanced in favour of the truth, or falsity, of a piece of testimony just is to be doxastically irresponsible. Neither route ends in success.

And here I will reiterate that an ascription of doxastic irresponsibility is not affected by one's believing truly or not. The villagers in *Continued Acceptance* demonstrate this. So, even supposing it is true that Jeff is great at tennis, he is doxastically irresponsible insofar as he discriminates between bits of testimony in the way that he does. Jeff's epistemic conduct is much like that of the villagers on the 50th cry. Whilst they may all enjoy true beliefs, it is the result of doxastic irresponsibility, and thus, prevents the acquisition of knowledge.

Perhaps there is a way to strengthen the case. Suppose that all hiring decisions in Jeff's company are made by Elene. Elene is very careful when hiring people; she only hires those she deems trustworthy. Furthermore, she lets employees know that they should only offer testimony they *know* to be true, or else their contracts will be terminated. This means that, when Jeff receives flattering testimony from his employee, there is a high probability that the employee is reporting something they know to be true. The question we must now ask is whether this revised case has any impact on our ascription of doxastic irresponsibility for Jeff.

I do not see how this impacts an ascription of doxastic irresponsibility for Jeff. Jeff's employee may well be saying something they know to be true, but doxastic irresponsibility is something which concerns the agent *forming* the belief. In the testimonial domain, it concerns the *audience* and not the speaker, therefore.

⁷ And if the antecedent is false, then premise one cannot be falsified. This is true given the truth-conditions for the conditional $p \rightarrow q$.

Irrespective of whether the speaker offers testimony they know to be true, Jeff still discriminates between bits of testimony on the basis of whether it is flattering to him. And insofar as the discriminations he makes are flattery-directed and not truth-directed, he fails to be sensitive to considerations which could be advanced in favour of the truth or falsity of the piece of testimony offered.

To allow that Jeff is *not* doxastically irresponsible is to allow contradiction. Allowing that one can *fail* to make truth-directed testimonial discriminations whilst being doxastically *responsible* requires that one can both *neglect* considerations which could be advanced in favour of the truth or falsity of a piece of testimony and simultaneously be sensitive to such considerations. Clearly, this is not possible.

With this, we can now reach the conclusion of my argument from doxastic responsibility:

(EP) If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge.

When we add to this the claim (EF): *that we acquire knowledge from testimony (by and large)*, we can reach the conclusion:

Epistemic Consequent (EC): Our testimonial discriminations must be truth-directed (by and large).

This is because, if it is the case that we acquire knowledge from testimony (by and large), we must satisfy the necessary conditions of knowledge acquisition (by and large). Since it is the case that we acquire knowledge from testimony by and large, it follows that *our testimonial discriminations must be truth-directed (by and large)*.

My *argument from doxastic responsibility* presents one route to supporting the *argument for truth-directed testimonial discriminations*. Through appealing to the ideas underlying that argument, I have made sense of truth-directed testimonial discriminations as a matter of doing as any reasonable person with the same aims would do. In the testimonial domain, we can understand those aims as exhausted

by the goal of reaching the truth. I now aim to provide a further defence by way of my *argument from epistemic luck*.

V

I will now provide a further defence of (EP), by way of my *argument from epistemic luck*:

1. If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then it would be a matter of veritic epistemic luck that the audience attained a true belief.
2. If the audience attains a true belief through veritic epistemic luck, that belief cannot amount to knowledge.
3. (EP) If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge.

Put otherwise, if we accept a piece of testimony on the basis of a discrimination which is not truth-directed, then any true belief we attain will be a consequence of veritic epistemic luck. As veritic epistemic luck is incompatible with knowledge, if an agent accepts a piece of testimony on the basis of a discrimination which is not truth-directed, that agent's resulting belief cannot amount to knowledge.

I will begin my defence the argument from epistemic luck by arguing in support of the second premise: *if the audience attains a true belief through veritic epistemic luck, that belief cannot amount to knowledge*. In doing so, I include the following Gettier case from Pritchard (2014: 153-154) (henceforth, *Sniper*):

Erika is in her apartment. She gazes out of the window into a nearby office block and sees, what appears to be, a sniper with her rifle pointing towards her. As a result, Erika forms the belief "there is a sniper in that building". However, what Erika sees is not a sniper. Rather, Erika is looking at a cardboard cut-out of a sniper which has been used for the

sake of advertising a new film. But behind that cardboard cut-out is a real sniper with her rifle pointing towards Erika.

Can we ascribe to Erika knowledge *that there is a sniper in that building*? Her belief is true; there really is a sniper in that building. Moreover, it seems that this belief is justified in that she has good reason to form the belief “that there is a sniper in that building”⁸. Nevertheless, it seems a mistake to regard Erika as *knowing* that there is a sniper in the nearby office block.

Sniper is a Gettier case (Gettier, 1963). It presents us with a case of justified true belief which does *not* amount to knowledge. Gettier cases act as counterexamples against a traditional, but long since abandoned, view that knowledge is justified true belief and nothing more. This view must be mistaken, Gettier argues, since these cases illustrate that an agent can have a justified true belief *which falls short of knowledge* (ibid. 123). Post-Gettier, figures have either sought to analyse the concept of knowledge by adding and amending conditions or by simply abandoning analysis and taking knowledge to be basic⁹. To do this first requires a diagnosis of the problem in cases like *Sniper*. We must ask *why* does Erika’s belief *that there is a sniper in that building* not amount to knowledge? One suggestion is that it is a matter of *luck* that her belief is true (Unger, 1968), (Lehrer and Paxson Jr., 1969), (Zagzebski, 1994), (Harper, 1996) (Engel Jr., 2000) (Pritchard, 2005), (2007), (Kelp, 2013). And given, as Pritchard puts it, the “seemingly universal intuition that *knowledge excludes luck*” (2005: 1), Erika should not be regarded as having knowledge here¹⁰.

⁸ We can add that the cardboard cut-out is incredibly realistic.

⁹ Goldman has defended a causal theory of knowing (1967). Later he argues for a “reliability” condition (1979). Armstrong includes a “no false lemmas” condition (1973). Nozick suggests that knowledge requires *sensitive* belief (1981). Sosa argues that knowledge requires *safe* belief (1999). Williamson takes it that knowledge is “fundamental” (2000: 185).

¹⁰ I will note that not everybody shares the intuition that knowledge and the kind of luck featured in Gettier cases are incompatible. See Turri (2012) and Hetherington (2019).

But why hold that knowledge and luck are incompatible? If they are indeed incompatible, then we shouldn't be able to construct any case wherein an agent comes to acquire knowledge whilst being lucky. But consider Unger's example of an employee overhearing his employer say that he will be fired tomorrow (henceforth, *Overheard Report*) (1968: 159). Imagine that the employee is walking past his employer's office at just the right moment to overhear his employer's testimony. If he had walked past a little bit earlier, or later, he wouldn't have heard anything. Is it not lucky that the employee overheard his employer's testimony? Unger's position is that, despite it being a matter of luck that the employee walks past his employer's office at just the right time to hear the report, "he may know whether by accident or not" (ibid.).

Overheard Report presents an uncontroversial case of an agent acquiring knowledge. But if knowledge and luck are compatible, we must explain the intuition that it is because of the luck present in *Sniper* that Erika doesn't acquire knowledge. Why does the presence of luck prohibit knowledge acquisition in one case but not the other? At this point, we must look more closely at the sort of luck featured in both cases.

In *Overheard Report*, it is lucky that the employee heard the employer's report. Had he not heard it, he would not have formed the belief that he will be fired tomorrow. And if he had not formed that belief, then he wouldn't have come to *know* that he will be fired tomorrow. In other words, the employee believes that p because he heard S 's report that p , and it is *lucky that he heard S 's report that p* . We can imagine a nearby possible world in which the employee walks past his employer's office a little later. In that world, the employee simply doesn't form the belief that he will be fired tomorrow. Contrast this with *Sniper*, in which Erika forms the belief that there is a sniper in a nearby building. In this case, it is lucky *that Erika's belief is true*. She isn't lucky to form the belief in the same way that the employee is lucky to form his belief. And this is because, it is a matter of luck, not that she *forms*

her belief, but that her belief is *true*. If the sniper had gone to a different building, she would have still formed the belief, but it would have been *false*.

The forms of luck here are quite different. So, perhaps we should retreat to the weaker claim that it is only *some* rather than *all* forms of luck which are incompatible with knowledge. This would allow us to view *Overheard Report* as a case of an agent knowing that *p* whilst rejecting *Sniper* as such a case. If this is right, then there are some forms of luck which are *knowledge-permitting* in that they are *compatible* with knowledge, and some which are *knowledge-impeding* in that they are *incompatible* with knowledge. My task now is to make clear the sort of luck that is incompatible with knowledge and to understand why that is so. Prior to this, however, a discussion of 'luck' is required. Though we have a basic grasp of what luck is and what it is to be lucky, a closer look is necessary.

'Lucky' has been used interchangeably with 'accidental' (Unger, 1968: 158), 'chance' (Harper, 1996: 273), and 'out of an agent's control' (Nagel, 1979: 25). As Pritchard notes, however, this is problematic (2005: 126-128). You are *lucky* when you win the lottery, but it isn't *accidental* that you won. It might be a *chance* event that there are four people called 'Elena' in the classroom, but it isn't *lucky* that this is so. And the orbit of Jupiter is *beyond our control*, but it is not a matter of *luck* that it orbits the Sun in the manner that it does. Though there is certainly some overlapping with these terms, 'luck' diverges in some way.

What we require then, is an account of luck. Coffman notes that there are three leading accounts: *the modal account; the control account; the mixed account* (2015: 3). My goal here is not to defend a particular account of luck, but to elucidate the term under examination. With the intention of doing this, let us consider Pritchard's *modal account* of luck. Pritchard claims that an event is lucky when the following two conditions are satisfied:

"(L1) If an event is lucky, then it is an event that occurs in the actual world but which does not occur in a wide class of the nearest possible worlds

where the relevant initial conditions for that event are the same as in the actual world" (2005: 128).

"(L2) If an event is lucky, then it is an event that is significant to the agent concerned (or would be significant, were the agent to be availed of the relevant facts)" (ibid. 132).

We can better understand Pritchard's model by using an example.

Suppose you win the lottery. This is the classic example of being lucky. On Pritchard's account, it must be the case that the event occurring in the actual world, viz., your lottery win, does *not* occur in nearby possible worlds, where the "relevant initial conditions" are the same as those of the actual world. Let us unpack this some more. For Pritchard, possible worlds are ordered by their likeness to the actual world, i.e., *this world*. A possible world which is very similar to the actual world is a "nearby possible world" and a possible world which is very dissimilar to the actual world is a "distant possible world". The "relevant initial conditions" are conditions which relate to the event, but which do not entail it¹¹. In this case, the relevant initial conditions are the purchase of the ticket, the fairness of the lottery, and so on (ibid. 128). So, the nearest possible worlds are those worlds in which all is the same as in the actual world, but for the winning numbers, and thus, your lottery win. As Pritchard states, "[a]lthough the agent has a winning ticket in the actual world, in most (if not all) of the nearest possible worlds at issue she will be in possession of a losing ticket" (ibid. 128). As such, L1 is satisfied.

¹¹ Pritchard notes that we could get rid of the "relevant initial conditions" clause due to it being problematic. The issue is in determining just what counts as a relevant initial condition. If, for example, one includes the way the lottery balls fall in the set of relevant initial conditions, then this will bring about the same winning numbers across all possible worlds. What is needed then is a definition which does not determine events in such a way nor ensures "in advance that the event in question is 'chancy' in just the respect that is needed if it is to be regarded as a lucky event" (ibid. 132). Following this, Pritchard notes that since the way in which possible worlds are ordered in such that those nearby possible worlds are ones in which conditions related to the event in question remain the same (e.g., the fairness of the lottery, the purchase of the ticket), there is no need to include the "relevant initial conditions" clause; ordering by likeness to the actual world does the job just fine. That said, Pritchard does not dispense the "relevant initial conditions" clause as he takes it to be explanatorily valuable (ibid.).

Understanding L2 better requires that we pry into Pritchard's use of the word 'significant' here. I will avoid this, as an intuitive grasp of it will serve our purposes¹². Intuitively, a lottery win would be regarded as 'significant' to the agent involved. Compare this to the following example: I have, presented before me, a bag with 100 marbles inside it. 90 are white and 10 are red. I am asked to select a marble at random without looking. I remove a red marble; an unlikely outcome given the ratio of white to red. In most nearby possible worlds, I remove a white marble. So, L1 is satisfied. But it is not clear how this event could be seen as significant to me. So, L2 is not satisfied. Regarding me as lucky in this case is mistaken. If, however, I receive an all-expenses paid trip to the Bahamas as a result of selecting the red marble, regarding me as lucky seems more appropriate. A lottery win is understandably significant to the agent. As such, a lottery win satisfies both L1 and L2 and is therefore lucky on Pritchard's model.

We can now return to the discussion concerning forms of luck. Recall, I suggested that the form of luck in *Overheard Report* is knowledge-permitting, whereas the form of luck in *Sniper* is knowledge-impeding. Why is this so? Utilising Pritchard's modal account of luck, we can say that, in nearby possible worlds, the employee in *Overheard Report* walks past his employer's office at a different time and, consequently, doesn't hear his employer say that he will be fired tomorrow. As a result, in nearby possible worlds, the employee does not form the belief that *p* and so, does not have knowledge. This is different to *Sniper*. In this case, Erika is not lucky to receive evidence in support of a belief and thus lucky to *form* a belief, but rather lucky to have a belief which turns out to be true. Consider, in nearby possible worlds, the sniper arrives at the building at a different time or chooses a different area from which to snipe. However, in those cases, she forms a *false* belief. So,

¹² Of course, what is 'significant to the agent' will depend on many factors. And what is significant to you at some time, *t*₁, might not be significant at *t*₂. It is not within the scope of this project to delve into this matter further, though I do appreciate the ambiguity here.

Unger's employee is lucky to get evidence whereas Erika is lucky to have a true belief.

The difference between these two forms of luck is recognized by Engel who distinguishes between *evidential epistemic luck* (EE luck) and *veritic epistemic luck* (VE luck) (1992: 67). Attempts at making such a distinction can be seen earlier in the work of Unger (1968) and Nozick (1981). Pritchard follows this distinction (2005). He writes that an agent is EE lucky when "[i]t is lucky that the agent acquires the evidence that she has in favour of her belief" (ibid. 136). In *Overheard Report*, the employee is EE lucky because he is lucky to receive evidence which supports his belief that he will be fired. Erika is not lucky in this sense.

An agent is said to be VE lucky when "[i]t is a matter of luck that the agent's belief is true" (ibid. 146). This requires further clarification, however. Appealing again to the modal account of Pritchard, we can say that an agent is VE lucky when:

"[T]he agent's belief is true in the actual world, but [...] in a wide class of nearby possible worlds in which the relevant initial conditions are the same as in the actual world - and this will mean, in the basic case, that the agent at the very least forms the same belief in the same way as in the actual world [...] - the belief is false" (ibid.)

This description seems to capture the sort of luck featured in *Sniper*. And this is because Erika is lucky to believe truly because it just so happens to be the case that, when she saw the cardboard cut-out of the sniper, there was a sniper in that building. Consequently, her belief turns out to be true. Things *could* have gone differently, of course. And in nearby possible worlds, they do. In nearby possible worlds, the sniper arrives later in the day, or sets up in a different building altogether.

Consider a nearby possible world in which the sniper arrives at the building later in the day. Say that all else remains. At some time, t_1 , Erika forms the belief *that there is a sniper in that building* from seeing the cardboard cut-out of the sniper. But

let's say that the sniper arrives at the building in question later, at t_2 . This results in Erika forming a belief which is false despite her believing in the *same manner* as in the actual world, in which she believes truly. So, Erika is lucky insofar as she inhabits the world in which the sniper is already behind the cardboard cut-out at t_1 ¹³. In nearby possible worlds, she forms a belief in the same way, on the same grounds, but believes *falsely*. It is *this* type of luck, VE luck, which is knowledge-impeding in that it blocks Erika's route to knowledge.

Further evidence for the incompatibility of knowledge and luck comes from Zagzebski, who argues that we can construct Gettier cases simply by adding some VE luck (1994: 69). If this is so, it reinforces the claim that knowledge excludes VE luck, for we are here modifying only one factor: VE luck. If a change in knowledge-ascription correlates with a change in the presence of VE luck, this speaks in favour of the thesis that VE luck and knowledge are incompatible.

Zagzebski's procedure runs like so: start with a case of justified false belief, where the justification is strong enough to allow for knowledge. Importantly, the falsity of the belief should be a matter of luck. The next step is to add a further element of luck which results in the belief being true. After that, you have your Gettier case (ibid.).

Consider Chisolm's "sheep on the hill" example (1989: 93). Stanley is out for a walk in the countryside. He spots a sheep in a field and forms the belief "there's a sheep in the field". He then spots another sheep, in a barn, and forms the belief "there's a sheep in that barn". Stanley's eyesight is functioning exactly as it should. But now imagine that, as he is walking, he spots what appears to be a sheep on a hill. He forms the belief "there is a sheep on that hill". Surely, we are to say that Stanley has a justified belief, given how things appear. What we need, however, is for Stanley's justified belief to be false through bad luck. So, let's say that what he

¹³ Though her luck is confined to her having a true belief. Being the target of a sniper is not a paradigmatic case of being lucky.

sees is not a sheep, but a dog in a very convincing sheep costume. We now have a case in which an agent has a justified false belief which is false by bad luck. The next step is to add a further element of luck which sees Stanley's belief come out as true. We can achieve this by stipulating that, on the other side of the hill, and consequently not visible to Stanley, there is a sheep grazing. So, when he formed the belief "there is a sheep on that hill", he formed a true belief. He therefore has a justified true belief, but which falls short of knowledge.

In this example, Stanley fails to know that there is a sheep on the hill for the same reason that Erika fails to know that there is a sniper in the nearby building, viz., because it is a matter of luck that he believes *truly*. In nearby possible worlds, Stanley continues to form the belief *that there is a sheep on the hill* on the basis of seeing the dog but believes *falsely*. In such worlds, Stanley walks past at a different time of the day, spotting the dog and forming the belief *that there is a sheep on the hill*, when no sheep is present. In such worlds, Stanley forms a belief in the same way but the belief he forms is *false*. What we see then is that, to make a Gettier case, the vital ingredient is VE luck. Through using Zagzebski's method, which focuses on the inclusion of VE luck, we are able to produce cases where agents have justified true beliefs which fall short of knowledge. This reinforces the claim that knowledge is incompatible with VE luck since it is the addition of such luck in these cases which renders them *Gettier cases*.

Now, in maintaining that knowledge is incompatible with VE luck, I am maintaining that there is an *anti-luck condition* on knowledge. Various anti-luck conditions have been proposed in an attempt to provide a theory of knowledge which is not undermined by Gettier cases. And while my task in this section is simply to defend premise two of my argument from epistemic luck (*that if the audience attains a true belief through veritic epistemic luck, that belief cannot amount to knowledge*), it will be helpful to outline some proposals.

Goldman has, at one point, defended a causal theory of knowledge in which it is required that the fact that *p* be “causally connected in an ‘appropriate’ way with *S*’s believing *p*” (1967: 369). Armstrong has forwarded a “no false lemmas” condition, in which “a condition is introduced that the justifying beliefs are known to be true” (1973: 153). Nozick has argued that knowledge must *track the truth* (1981: 179). He writes that a person “knows that *p* when he not only does truly believe it, but would truly believe it and wouldn’t falsely believe it” (ibid. 178). And Sosa has argued that knowledge entails *safe belief*, meaning that one’s belief “has a basis that it would (likely) have only if true” (2007: 26).

Each of these conditions help to explain why Erika fails to have knowledge in *Sniper*, but in different ways. An advocate of Goldman’s causal account, for example, would say that Erika fails to have knowledge since the fact that there is a sniper in the building is not what *causes* Erika to believe that there is a sniper in that building (1967). An advocate of Armstrong’s position, on the other hand, would say that Erika doesn’t have knowledge in *Sniper* because her belief comes from a falsehood, i.e., that she is looking at a real sniper (1973). And an advocate of Nozick’s theory would hold that Erika fails to know that there is a sniper in the nearby building because her belief is not *sensitive*. Erika would believe that there is a sniper in that building even if there was no sniper there, since her belief is formed on the basis of seeing the cardboard cut-out. As such, it is not the case that if there *wasn’t* a sniper in that building, she *wouldn’t* believe it (1981). Her belief does not “track the truth”.

Though these accounts can help to explain why Gettier cases are cases in which agents fail to have knowledge, I will not discuss them in detail. This is because, rather than seeking to defend a given account, my focus is with the claim that knowledge cannot be attained by VE luck. Importantly, these accounts agree that Gettier cases are *not* cases of knowledge. So, though there may be disagreement regarding the nature of the anti-luck condition, these accounts do not go against

the claim that knowledge and VE luck are incompatible. What will be useful, however, is an application of one of these proposed conditions. In doing so, we can better understand the epistemic failings present in Gettier cases. I will proceed by appealing to the *safety condition* (Sosa, 1999) (2007), (Williamson, 2000) (Pritchard, 2005) (2014).

As we have seen, Pritchard makes sense of luck by appealing to *possible worlds*. He understands a lucky event, such as a lottery win, as an event occurring in the actual world but *not* occurring in a wide class of nearby possible worlds (2005: 128). He makes sense of the safety condition by taking a similar modal approach. He writes that one's belief that *p* is *safe* when "one has a true belief that *p* such that, in close possible worlds, if one continues to form a belief on the same basis as in the actual world, then one's belief continues to be true" (2014: 156). By contrast, one's belief is *unsafe* when one has a true belief that *p* in the actual world but, in nearby possible worlds, beliefs formed on the same basis end up false. Since Gettier cases are cases of agents experiencing VE luck, they can be regarded as cases of events occurring in the actual world, viz., believing truly, which do not occur in a wide class of nearby possible worlds. Insofar as Gettier cases feature luck in this sense, they are cases of *unsafe belief*, and consequently, not cases of knowing. This is because, when an agent is VE lucky, that agent forms a belief on a basis such that it could have easily been false. Indeed, in most nearby possible worlds, that agent's belief is false.

Notice that the safety condition presented above does not require that the truth of one's belief that *p* be evaluated across possible worlds. Instead, what is evaluated is the truth of beliefs formed on a certain *basis*. There is an important reason for this which can be appreciated when we consider Pritchard's example of the individual who forms beliefs about mathematical claims on the basis of flipping a coin (2014: 158). When presented with a claim, he flips a coin to decide whether to believe it. Let's suppose he is presented with the claim that $2+2=4$ and the result

of the coin flip pushes him to believe this. If an agent's belief that p is safe when an agent forms *the same belief* across nearby possible worlds and believes truly, we are forced to say that this individual's belief is *necessarily safe*, since there is no possible world in which $2+2=4$ is false. But this is clearly mistaken. We do not want to allow cases like this as cases of knowing. Intuitively, this agent is still lucky to have a true belief. After all, this individual would easily believe that $2+2=5$ on the basis of the belief-forming method they use. As such, we need to focus on the way agents form beliefs rather than the specific beliefs they hold when assessing safety. This is because, intuitively, it is the *method* of forming beliefs that is suspect here¹⁴.

Williamson (2000), Sosa (2007), and Pritchard (2005) (2014) have all defended, what Sosa calls, *basis-relative safety*:

"What is required of one's belief, if it is to constitute knowledge, is at most its having some basis that it would not easily have had unless true, some basis that it would (likely) have had only if true. When your belief that you are in pain is based on your excruciating pain, it satisfies this requirement: it would not easily have been so based unless true, it would (likely) have been so based only if true. And this is so despite its *not* being safe outright, since you might too easily have believed that you were in pain while suffering only discomfort and not pain" (Sosa, 2007: 26).

Basis-relative safety can be read as requiring that, to know that p , an agent's belief that p must be formed on a basis such that, in nearby possible worlds when one continues to form beliefs on such a basis, one continues to believe truly. As such, what is key here is the *way in which* or *method by which* beliefs are formed.

The safety condition has been presented in slightly different ways in the literature. Sosa, for example, writes that "[w]hat is required for the safety of a belief

¹⁴ It is worth noting that Pritchard has also presented a safety condition which deals only with contingent propositions. This affords us another way to respond to the issue raised concerning necessary truths (2005: 163).

is that not easily would it fail by being false" (2007: 25). This idea can be appreciated further by appealing to an example from Williamson:

"Imagine a ball at the bottom of a hole, and another balanced on the tip of a cone. Both are in equilibrium, but the equilibrium is stable in the former case, unstable in the latter. A slight breath of wind would blow the second ball off; the first ball is harder to shift. The second ball is in danger of falling; the first ball is safe. Although neither ball did in fact fall, the second could easily have fallen; the first could not. The stable equilibrium is robust; the unstable equilibrium, fragile" (2000: 123).

The thought is that, like how a ball can be in an "unstable" or "unsafe" position insofar as it could easily fall, a belief can be "unsafe" insofar as it could easily have been false. As Williamson states, a slight breath of wind would affect the ball's position. In nearby possible worlds, that ball's position is otherwise given its unstable position. The unsafe belief is similar in the sense that very little needs to change for a true but unsafe belief to be a false unsafe belief. In nearby possible worlds, that unsafe belief is false. By contrast, a safe belief couldn't easily be false.

This idea of a safe belief being a belief which couldn't easily be false captures, writes Pritchard, "an intuition about our tolerance of the risk of error" (2014: 157). If your belief that p is such that it could easily have been false, there is a high risk of error associated with this belief in the sense that very little needs to change for it to be false. Pritchard comments that "[i]n the very closest nearby possible worlds we are extremely intolerant when it comes to such epistemic risk [...] In far-off possible worlds, however, we are extremely tolerant about such epistemic risk" (ibid.). The idea here is that we can afford to be tolerant of epistemic risk in far-off possible worlds because so much must change for these possibilities to obtain. Nearby possible worlds, by contrast, look very much like the actual world. If in a nearby possible world one believes on the same basis but believes falsely, very easily could one have been in exactly that situation. This is what explains the high risk of error

associated with this sort of believing and our response of intolerance to it. In short, when there is a high risk of error, the belief is unsafe. When that risk is low, one's belief is *safe*.

Consider the following example from Pritchard (ibid.): let's say you form a belief that p by reading a reliable newspaper. This belief is safe because in nearby possible worlds, the newspaper continues to publish events accurately. This is because there are so many processes in place to ensure that the newspaper does exactly this. And since possible worlds are ordered by their likeness to the actual world, the nearby possible worlds will be those worlds in which very little changes. In a far-off possible world, a reporter for the newspaper makes a mistake which is not spotted by an editor. The newspaper goes to press, and the error gets missed multiple times, and so on and so forth. This is a far-off possible world because it differs so much to that of the actual world. So many things must go wrong for the newspaper to print an event inaccurately, and the possibility of all these events occurring is low. As such, the risk of error is low.

Compare the newspaper example to *Sniper*. Reflecting on the risk of error, we can ask: what must go wrong for Erika to believe falsely instead of truly? The answer is very little. As previously stated, if the sniper were to arrive later in the day, then Erika would still form her belief on the same basis *but believe falsely*. So, unlike the newspaper example, Erika's belief could have very easily been false. In most nearby possible worlds (if not all), her belief that there is a sniper in that building is false. Consequently, there is a very high risk of error here and so, Erika's belief is unsafe and cannot amount to knowledge.

If instead Erika had formed her belief that there is a sniper in that building on the basis of seeing a real sniper, her belief would be safe. This is because, in nearby possible worlds where the sniper arrives later in the day, Erika wouldn't form the belief *that there is a sniper in the building* or would only form that belief when she spots the sniper. So, in nearby possible worlds where she continues to form beliefs

on the same basis as the actual world, she would continue to believe truly. As such, there would be a low risk of error for Erika in this case. Moreover, we could not regard her as VE lucky. We see in *Sniper*, however, that Erika is VE lucky because the conjunction of events that result in her believing truly do not happen in nearby worlds. This is what renders her belief *unsafe* and thus, not knowledge.

The safety condition attempts to make sense of the anti-luck intuition about knowledge. If it is to do this, however, it must be the case that when an agent is VE lucky, they also have an unsafe belief. If safety fails to achieve this, then this allows for cases in which one can have a lucky, but safe, belief. And if that is so, then safety fails to do the job for which it was employed. Consequently, I suggest that, if we are to utilise basis-relative safety, we must utilise a corresponding notion of VE luck. We can amend Pritchard's definition presented earlier such that:

"[T]he agent's belief is true in the actual world, but [...] in a wide class of nearby possible worlds in which the relevant initial conditions are the same as in the actual world – and this will mean, in the basic case, that the agent at the very least forms [a belief] in the same way as in the actual world [...] – the belief is false" (2005: 146).

So, rather than ask whether an agent forms the same belief in the same way in nearby possible worlds and believes falsely, we should ask whether beliefs formed by the agent *in the same way* in nearby possible worlds are false or not.

Through utilising the safety condition, we can make sense of the incompatibility between VE luck and knowledge. VE luck is incompatible with knowledge insofar as it indicates a highly risky method of forming beliefs. The safety condition captures this intuition by demanding that knowledge entails safe belief, which can be understood as the requirement that the basis of one's belief be such that, were one to form beliefs on such a basis in nearby possible worlds, then one would continue to believe truly.

This serves as my defence of the second premise of my argument from epistemic luck, *that if the audience attains a true belief through veritic epistemic luck, that belief cannot amount to knowledge*. I will now focus of the first premise, *that if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then it would be a matter of veritic epistemic luck that the audience attained a true belief*.

VI

To defend premise one of my argument from epistemic luck, it will be beneficial to look once more at the case of *Flattering Testimony*:

Jeff is a petty boss who looks fondly only on those who flatter him. As a result, the best way to succeed in the company is to flatter him. Jeff also has a particularly strange approach to responding to testimony: he accepts what his employees say *only when it flatters him*, otherwise he rejects it. Jeff is told by an employee that he is great at tennis. As a result, Jeff forms the belief that he is great at tennis.

As Jeff discriminates between bits of testimony on the basis of whether it is flattering to him or not, his discriminations are *flattery-directed* and not truth-directed. So, the antecedent of premise one is true. If it is *not* the case that it would be a matter of veritic epistemic luck that Jeff attained a true belief, then the consequent of premise one is false. If that is so, then my argument of epistemic luck fails.

As the falsity of the consequent necessitates Jeff's belief being true, let us imagine that Jeff is a very skilled tennis player. However, we will have to suppose that Jeff has never reflected on his abilities until an employee tells him that he is great at tennis. As such, when he receives his employee's testimony, he forms a true belief *that is testimonially based*. The question we must now consider is whether it is a matter of luck that Jeff has a true belief.

One way to answer this question is to ask whether Jeff *knows* that he is great at tennis *on the basis of his employee's testimony*. This is because, as I have argued, VE luck and knowledge are incompatible. It follows, therefore, that if Jeff knows that he is great at tennis, he cannot be VE lucky. The case here looks quite odd though. We must imagine someone humble enough not to recognize that he is good at tennis but vain enough to be undiscerning about flattery. If the reader is unable to imagine such a case, please note that my defence of the first premise is not dependent on the intuitions that this case elicits.

If Jeff doesn't believe that he is great at tennis initially, then he forms a true belief when he receives the flattering testimony that he is great at tennis from his employee. Intuitively, however, Jeff doesn't *know* that he is great at tennis. One simple way to see this is to recall my defence of premise one from my *argument from doxastic responsibility*. In failing to be sensitive to considerations which could be advanced in favour of the truth, or falsity of his employee's testimony, Jeff is doxastically irresponsible. And since doxastic irresponsibility prohibits knowledge acquisition, it follows that Jeff cannot know that he is great at tennis from his employee's testimony. A second, and related suggestion, as to why Jeff doesn't attain knowledge here is that his belief lacks *justification*. Insofar as he neglects considerations which have a bearing on the truth or falsity of the claim *that he is good at tennis*, he is without good reason to believe that he is great at tennis and thereby fails to have a justified belief. Whilst the belief he forms may be true, this does nothing to change the fact that he neglects considerations to which he *should* be sensitive.

Let us now directly consider the question of whether Jeff is VE lucky. Appealing to our previous discussion of safety, we can say that there is a high risk of error associated with Jeff's testimonial beliefs. This is because we can imagine nearby possible worlds in which Jeff makes flattery-directed testimonial discriminations *but believes falsely*. In the nearby possible world in which he is told

that he is great at tennis, but is not, he would form the same belief on the same basis but believe *falsely*. We can also imagine possible worlds in which Jeff's employees know about his strange practice of discriminating between bits of testimony in a flattery-directed manner and so, they flatter him in an attempt to progress in the company. Again, in these worlds, Jeff can very easily form false beliefs. As such, whenever he forms a true belief through employing this method of forming beliefs, *he is lucky to have a true belief*. Otherwise put, he is VE lucky and, as argued, beliefs which are VE lucky cannot amount to knowledge.

The extent to which a piece of testimony is flattering is completely disconnected from the truth, or likely truth, of that piece of testimony, and so a belief based on such a discrimination could easily be false, irrespective of whether it is false. As such, if one discriminates between bits of testimony in such a way, one would be lucky to form a true belief via testimony. However, we can amend *Flattering Testimony* to present a greater threat to my argument:

Suppose that all hiring decisions in Jeff's company are handled by Elene. Elene is incredibly sensitive to detail and only hires those she regards as trustworthy. Moreover, she has made it company policy that employees only report that which they know to be true. If any worker breaches company policy, she terminates their contracts without delay. All workers are aware of this policy and tend to act in accordance with it.

In Jeff's company, there is a strong culture of only reporting that which you know to be true. As such, when Jeff receives testimony from a worker, the chance that it is true is very, very high. This means that, despite making flattery-directed testimonial discriminations, when Jeff forms a belief as the result of an employee's testimony, it is very likely that that belief will be true since his employee wouldn't say that he is great at tennis unless they know it to be true. And so, it is not the case that Jeff's testimonial beliefs could easily be false as there is a policy in place to decrease the risk of Jeff forming false testimonial beliefs.

Given this amended version of *Flattering Testimony*, what are we to make of Jeff's beliefs resulting from his flattery-directed testimonial discriminations? Are we to say that he is lucky to form true beliefs? We can go in different directions here depending on our intuitions. Let us begin with the intuition that Jeff is VE lucky; that in nearby possible worlds in which he forms testimonial beliefs on the basis of flattery-directed testimonial discriminations, he believes falsely. Adopting this position, one might draw attention to the epistemically unusual environment that exists within Jeff's company.

If we suppose that the world Jeff inhabits is much like ours, then his office environment looks significantly different, epistemically, to the environment outside of his office. There is a very strong pressure on speakers in the office to report only that which they know to be true where a comparable pressure is absent outside of the office. Admittedly, there is *some* pressure on speakers outside of Jeff's office to report truthfully, but not all speakers succumb to that pressure. We know that people sometimes report falsities. Given this, one might say that Jeff is lucky whenever he forms a true testimonial belief because, if he was outside of his office, he would form a testimonial belief in the same way, but that belief could very easily be false. Alternatively, one might introduce the possibility of an angry worker who flouts company policy, or a visitor to the company who is unaware of the policy. When we consider these possible scenarios, we see that the possibility of Jeff attaining a false testimonial belief is still a real concern in this case. And insofar as these possibilities can arise in nearby possible worlds, we are committed to saying that Jeff is lucky to form true testimonial beliefs because, if he received flattering testimony from someone other than a policy-abiding employee, he could very easily end up with a false belief given the manner in which he discriminates between bits of testimony.

However, one might be of the view that Jeff is not VE lucky. In response to the worry above, perhaps one might extend the scope of the amended case of

Flattering Testimony. Let's suppose that the policy enforced by Jeff's company is not isolated to his company alone but is heavily enforced by a tyrannical government. In this world, speakers have immense pressure on them to report only that which they know to be true, lest they suffer severe punishment. In such a world, Jeff cannot be regarded as lucky in forming true testimonial beliefs given his flattery-directed testimonial discriminations because, in nearby possible worlds, Jeff would continue to believe truly on the basis of making flattery-directed testimonial discriminations. As such, we have a case in which an agent forms a true belief on the basis of a discrimination which is not truth-directed *but which is not true by luck.*

Let us take seriously the concern raised here. In this imagined world, the epistemic heavy lifting is done entirely by the speakers. There is no need for audiences to pay attention to truth-conducive considerations because the probability of the speaker reporting truthfully is immensely high. But this world looks incredibly different to our world. So, perhaps the truth-directed testimonial discriminations condition is a condition only on worlds like ours; worlds in which speakers behave in a particular way. In this world, speakers sometimes tell the truth and sometimes fail to do so. Given this, it makes sense why there is a need to be sensitive to considerations which could be advanced in favour of the truth or falsity of a piece of testimony, viz., not all speakers transmit knowledge (be it intentionally or otherwise). In light of this, one route to take is to make it explicit that the truth-directed testimonial discriminations condition is a condition on justified testimonial belief and testimonial knowledge *only in worlds like ours.* This would mean that, when questioning whether an agent is VE lucky with the aim of establishing the falsity of premise one of my argument from epistemic luck, we need to look at worlds like ours; worlds in which speakers sometimes tell the truth and sometimes don't.

Consider an analogy. Imagine the world in which only one type of bird exists: the buzzard. In this world, getting the slightest glimpse of a bird is sufficient to allow

one to gain a justified belief, or even knowledge, that one just saw a buzzard. But this is only because of the way that world looks. The world we currently inhabit, by contrast, is different. This world is one in which there are many types of bird. To know that the bird you see is a buzzard requires something of you which is not required in the world where the only type of bird is the buzzard. In this world, to know that the bird you see is a buzzard requires an ability to discriminate between buzzards and other sorts of birds simply because the bird you see might not be a buzzard. Our world is similar with respect to testimony; the testimony you are receiving might be veridical, but it might not. This fact puts an epistemic burden on the audience. One can't just consume testimony carelessly, rather, one must discriminate between bits of testimony in a truth-directed manner because there is a risk that the testimony one receives is not veridical even though the speaker may present it as such.

We are now well-placed to respond to the concern raised by the amended case of *Flattering Testimony*. The concerns raised by the counterexample fail because they present a case for which the truth-directed testimonial discriminations condition is not designed. If we think of our capacity to make truth-directed testimonial discriminations as an evolved ability, then we can push this point further. The point is that such an ability would not exist were the environment substantially different. The fact that we have this ability speaks in favour of the idea that the environment we inhabit is one in which speakers can sometimes report falsities. It is *because* of this that we have an ability to discriminate in a truth-directed manner. If our world was different, so would this ability be.

In response to this, I suggest that we amend premise one so that it reads like so:

1. If, in an environment wherein speakers sometimes tell the truth and sometimes fail to do so, a given piece of testimonial acceptance is based

on a discrimination which is not truth-directed, then it would be a matter of veritic epistemic luck that the audience attained a true belief.

Given the revision of premise one, the amended case of *Flattering Testimony* fails to prove a problem to my argument because it involves a different kind of epistemic environment.

Whilst premise one of my argument from epistemic luck maintains that any testimonial acceptance on the basis of a discrimination which is not truth-directed will entail that any resulting true testimonial belief will be VE lucky, it is important to note that I do not take the truth-directed testimonial discrimination condition to be an anti-luck condition. One can do their best to discriminate between bits of testimony but come to believe through VE luck; Gettier cases demonstrate exactly this. Nevertheless, given my arguments from this section, I take it that the anti-luck condition, be it safety or something else, should be seen as a substantial feature of knowledge. It is not simply a way to sidestep troublesome Gettier cases; it is an essential part of knowledge. I have argued that truth-directed testimonial discriminations are *necessary*, but not *sufficient*, to avoid VE luck. Therefore, the truth-directed testimonial discriminations condition is a necessary condition on justified testimonial belief and testimonial knowledge. At this juncture, I take it that the truth-directed testimonial discriminations should be viewed as something like a *justification condition*, though I do not commit myself to this.

Having defended premises 1 and 2, we can now reach the conclusion of my argument from epistemic luck: (EP) *that if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge*. However, given that I restricted the scope of premise one of my *argument from epistemic luck* to worlds like ours, EP should be likewise restricted:

(EP) If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, in an environment wherein

agents sometimes tell the truth and sometimes false to do so, then the audience's resulting belief cannot amount to knowledge.

EP features as a premise in my *argument for truth-directed testimonial discriminations*. Taken with the claim (EF): *that we acquire knowledge from testimony (by and large)*, we can reach the conclusion:

Epistemic Consequent (EC): Our testimonial discriminations must be truth-directed (by and large).

We can arrive at this conclusion because EP states that a necessary condition of acquiring testimonial knowledge is that testimonial acceptance be based on a truth-directed testimonial discrimination. EF states that, we acquire knowledge from testimony (by and large). Since it is the case that we acquire knowledge from testimony by and large, it follows that EC: *our testimonial discriminations must be truth-directed (by and large)*.

VII

Before considering some objections to my proposal, it is important to return once more to the definition of truth-directed testimonial discriminations presented earlier. Following my argument from doxastic responsibility, I suggested that we should think of truth-directed testimonial discriminations as a matter of doing one's best to get to the truth. Recall the example in which you are trying to get to the cathedral in an unfamiliar city. If you follow signs which are misleading, there is a sense in which your travel is "cathedral-directed" and a sense in which it isn't. It is *not* cathedral-directed in the sense that you are not travelling towards the cathedral. But it is cathedral-directed in the sense that the moves you make are in keeping with the aim of getting to the cathedral. I suggested that we should think of truth-directed testimonial discriminations in the latter sense. The arguments from this chapter give us reason to revise this understanding.

Imagine again that you are in an unfamiliar city, and without any kind of map, you check for street signs to help you locate the cathedral. You follow the street signs and eventually reach the cathedral. Are you lucky to reach the cathedral? It doesn't seem so. But what if, like in the earlier case, those signs were put there by someone trying to confuse tourists? The signs you were following were *misleading signs*; they lead away from the cathedral. Despite this, you somehow managed to reach the cathedral. What are we to say now? Surely, we must say that you were lucky. You were lucky because following misleading signs is not a good way to reach one's destination. The result is that, insofar as we regard you as *lucky*, we cannot give you any *credit*. It would be improper to attribute your success in reaching the cathedral to any skill or ability you exercised. Rather, you reached the cathedral by dumb luck. And achieving things purely by luck is not something creditable.

Contrast the case above with the scenario in which you reach the cathedral by following reliable signs. In this case, it seems that we can credit you. We can give you credit because you rely on a good method to reach the cathedral. It was not through luck that you reached the cathedral, it was the result of you using something which has a clear connection to the location of the cathedral, viz., signs which direct you to the cathedral. Through appealing to a similar intuition concerning knowledge as the result of cognitive ability and, therefore, a kind of epistemic "success" or "achievement", we can draw an analogy here.

We can utilise the virtue-reliabilist position discussed earlier to maintain, as does Greco (2003), Sosa (2003), Pritchard (2012), Kelp (2013) and Carter, Jarvis & Rubin (2015), that knowledge is a *creditable success resulting from the exercise of one's cognitive abilities to get to the truth*. The idea here is that knowledge is not something which happens accidentally, rather it is something resulting from your *abilities to get things right*. Consider again Sosa's example of the archer at a target practice. Imagine two scenarios: one in which she hits her target through her skill and one in which a strong gust of wind blows the arrow in such a way as to cause it

to strike the target in exactly the same place. Our response to the question of why she hit the target differs from one case to the other. In only one of these cases do we credit the archer's hitting the target to her abilities. In the other case, we explain the archer's hitting the target by referencing the gust of wind. We can apply this thought to the matter at hand.

For Greco, to know that p requires that you reach the belief that p in *the right sort of way* (2003). This means arriving at the belief that p thanks to your own abilities and not luck or any other intervening factors¹⁵. He writes that:

"[T]o say that someone knows is to say that his believing the truth can be credited to him. It is to say that the person got things right owing to his own abilities, efforts, and actions, rather than owing to dumb luck, or blind chance, or something else" (2003:111).

Recall the virtue-realist position that virtues consist in abilities, or faculties, which produce more true than false beliefs. Understood in this way, Greco suggests that knowledge entails belief resulting from one's ability to get to the truth *reliably*. That is, an ability to *tend to get to the truth* (Goldman, 1979/2012: 37).

When you form the true belief that the cathedral is in the old part of the city by reading a map, or following reliable signs, you satisfy this "ability condition" because your belief results from your ability to demonstrate a sensitivity to considerations which could be advanced in favour of the truth of that claim. It is *because* you reasoned in this way that you have a true belief and because of that, you deserve credit. With respect to testimony, we can credit an agent who comes to know that p from a speaker's testimony that p on the grounds that this agent has exercised a cognitive ability in the form of demonstrating a sensitivity to considerations which could be advanced in favour of the truth, or falsity, of a piece of testimony. In short, we can credit the agent because they accepted a piece of

¹⁵ For example, inhabiting an environment in which those around you do all the epistemic heavy lifting, as in the amended case of *Flattering Testimony*.

testimony on the basis of a discrimination which was truth-directed. This suggests that we should think of making truth-directed testimonial discriminations as an ability to get to the truth.

But one might, as Lackey does, object to the thought that knowledge is deserving of credit insofar as it evidences belief resulting from one's ability (Lackey, 2007). For Lackey, there exist cases in which one knows that p but is not deserving of credit for it. Consider the following case she puts forward:

"CHICAGO VISITOR: Having just arrived at the train station in Chicago, Morris wishes to obtain directions to the Sears Tower. He looks around, approaches the first adult passerby that he sees, and asks how to get to his desired destination. The passerby, who happens to be a lifelong resident of Chicago and knows the city extraordinarily well, provides Morris with impeccable directions to the Sears Tower by telling him that it is located two blocks east of the train station. Morris unhesitatingly forms the corresponding true belief" (2009: 29).

Lackey writes that, "while Morris clearly knows on the basis of testimony that the Sears Tower is two blocks east of the train station, he does not deserve the requisite kind of credit for truly believing this proposition" (ibid.). If Lackey is right, then there is an issue with the current proposal.

My response to this objection is that this example fails to reach the conclusion Lackey suggests it does. The problem is that Lackey focuses on Morris's *belief*, neglecting the *route* to belief and, consequently, his discriminative capacity. If we ascribe to Morris knowledge, it is not clear what prevents us from replying that he believes through his capacity to make truth-directed testimonial discriminations. To offer a more detailed response, however, I suggest that we lack the sort of information we need to determine whether Morris has knowledge or not. We must ask whether Morris is disposed to accept *whatever* was told to him. Would he, for example, accept that the Sears Tower is at the bottom of the Monroe Harbour?

Would he accept the testimony of a person worriedly looking at a map? If he is so disposed, then I take it that Morris wouldn't attain knowledge from that speaker's testimony. This would be explained by Morris's failure to exercise his cognitive abilities to get to the truth.

With that said, I am happy to agree with Lackey that Morris gains knowledge here. Contrary to Lackey's suggestion, however, I take it that he does deserve credit. One of the reasons why can be found, I propose, in the work of Zagzebski:

"The particular kind of praiseworthiness that applies to virtue, then, reflects the fact that the virtuous person might have been vicious instead" (1996: 105).

Zagzebski's position is that one need not be "virtuous"; one could do otherwise. This thought is common in the moral domain. We can understand why the generous person is deserving of credit by considering that they didn't have to be generous; they could have been selfish instead. Applied to the matter at hand, we can say that Morris *could have* engaged in epistemically dubious practices, but he didn't. As such, when one exercises one's cognitive abilities to get things right, one is deserving of credit. When you reach the cathedral by using a good method, you deserve credit. When you reach the cathedral by dumb luck, you don't. The same is true in the epistemic domain.

Nevertheless, I recognize the intuition which informs Lackey's example. Per Lackey, insofar as the speaker, *S*, knows the location of the Sears Tower and passes this knowledge to Morris, it is *S* and not Morris who is deserving of credit (2007: 358). My response does not force one to disagree with the notion that *S* deserves credit. Instead, I mean merely to argue that Morris deserves credit too because, on the route to knowledge, he employs a cognitive ability to get things right.

But one might ask: what sort of ability could Morris possibly be exercising in the process of coming to form the belief that the Sears Tower is located two blocks east of the train station? My response is that, in the scenario wherein Morris attains

knowledge, he accepts a piece of testimony on the basis of a truth-directed testimonial discrimination. But if knowledge entails belief through a cognitive ability to get things right, this means that making truth-directed testimonial discriminations must be such an ability; it must be a *reliable* ability in that it tends towards true belief (Goldman, 1979/2012: 38). If we make this specification, we can say that, in making truth-directed testimonial discriminations, Morris is doing exactly what he should do, insofar as he is exercising a cognitive ability to get to the truth. Through making truth-directed testimonial discriminations and thereby demonstrating a sensitivity to considerations which could be advanced in favour of the truth or falsity of a piece of testimony, Morris can be viewed as deserving at least *some* credit when he achieves knowledge. This is because, Morris is disposed such that, if he had reason to reject the speaker's testimony, he would have done so. A gullible agent would not act in such a way.

How does this affect our understanding of what it is to make truth-directed testimonial discriminations? I see it like so: simply doing one's best and acting as the reasonable person would act isn't enough. We need to accommodate the anti-luck intuition, thereby ruling out cases wherein agents are lucky to get to the truth by means an epistemically flawed belief-forming method. What we need is an understanding of truth-directed testimonial discriminations which will not allow for *systematic error*. By this, I mean that it should not be the case that, in making truth-directed testimonial discriminations, one can get things wrong more often than one gets things right. Put another way, I suggest that we add to our current understanding that truth-directed testimonial discriminations must be *reliable* in the sense described by Goldman (ibid.). That is, our ability to make truth-directed testimonial discriminations must be such that it tends towards true belief. Our understanding of truth-directed testimonial discriminations as a matter of doing as any reasonable person would do to get to the truth should be revised to accommodate this. I suggest that we think of making truth-directed testimonial

discriminations as a matter of doing what the reasonable person would do to get to the truth by way of a reliable cognitive ability.

Notice that, in understanding truth-directed testimonial discriminations in such a manner, we can accommodate internalist and externalist intuitions about justification and knowledge. In my discussion of doxastic responsibility, I dealt primarily with the intuition that agents must be sensitive to considerations which have a bearing on the truth of some piece of testimony, p ; that is, I dealt with the intuition that one cannot come to know that p whilst being neglectful. The villagers in *Continued Acceptance* fail in this respect. They are gullible and, consequently, neglect important considerations. The guiding thought here motivates an *internalist* view, in that as it advances a requirement which makes no reference to factors external to an agent's mental life. To be doxastically responsible is to respond to the rational force of considerations which concern the truth of p , irrespective of whether some of those considerations may be misleading. It is, in a sense, to *try* to get to the truth. This intuition drives forward the idea that making truth-directed testimonial discriminations is a matter of doing what any reasonable person would do to get to the truth.

In my discussion of epistemic luck, I dealt primarily with the anti-luck intuition that one cannot come to know that p if it is a matter of luck that one's belief that p is true. Erika in the *Sniper* example fails in this respect. Through no fault of her own, she forms her belief *that there is a sniper in that building* on the basis of seeing a cardboard cut-out of a sniper. In doing so, she is lucky to have a true belief since, in nearby possible worlds, where she continues to form beliefs on such a basis, she believes falsely. The guiding thought here motivates an *externalist* view, in that it advances a requirement which appeals to at least some factors *external* to an agent's mental life. Consider, whether one uses a belief-forming practice which has a high risk of error, in that, in nearby possible worlds where one continues to form beliefs on the same basis as in the actual world, one believes falsely, is not

determined purely by factors dependent on the agent. What we see from cases like *Sniper* is that simply doing as any reasonable person would do to get to the truth isn't always enough. Sometimes, one is lucky to get to the truth. In response, we must accommodate the anti-luck intuition by realising truth-directed testimonial discriminations as a matter of doing what the reasonable person would do to get to the truth by way of a reliable cognitive ability. Let us now turn to some objections.

VIII

1: Children as Undiscriminating Consumers of Testimony

I have argued that a necessary condition on the attainment of testimonial knowledge and justified testimonial belief is that we make *truth-directed testimonial discriminations*. But children are often presented as indiscriminating consumers of testimony, simply accepting whatever they are told. This implies that testimonial knowledge is beyond the grasp of children since they would fail to satisfy this necessary condition on the attainment of testimonial knowledge. The alternative position is to suggest that children can attain testimonial knowledge, which then puts pressure on my proposal. I intend to show that this challenge is unproblematic.

One straightforward response is to accept the suggestion that testimonial knowledge is beyond the grasp of children. Another is to propose that the requirements on children for the attainment of testimonial knowledge and justified testimonial belief are weaker than those for adults. Maybe it is precisely *because* children are indiscriminating consumers of testimony that we must be more lenient. Neither approach is necessary, however. Instead, I maintain that children are perfectly capable of making truth-directed testimonial discriminations, as the empirical data indicates.

Whilst some studies have found that three-year-olds struggle to consistently discriminate between veridical and false testimony, or between trustworthy and untrustworthy individuals, those results differ when looking at five-year-olds, or even

four-year-olds (Mascaro & Sperber, 2009) (Jaswal, Croft, Setia & Cole, 2010). This is not to pretend, however, that there is unanimous agreement on the matter (Vosniadou & Brewer, 1992).

Jaswal, Croft, Setia & Cole suggest that three-year-olds have a “highly robust bias to trust testimony” (2010: 1546). In the *lying experimenter* study, a three-year-old participant was informed by a confederate that they would play a game, the objective of which was to find as many stickers underneath upside-down cups as they could. The experimenter sat down opposite the child and saw her place two differently coloured cups down on a tray. The cups were then concealed by a small screen. The experimenter, who was not obscured by the screen, showed the participant a sticker and then concealed it underneath one of the cups, acting as though she hid it under both to rule out the possibility that the participant relied on perceptual cues. Next, the screen was removed, and the experimenter asserted that the sticker was in the opposite cup, e.g., she asserted it was in the red cup when it was in the yellow cup. The experimenter then pushed the tray forward to allow the child to select a cup, with the aim of collecting a sticker. There were 8 trials of this experiment.

A further study, called the *arrow condition*, was conducted which differed very slightly to this. In this variant, children were trained to recognise that a cardboard arrow, attached to a stand, pointed to a box with puppets in it. After a while, the experimenter entered the room, and they played the sticker game again. However, instead of offering testimony about the location of the sticker, the experimenter positioned the arrow so that it pointed towards the empty cup. There were 8 trials of this experiment.

Jaswal, Croft, Setia & Cole found that there was a 0% success rate in the first trial of both experiments. However, this quickly changed. Children in the arrow condition performed much better than those in the testimony condition. 56% of children in the testimony condition *never found a sticker*; that is, they accepted what

the experimenter said and acted in accordance with her testimony *in every trial*. Jaswal, Croft, Setia & Cole write that:

“Children’s difficulty in the testimony condition relative to the arrow condition suggests their willingness to believe what they are told is not merely the result of a general inclination to trust people. Instead, it seems to reflect a more specific, highly robust trust in testimony” (ibid.).

That is, there seems to be something about testimony that makes it particularly difficult for young children to avoid accepting what is reported to them. Consequently, making testimonial discriminations, let alone *truth-directed testimonial discriminations*, might be a struggle for young children.

Now consider some empirical work from Mascaro & Sperber. They conducted a range of experiments to test the *epistemic vigilance* of children from the age of three to the age of five. In testing epistemic vigilance, these studies tested “an ability aimed at filtering out misinformation from communicated contents” (2009: 367). These studies might also be seen as assessing the ability of children to make truth-directed testimonial discriminations, therefore.

In the *big liar false communication study* (BLFCS), the experimenter explained to the child participant (either a three-year-old or a four-year-old), that he would hide a sweet inside a box. The child was then asked to turn around and, after hiding the sweet, was told to turn around again. Next, a frog puppet was introduced and shown to investigate the contents of boxes. After this, the experimenter warned the child “Now, the frog is going to talk to you, but be careful! The frog is a big liar! It always tells lies” (ibid. 371). To ensure that the children comprehended that the frog is a liar, the experimenter explicitly asked the children “Is it a liar? Does it tell lies?” (ibid.). Then, the experimenter made the frog puppet touch one of the boxes and say in a distinctive voice “The sweet is in the red box!” (ibid.). And after this, the child was asked where the sweet is. The game was then repeated afterwards.

A further study was conducted, which was identical to BLFCS except that it did *not* involve a warning about the frog's deceptive tendencies. The child participants had no information about whether the frog tells truths or lies. This is the *trust baseline assessment*.

In the trust baseline assessment, Mascaro and Sperber found that "all children trusted the informant" (ibid.). Interestingly, in BLFCS Mascaro and Sperber found that:

"All 3-year-olds trusted the lying character, thus leading to a performance score of 0%, significantly below a chance score of 50% of correct answers [...] and identical to the baseline trust level. Conversely, 4-year-olds gave 77% of correct answers. This result was significantly above that of 3-year-olds" (ibid. 372).

This suggests that, while three-year-olds might lack a capacity to make truth-directed testimonial discriminations, by the age of four, that capacity has started to develop.

Now consider the *kind vs. mean puppet* study in which three-year-old children were presented with a closed container sat between two puppets on a table. The goal was for the child to find out the contents of the container. One of the puppets was said to be "kind" and the other was said to be "mean". The lid was opened such that the puppets could look into it whilst the child could not. After the puppets looked into the box, each puppet gave testimony as to the contents of the box. The kind puppet always gave veridical testimony, e.g., "there is a spoon in the box" when there is, whilst the mean puppet always gave false testimony, e.g., "there is a fork in the box" when there is a spoon in it. After playing this game once, the experimenter informed the child that they would play again but with a new box. The experiment was then repeated with the puppets delivering testimony in accordance with their status as "kind" or "mean".

Mascaro and Sperber write that “[c]hildren performed above chance on this measure of performance with a mean score of 1.36” (ibid. 370). However, of the 25 children tested, 10 gave inconsistent results insofar as they “did not trust the same character over the two tasks, but changed their preference (i.e. selected the testimony of the kind character once and the testimony of the mean character once)” (ibid.). This suggests that, at the age of three, children can make truth-directed testimonial discriminations, though the capacity to make such discriminations is not fully realised. Nonetheless, Mascaro and Sperber take it that:

“This, together with earlier work showing that children take into account an informant’s competence, establishes that young children are not simply gullible and do adjust their level of epistemic trust in a variety of ways” (ibid.).

There is reason, based in empirical data, to reject the view that children are indiscriminating consumers of testimony then. But the results for three-year-olds are certainly poorer than those for four-year-olds.

Given the empirical data discussed above, we must reject the idea that children are indiscriminating consumers of testimony, gullibly accepting whatever they are told. In response, one might just narrow the scope of this challenge. Rather than *children* being indiscriminating consumers of testimony, one might argue that *young children* are indiscriminating consumers of testimony. From here, the original criticism follows. I see this bullet as far less painful to bite. The set affected is much smaller than the set of *children* insofar as it would seem to contain those of three years of age and under. It may be that testimonial knowledge is beyond the reaches of three-year-olds, but available to four-year-olds.

The findings here accommodate well the suggestion that the capacity to make truth-directed testimonial discriminations consists in a cognitive ability, or a kind of *know-how*. That cognitive ability, at a very young age, is not fully realised. No wonder then, that young children perform poorly in these experiments.

2: The Presence of Implicit Bias

A second objection comes from the wealth of data accumulated from studies into *implicit bias*. These studies show that we can demonstrate biases without endorsing them, or even being aware of them (Greenwald, et al., 1998). Such biases have been shown to affect behaviours, such as hiring decisions (Steinpreis, Anders and Ritzke, 1999). If such biases indicate underlying preferences, this may impact the credibility we give to speakers. We may, for example, give more credibility to a speaker because they belong to a particular social group. And if this is so, then perhaps it is instead *implicit biases* which push us to accept and reject testimony and not a capacity to make truth-directed testimonial discriminations.

This objection prompts, what I will call, the *how-possible question*: How can we acquire testimonial knowledge given that so many biases are in play in the testimonial domain? The issue with this is that it questions something I take for granted. I take it for granted that we attain knowledge from the testimony of others, so any objection to this point is something with which I will not engage. No doubt, there is a time and place for dealing with such questions, but it not within the scope of this project to deal with such matters.

Nevertheless, one might ask: does the fact that we have so many biases which drive testimonial acceptance not prove troublesome for a proposal requiring that testimonial knowledge necessitates the making of truth-directed testimonial discriminations? And to this, I will provide a brief response. The data attained from years of research into implicit bias indicates that we have biases. I am happy to concede that the existence of such biases might influence given instances of testimonial acceptance through disposing us to treat social-group membership as a relevant consideration when it comes to testimonial acceptance or rejection. The question is whether those biases can coexist with the proposed ability to make truth-

directed testimonial discriminations. And I do not see any reason to believe that this is not possible.

One suggestion, though I am not committed to it, is that whatever is cognitively responsible for truth-directed testimonial discriminations comes apart from the process/es responsible for implicit bias. To be clear, this is not to deny the presence of implicit bias but simply to stipulate that the process of making truth-directed testimonial discriminations need not be affected by the presence of implicit bias. Instead, what might occur is the “overlying” of biases such that when a given agent makes truth-directed testimonial discriminations, their biases lay on top of the discriminations, rather than infuse with them. What this means is that, whilst biases might affect testimonial acceptances in *some environments*, viz., those environments which activate our biases, it need not follow that such biases affect our capacity to make truth-directed testimonial discriminations on a global scale¹⁶. Given this position, it is possible to both recognize the presence of implicit bias and defend the truth-directed testimonial discriminations condition.

IX

In this chapter, I have argued for (EC): *that our testimonial discriminations must be truth-directed (by and large)*. I reached this by way of my *argument from doxastic responsibility* and my *argument from epistemic luck*.

In my argument from doxastic responsibility, I argued that doxastic irresponsibility is incompatible with knowledge. In arguing this, I highlighted some work in virtue epistemology. I argued that this broad position is helpful in explaining why epistemic vices, such as gullibility and stubbornness, fail to allow for the attainment of knowledge. Appealing to McDowell (1998), I argued that, instead,

¹⁶ This allows for the possibility that biases might close some opportunities for testimonial knowledge. Those testimonial encounters in which the operation of bias overly limits an audience’s capacity to make truth-directed testimonial discriminations will be encounters where the door to knowledge is shut.

agents must be *doxastically responsible*. Enforcing this as a necessary condition for knowledge and justified belief allows us to capture the intuition that knowledge is the result of believing *in the right sort of way*.

After defending premise two, I turned my attention to premise one, that if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience behaves in a doxastically irresponsible manner. Prior to defending this, however, I gave some detail as to how we should understand truth-directed testimonial discriminations. I argued that, to make truth-directed testimonial discriminations is to be doxastically responsible. One can therefore regard truth-directed testimonial discriminations as a class instance of doxastic responsibility. I noted that we can further understand *truth-directed testimonial discriminations* in a number of ways; the cathedral example helps to illustrate this. Following on from that example, I suggested that, in line with the argument from doxastic responsibility, we should think of truth-directed testimonial discriminations as a matter of doing as the reasonable person would do to get to the truth.

Given the connection between truth-directed testimonial discriminations and doxastic responsibility, the defence of premise one was swift. I suggested that, as testimonial discriminations which are not truth-directed are not concerned with considerations that could be advanced in favour of the truth, or falsity, of p , premise one follows. From here, I was able to conclude *that if a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge*.

In §V, I introduced my argument from epistemic luck. I then worked to defend the second premise, *that if the audience attains a true belief through veritic epistemic luck, that belief cannot amount to knowledge*. This involved a discussion of differing forms of luck in an attempt to establish which forms of luck are knowledge-impeding. Thereafter, I argued in support of premise one, *that if a given piece of testimonial acceptance is based on a discrimination which is not truth-*

directed, then it would be a matter of veritic epistemic luck that the audience attained a true belief. Defending this claim required a small revision to the argument. Insofar as there exist environments in which the epistemic heavy lifting can be carried out by others, there exist environments in which an audience can come to attain true beliefs from others by putting in little to no work. That is, one can fail to demonstrate any sensitivity to truth-conducive considerations, yet tend to form true beliefs due to the epistemic work others have put in. This is problematic for my proposal as it allows for cases in which the antecedent of premise one obtains while the consequent does not. In response, I suggested that the kinds of environment in which the truth-directed testimonial discrimination must be assessed are those environments in which speakers sometimes tell the truth and sometimes fail to do so. In short, environments like the ones we typically experience. In such environments, an audience would be lucky to form a true belief from accepting a piece of testimony on the basis of a discrimination which is not truth-directed because the audience's belief could easily have been false. The consequence of this is that premise one must be read as a claim about worlds, or epistemic environments, like ours.

The work carried out in the presentation of my argument from epistemic luck forced me to reconsider my understanding of truth-directed testimonial discriminations. In §IV, I suggested that we understand truth-directed testimonial discriminations as a matter of doing what the reasonable person would do to get to the truth. The issue with this, I suggested, is that if one engages in an epistemic practice which tends to result in failure, then one will be lucky whenever one gets to the truth. To overcome this issue, I revisited the virtue-realist position, remarking that knowledge can be considered as a state deserving credit insofar as it results from one's cognitive ability to get things right. Through appealing to this idea, I suggested that truth-directed testimonial discriminations must accommodate this thought or else be vulnerable to the challenge that one can make truth-directed

testimonial discriminations and suffer systematic error. The result of this discussion saw me revise my understanding of truth-directed testimonial discriminations. I take it that making truth-directed testimonial discriminations is a matter of doing what the reasonable person would do to get to the truth by way of a reliable cognitive ability.

Next, I introduced some objections to my proposal that there is a truth-directed testimonial discriminations condition on the attainment of justified testimonial belief and testimonial knowledge. I found that the objection concerning children and testimony relies on a misunderstanding of the capabilities of children. We have good reason, given current empirical data, to believe that children are *not* indiscriminating consumers of testimony. The second objection, concerning implicit bias, introduces, what I have called, the how-possible question, viz., the question of how it is possible that we can attain testimonial knowledge given the presence of implicit bias. While I acknowledge the importance of this question, I do not purport to deal with it in this thesis. Nevertheless, I maintain that the presence of implicit biases does not necessarily impact the making of truth-directed testimonial discriminations. It is entirely possible that whatever is cognitively responsible for our truth-directed testimonial discriminations comes apart from whatever is cognitively responsible for our implicit biases.

With both arguments complete, we can now return to my argument for truth-directed testimonial discriminations:

Social Fact (SF): We discriminate between bits of testimony.

Epistemic Principle (EP): If a given piece of testimonial acceptance is based on a discrimination which is not truth-directed, then the audience's resulting belief cannot amount to knowledge.

Epistemic Fact (EF): We acquire knowledge from testimony (by and large).

Epistemic Consequent (EC): Our testimonial discriminations must be truth-directed (by and large).

To this, we can ask the question: *What makes our discriminations truth-directed? How is it that they are directed so?* In doing so, we are attempting to account for the *truth-directedness* of our testimonial discriminations.

The rest of this thesis serves as an attempt to explain the truth-directedness of our testimonial discriminations. In doing so, I will survey the following positions:

Non-Reductive Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to counterevidence.

Reliabilist Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process.

Assurance Hypothesis: Our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said.

Evidential Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence.

I now conclude chapter one. The arguments provided in this chapter serve as a prelude to the rest of this thesis. In this chapter, I have defended the claim *that our testimonial discriminations must be truth-directed (by and large)*. The remainder of this thesis serves to analyse the explanatory hypotheses presented above. I will begin that analysis with the non-reductive hypothesis in the next chapter.

2. The Non-Reductive Hypothesis

I

In this chapter, I examine the first of four explanatory hypotheses for the truth-directedness of our testimonial discriminations: the non-reductive hypothesis. According to this hypothesis, our testimonial discriminations are truth-directed because they demonstrate a sensitivity to counterevidence. I will argue that the truth of this hypothesis necessitates a clear distinction between evidence and counterevidence. The problem is that no such distinction exists; that which is evidence that p can, depending on the context, be counterevidence that $\sim p$. The upshot is that the non-reductive hypothesis fails to provide an explanation of the truth-directedness of our testimonial discriminations which is distinct from the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. It can therefore be eliminated due to its redundancy.

In the previous chapter, I argued that a necessary condition on the attainment of testimonial knowledge is that we make truth-directed testimonial discriminations. If we tended to accept testimony on the basis of discriminations which were *not* truth-directed, we would not be able to acquire testimonial knowledge, by and large. Since it is clearly the case that we acquire knowledge from testimony (by and large), it must also be the case our testimonial discriminations are truth-directed (by and large). But what explains the truth-directedness of our testimonial discriminations? The remainder of this thesis serves as an analysis of the following explanatory hypotheses:

Non-Reductive Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to counterevidence.

Reliabilist Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process.

Assurance Hypothesis: Our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said.

Evidential Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence.

In this chapter, I focus on the non-reductive hypothesis.

In §II, I examine the non-reductive position concerning justified testimonial belief and testimonial knowledge. In doing so, I survey the work of Reid (1764), Burge (1993), and Fricker (1994) with the aim of ascertaining the conditions under which testimonial acceptance is, and is not, justified. Doing so helps us to understand how the non-reductive hypothesis receives support. In §III, I focus specifically on the conditions which *defeat* justified testimonial acceptance. For non-reductionists, we have some sort of *epistemic right* or *entitlement* to accept testimony without any need for supplementary premises. However, this entitlement can be *defeated*. In this section, I focus on those defeating conditions, arguing that non-reductionists should maintain that *normative defeaters* can override the entitlement to accept a piece of testimony. Next, in §IV, I focus more specifically on the non-reductive hypothesis and introduce my redundancy argument.

I argue that there is no difference in principle between evidence and counterevidence. Depending on the claim it is put forward to support or undermine, a proposition, *p*, can be evidence or counterevidence. But if there is not a difference in principle between evidence and counterevidence, then the position afforded to us by the non-reductive hypothesis is also offered by the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. As such, we see that the non-reductive hypothesis is redundant. The upshot is that we can eliminate the non-reductive hypothesis from the pool of hypotheses under consideration. This leaves us with the reliabilist hypothesis, the assurance hypothesis, and the evidential hypothesis.

II

Non-reductionists, contrary to *reductionists* (e.g., Hume¹⁷ (1748/1999: 77-78), (Fricker¹⁸ (1994) and (2017)), maintain that that which justifies testimonial acceptance is *unique* and *irreducible*. Whatever justifies testimonial acceptance does not reduce “to other, allegedly more basic kinds of epistemic justification” (Goldberg, 2007: 144). In other words, what justifies an audience’s acceptance of a speaker’s testimony that *p* does not depend on, or reduce to, some reason the audience could possibly cite in support of *p* (Faulkner, 2011: 79). Instead, the non-reductionist forwards, much like we are justified in forming perceptual beliefs by taking appearances at face value, so too are we justified in accepting the word of others, *all else being equal* (Burge, 1993: 285) (Goldberg, 2007: 148). The result is a position which sees justified testimonial acceptance as the “default position” (Burge, 1993: 281). Appealing to an example will assist us here.

Consider the following case: I look at my desk and see that there is a red book on it. Intuitively, I am justified in believing “that there is a red book on my desk” just by taking the appearance of the red book on my desk as representative of how the

¹⁷ For Hume, the justification we have for accepting a piece of testimony comes from “no other principle than our observation of the veracity of human testimony, and of the usual conformity of facts to the reports of witnesses” (1748/1999: 77). In short, it has an inductive basis and so, that which justifies testimonial acceptance can be seen as *reducible* to that which justifies induction. Further, on Hume’s reductive position, to be justified in accepting a speaker’s testimony, an audience must have *positive reasons* to do so (Lackey, 2008: 144). The reason or reasons in question for accepting a given piece of testimony comes from past observations of the truth of testimony of a *particular kind* (Hume, 1748/1999: 78) (Faulkner, 1997: 304-305). The inference made by the audience is then *inductive*. The audience reasons that, since past reports of this particular kind have been veridical so is this particular report veridical (ibid.).

¹⁸ On Fricker’s model, “[t]he epistemic means a recipient [of testimony] has available to assess a speaker for trustworthiness is to construct a mini-psychological interpretation of him. One should trust what he tells one just if one’s best explanation of his utterance, as afforded by this mini-theory of his psychology, entails the fact that what he states is true” (2017: 10). Fricker’s “local reductionist explanationist” account argues that we build a mini-psychology-theory about our interlocuter and then determine whether to believe the offered report by way of an *inference to the best explanation*. As such, Fricker’s proposal sees the justification for testimonial acceptance as reducible to that of inference to the best explanation.

world really is. I do not need to employ any supplementary premises to support the truth of the claim that there is a red book on my desk. I do not need to provide some argument in support of, say, the accuracy of my vision. Rather, my perceptual experience of things appearing to me thus-and-so seems enough to justify the formation of a corresponding perceptual belief. However, appearances alone are not *always* enough to bring about justified belief.

Suppose that the room in which my desk is situated is illuminated by a red lightbulb. Suppose further that I know this. In such a case, I have good reason to reject the proposition *that there is a red book on my desk* because of what I know. In this case, appearances alone fail to furnish me with justified belief since, despite it appearing to me that there is a red book on my desk, I have good reason to believe that the way things appear might fail to be representative of how the world really is.

From the examples discussed above, we might suggest that, with respect to perceptual belief, there is a *default position* of justified belief. That is, by default, one is justified in taking things to be as they appear to be. This explains the intuition that, in the scenario *without* the red lightbulb, I am justified in believing that there is a red book on my desk. As we have seen, however, that default position can be overridden. The scenario in which I know that my room is being illuminated by a red lightbulb is such a case. In that scenario, I cannot gain a justified belief by taking appearances as a true representation of how the world is, given my knowledge of the lighting conditions.

The notion of a default position of justified belief echoes Pollock's concept of *prima facie justification*. He writes that:

"If a belief is *prima facie* justified, one does not need a reason for believing it, but it may be possible to have a reason for disbelieving it. For example, one might urge that my belief that something looks green to me is *prima facie* justified (for me) because (i) in the absence of any

reasons for thinking it does not look green to me it would be justified, but (2) it might be possible for me to have reasons for thinking that it does not look green to me which would make the belief unjustified" (1971: 293).

The idea Pollock is forwarding here is that, with respect to perception, one is justified in believing that *p* when one does *not* have reason to doubt that *p*. In other words, the *default position* is one of justified belief. This default position can be overridden, however¹⁹. Indeed, this is exactly what happens when one has a reason for thinking that that which appears to one a certain way might not be so. For example, that the red book atop one's desk might not be red. The non-reductive position is that testimonial acceptance works similarly in this respect²⁰.

Non-reductionists maintain that, by and large, an audience does not need any supplementary premises to be justified in accepting a speaker's testimony (Reid, 1764/1993: 193-194), (Coady, 1992: 145), (Burge, 1993: 281), (Weiner, 2003: 257), (Goldberg, 2020: 103). Rather, like we see in the case of perception, agents are justified in taking "appearances" to be an accurate representation of how the world really is. Applied to the testimonial domain, we can say that agents are justified, *by default*, in accepting what they are told by speakers. This is not to say that agents are *always* justified in accepting what they are told. Just as we saw in the case of perception, that default position can be overridden. So, provided that one

¹⁹ One might question my use of the term "default position" given that I reference Pollock. Why not just use Pollock's language of *prima facie justification*? This is due to the term "default position" being far more common in the literature on testimony. Given this, I opt for the term "default position" to provide a consistent discussion throughout this chapter.

²⁰ It is worth noting, however, that there is an important distinction between justified testimonial *acceptance* and justified testimonial *belief*. This is because, one can hold that what justifies an agent testimonial acceptance is not identical to that which justifies an agent's testimonial belief, insofar as the latter has something to do with the speaker. The thought underlying this is that justified belief, or knowledge, is *transmitted* from the speaker to the audience. On this point, Faulkner writes that "the idea that testimony transmits warrant is the idea that there is a way of being warranted in belief through testimony that necessitates that a prior speaker in the chain of testimony was warranted in belief" (2011: 62). To be clear, my focus is with testimonial acceptance. There is an ongoing debate concerning transmission and it is not within the scope of this project to delve into this (see Welbourne (1994), Faulkner, (2000) (2011), Keren (2007), Lackey (2008), Wright (2014) (2015)).

does not have reason to believe otherwise, receiving testimony that p (and comprehending it) permits the audience to justifiably accept the speaker's testimony that p ²¹.

The possession of reason to the contrary can override an audience's default position of justified acceptance. This can occur in cases wherein an audience knows that p and is told that $\sim p$. Suppose, for example, that an audience, A , who is a resident of York, is told by a speaker, S , that *York was affected by an earthquake yesterday*. A knows that this is not true as she was in York yesterday. Further, she has not seen anything in the local news about any earthquake. Given this, A has very good reason to reject S 's testimony. She is therefore moved out of the default position of justified acceptance. There are, however, some cases in which one can gain a justified belief that $\sim p$ despite believing, or knowing, that p .

Let's suppose I *know* that the shop closes at 5pm. Suppose further that you tell me it closes at 4pm and I regard you as a highly trustworthy testifier. Your report *can* cause me to doubt my belief because I regard you as so trustworthy. I might gain a justified belief from your testimony despite having knowledge because you might cause me to reassess the truth of my belief when you offer your report. In this case, I know that p and am told that $\sim p$ but can gain a justified belief that $\sim p$. As such, it is not the case that the possession of reason to the contrary of p will *always* prevent the attainment of justified belief that $\sim p$.

If the non-reductionists are right, then justified acceptance is the default position in the testimonial domain. This means that, so long as I do not have reason to believe otherwise, I can accept what the speaker says and attain a justified belief. To gain a better understanding of the non-reductive position, and to appreciate how the non-reductive hypothesis receives support, it will be beneficial to assess

²¹ I take it as necessary for the audience's formation of the belief that p from a speaker's testimony that p that the audience successfully comprehends the speaker's testimony. When I speak of an audience receiving testimony, it should be read as also involving the successful comprehension of that testimony unless otherwise stated.

some non-reductive accounts concerning justified testimonial acceptance. I will begin by looking at the work of Reid (1764).

In proposing his non-reductive theory, Reid suggests that we would believe little to nothing if we always assessed testimony prior to accepting it. He writes:

“It is evident that, in the matter of testimony, the balance of human judgment is by nature inclined to the side of belief; and turns to that side of itself, when there is nothing put into the opposite scale. If it was not so, no proposition that is uttered in discourse would be believed, until it was examined and tried by reason; and most men would be unable to find reasons for believing the thousandth part of what is told them” (Reid, 1764/1993: 194).

For Reid, our default response to testimony is to accept it. Rejection is the response when something is put “into the opposite scale”; that is, when we have reason to believe that the speaker’s report is not veridical (ibid.). Reid states further that, if we assessed reports for truthfulness prior to accepting anything, we would accept very little, given the apparent difficulty in finding reasons for belief.

Now, according to Reid, we are disposed to accept rather than reject the testimony of others. Further, he maintains that, if we were to accept testimony *only* after examining it, then we would find it very difficult to accept much at all. Add to this Reid’s claim that we come to know a great deal from testimony²² (ibid. 190), and we can conclude that it cannot be the case that our default response is to examine testimony for truthfulness prior to acceptance. We can conclude further that accepting testimony without examining it for truthfulness proves no obstacle to the attainment of testimonial knowledge or justified testimonial belief.

²² Such a claim can be identified in Reid’s suggestion that, “[t]he objects of human knowledge are innumerable, but the channels by which it is conveyed to the mind are few. Among these, the perception of external things by our senses, and the informations which we receive upon human testimony, are not the least considerable” (ibid. 190).

But one might accuse Reid of being too focused on the descriptive issue here. Whatever our actual practices may be, one might argue, if justified testimonial belief necessitates supporting evidence, then accepting whatever we are told without evidence will not allow for justified testimonial belief. Reid's introduction of *the principle of veracity* and *the principle of credulity* help to overcome this concern:

"The wise and beneficent Author of Nature, who intended that we should be social creatures, and that we should receive the greatest and most important part of our knowledge by the information of others, hath, [...], implanted in our natures two principles that tally with each other.

The first of these principles is, a propensity to speak truth, and to use the signs of language, so as to convey our real sentiments.

[...]

Another original principle implanted in us by the Supreme Being, is a disposition to confide in the veracity of others, and to believe what they tell us. This is the counter-part to the former; and as that may be called *the principle of veracity*, we shall, for want of a more proper name, call this *the principle of credulity*" (ibid. 193-194).

In short, the principle of veracity holds that we are disposed to report that which we sincerely believe to be true. The principle of credulity, on the other hand, holds that we are disposed to accept speakers' reports as true.

Wolterstorff notes an issue with Reid's account, however. The issue is that the principle of veracity, so conceived, allows for the possibility of speakers conveying their true beliefs whilst simultaneously offering widespread falsehoods. Wolterstorff suggests that Reid's account requires something more. He writes that:

"Reid's full account of testimony requires that the principle of veracity be understood not just as the disposition to assert only what one believes, but as that disposition combined with some sort of tendency to get it

right. It would be better to call the explanatory principle the principle of *verisimilitude* than the principle of veracity" (2001: 176).

Through adding a *principle of verisimilitude* to Reid's account, we can arrive at a non-reductive account of justified testimonial belief and testimonial knowledge.

Following Reid, we can maintain that, given our dispositions as described by the principles of credulity and verisimilitude, we are *a priori* justified in accepting testimony, all things considered. This comes from the fact that speakers tend to be sources of true beliefs. When one receives testimony from a speaker, there is a good chance that one will form a *true* belief. However, this default position of justified testimonial acceptance can be defeated. We can therefore understand the *a priori* justification to accept testimony as a *defeasible* a priori justification.

Reid argues that, when we are in possession of reason against the veracity of the speaker's report, we tend to reject rather than accept the speaker's report (1764/1993: 194). Here, Reid might be seen as making a *descriptive* claim about our testimonial practices, but this can be ported to allow for the *normative* claim concerning the defeating conditions for the default justification to accept testimony. That default justification is defeated when there is something "put into the opposite scale" (ibid.). Consequently, we can say that there is a kind of requirement on the audience: an audience, *A*, must *not* be in possession of sufficiently strong reason against the veracity of a speaker's, *S*'s, report. As such, we see that, at least for Reid, the non-reductive position requires something of the audience, in a sense.

Burge's more recent contribution explicitly details the conditions under which testimonial acceptance is, and is not, justified. This is revealed in his *Acceptance Principle*:

(AP) - "A person is entitled to accept as true something that is presented as true and that is intelligible him, unless there are stronger reasons not to do so" (Burge, 1993: 281).

Here we see that, rather than justification, Burge talks of *entitlement*. He writes:

“The distinction between justification and entitlement is this: Although both have positive force in rationally supporting a propositional attitude or cognitive practice, and in constituting an epistemic right to it, entitlements are epistemic rights or warrants that need not be understood by or even accessible to the subject. We are entitled to rely, other things equal, on perception, memory, deductive and inductive reasoning, and on - I will claim - the word of others” (ibid. 272-273).

So, whilst Burge makes a distinction between *justification* and *entitlement*, he maintains that they are alike insofar as they can both be regarded as providing rational support to a belief, thereby conferring a positive epistemic status to it. We can label both under the broader term *warrant*. The important difference between them has to do with *cognitive access*.

In allowing that both *justifications* and *entitlements* can grant one warrant to believe that *p*, Burge takes it that the path to knowledge does not necessitate *justified* belief. Rather, it necessitates *warranted* belief. One can be justified in believing that *p*, or one can be entitled to believe that *p*; both constitute kinds of warrant.

In response, one might say that Burge is mistaken since it is very clear that we have justified beliefs from perception. When I see the red book on my desk, I form a *justified belief* that there is a red book on my desk. The disagreement here, I take it, focuses solely on the terminology in play. The issue is that the term “justification” has become so commonplace, that it is hard to do without it. Using “justification” in this commonplace sense, Burge will surely agree that perception allows for justified belief. So, I take it that point that he is making here, and which *requires* a distinction, is that there are two pathways to satisfy the classically understood *justification condition* on knowledge. One path sees an agent attain a positive epistemic status which need not be understood by her or even accessible to her whilst the other seemingly requires this. We might classify both as forms of justification, but Burge

has opted for a distinction in terms. To prevent confusion, during my discussion of Burge's work I will follow his use of the terms "justification", "entitlement", and "warrant". Burge argues that warranted belief is necessary for knowledge. But that condition can be satisfied by an agent having an entitlement to believe that *p* or by being justified in believing that *p*. Let us consider an example to better make sense of this.

For Burge, when I see the red book on my desk, I am *entitled* to believe *that there is a red book on my desk*. I do not need the support of any supplementary premises for the belief *that there is a red book on my desk* to gain a positive epistemic status. Rather, I am "entitled" to take it that appearances accurately reflect reality; nothing extra is needed (ibid. 273). This is not to say that I *can't* be justified in believing *that there is a red book on my desk* though. For Burge, "[j]ustifications, in the narrow sense, involve reasons that people have and have access to" (ibid.). Insofar as one can access one's perceptual experiences, one could, by reflection, use that experience as a reason for belief. At any rate, the important point here is that, for Burge, we have, by default, an *entitlement* to believe that which appears to us thus-and-so.

Contrast the perceptual case above with the case of me believing *that the butler did it* whilst reading a murder mystery. After reading a good portion of the book, I try to piece things together and infer that the best hypothesis is that the butler committed the murder. But this belief, it seems, is only justified when I have evidence in support of it. If I believe *that the butler did it*, there is an expectation that I be able to provide some reason for believing so. If you ask why I believe *that the butler did it* and discover that it is not based on any such reason, then, intuitively, you can regard my belief as unjustified. So, we see a clear difference between two pathways to justified belief: *justifications* require reasons to which the agent has access, or at least, *could* have access; *entitlements*, on the other hand, have no such

requirement. Importantly, Burge is concerned with entitlements in his discussion of testimonial belief.

Burge provides an intricate defence of his claim that we are *a priori* entitled to accept what speakers tell us unless there are stronger reasons not to do so. It is not within the scope of this chapter to do proper service to this nuanced account as my interest is with the conditions that defeat that entitlement and the epistemic requirements put on the audience. Nevertheless, to say nothing would be remiss.

Burge's argument starts from the position that sources of knowledge, such as *perception, memory, and reason* are things on which we are rationally entitled to rely. Testimony, he argues, works similarly. We are rationally entitled to rely on the word of others when it is intelligible and presented as true because, in presenting content intelligibly and as true, we can presume that the source of that content, viz., the testifier, is rational (ibid. 292). Summarising his view, Burge writes:

"We are *a priori* prima facie entitled to accept something that is prima facie intelligible and presented as true. For prima facie intelligible propositional contents prima facie presented as true bear an *a priori* prima facie conceptual relation to a rational source of true presentations-as-true: Intelligible propositional expressions presuppose rational abilities and entitlements; so intelligible presentations-as-true come prima facie backed by a rational source or resource for reason; and both the content of intelligible propositional presentations-as-true and the prima facie rationality of their source indicate a prima facie source of truth" (ibid. 285).

Given that prima facie intelligible presentations-as-true are, per Burge, *a priori* related to a rational source of *true* presentations-as-true, we are thereby entitled to accept what speakers say, *unless there are stronger reasons not to do so*.

We see in Burge's *Acceptance Principle* that, by virtue of us receiving some testimony, which is intelligible, and presented as true, we are thereby entitled to

accept it as true unless there are stronger reasons not to do so (ibid. 281). As such, Burge's account produces the following audience requirement for warranted testimonial acceptance: *there must not be stronger reasons for an audience not to accept the speaker's testimony*. Here, one might ask whether those reasons must be known by the audience or not. Further, one might ask what *kind* of reasons defeat the entitlement for warranted testimonial acceptance. I will return to these important questions in §III, when I focus specifically on "defeaters". I am, in this section, concerned with making clear the conditions under which testimonial acceptance is, and is not, warranted, per non-reductionism. Burge's view affords us the position that an audience has a defeasible entitlement to accept what she is told. This amounts to a default position of warranted testimonial acceptance, all things considered. However, there is an audience condition on warranted testimonial acceptance, viz., there must not be stronger reasons against accepting what a speaker says. To avoid *unwarranted testimonial beliefs*, this means that the audience will have to demonstrate a sensitivity to such reasons. That is, the audience will have to demonstrate a sensitivity to *counterevidence*. If they fail to do so, then they will fail to be sensitive to the conditions under which the entitlement to accept what a speaker says is defeated.

According to Fricker, who argues against non-reductionism, any non-reductive theory must subscribe to the *presumptive right thesis*:

"PR thesis: On any occasion of testimony, the hearer has the epistemic right to assume, without evidence, that the speaker is trustworthy, i.e. that what she says will be true, unless there are special circumstances which defeat this presumption. (Thus she has the epistemic right to believe the speaker's assertion, unless such defeating conditions obtain.)" (1994: 125).

Commenting further, Fricker writes that:

"[The PR thesis] is a normative epistemic principle, amounting to the thesis that a hearer has the epistemic right to believe what she observes an arbitrary speaker to assert, just on the ground that it has been asserted: she need not attempt any assessment of the likelihood that this speaker's assertions about their subject matter will be true, nor modify her disposition to believe according to such an assessment" (ibid. 127).

We have seen in the accounts offered by both Reid and Burge the inclusion of some kind of PR thesis. Both maintain that justified, or warranted, testimonial acceptance does not *necessitate* the possession of any supporting evidence. Put otherwise, to be warranted in accepting a speaker's testimony that *p*, the audience does not need to do any "positive epistemic work", as Lackey puts it (2008: 156). And this is because we have, *by default, or a priori*, a defeasible epistemic right to accept what we are told "just on the ground that it has been asserted" (Fricker, 1994: 127). Moreover, as we have seen, both Reid and Burge grant that this epistemic right is lost when "defeating conditions obtain" (ibid.).

Insofar as the presence of defeating conditions can override an audience's entitlement to accept a speaker's testimony, it is important that we ascertain what those conditions are. One might take it that defeating conditions obtain only when the audience has an awareness of, or an access to, such conditions (Reid, 1764/1993: 194) (Burge, 1993: 297). Alternatively, one might maintain that warranted testimonial acceptance requires that defeating conditions do not obtain, irrespective of what the audience believes (Goldberg, 2007: 159-160). Either way, if one's entitlement to accept testimony is defeated, then testimonial acceptance cannot be warranted *by way of the entitlement, or epistemic right, to accept testimony*²³. As such, the acquisition of testimonial knowledge and warranted

²³ This is not to suggest that the presence of defeating conditions necessarily presents a barrier on the path to knowledge. Rather, it is to suggest that, to know that *p* when there is a presence of defeating conditions regarding the truth of *p*, one will have to take a different route to come to know that *p*. For example, one might have to acquire evidence in support of *p*.

testimonial belief necessitates the *absence* of defeating conditions, however it is they are realised. If such conditions do obtain, then the audience's belief formed via testimony, even if it is true, is not warranted and, consequently, cannot amount to knowledge. To avoid the formation of unwarranted testimonial beliefs, audiences must be sensitive to those defeating conditions. In short, they must be sensitive to *counterevidence*.

Now, insofar as I am interested in generating explanatory hypotheses for the truth-directed testimonial discriminations condition, I am interested in looking at *audience conditions*. I am therefore interested in constraints on the audience, given the non-reductive framework. As such, I will now focus on defeating conditions. I do so with the aim of establishing that any non-reductive account must recognize that *normative defeaters* can defeat the default epistemic right to accept a speaker's testimony. I will also note here that, as I proceed, I will be using the term "justified" as I did in the previous chapter²⁴.

III

The non-reductive position allows for justified testimonial acceptance without the need for supporting reasons, as the 'default' position. But if, as Weiner writes, "there is something fishy about the testimony - either we have specific evidence against its trustworthiness, or general experience tells us that this sort of testimony is not reliable - then we need positive evidence before we are justified in believing what we are told" (2003: 264). Put otherwise, if defeating conditions do *not* obtain, then acceptance is justified, given the default position of justified acceptance. If defeating conditions do obtain, however, then acceptance is not justified. That is,

²⁴ I use the terms "justified" and "justification" throughout the thesis. With the aim of consistency in mind, I judge that constantly switching between "warrant" and "justification" will only lead to confusion. As I have now provided an exposition of Burge's work, it is no longer necessary to employ terms in the manner he does. My use of the term "justified", "justification", etc., should be understood as referring to a positive epistemic status, the kind of which is required for knowledge.

unless we acquire, what Weiner calls, "positive evidence". On this point, Lackey states that:

"The underlying thought here is that so long as there is no available evidence *against* accepting a speaker's report, the hearer has no positive epistemic work to do in order to accept the testimony in question. Otherwise put, non-reductionists maintain that, so as long as there are no relevant undefeated defeaters, hearers can acquire testimonially based knowledge *merely* on the basis of a speaker's testimony" (2008: 156-157).

According to Lackey, the non-reductive position sees the audience as having "no positive epistemic work to do" for justified testimonial acceptance, *all things considered* (ibid.). Given the entitlement to accept testimony, we need not gather and assess evidence to supplement speakers' reports. If, however, an agent is in possession of sufficiently strong reason against the speaker's report, then positive epistemic work *is* required, as noted by Weiner (2003: 264). As such, there is a requirement that agents not be in possession of such reasons, or, as Lackey puts it, there is a requirement that agents not be in possession of any relevant *undefeated defeaters*.

But defeaters come in different shapes and sizes. We can follow Pollock in appreciating a distinction between those defeaters that give us reason to believe that the proposition in question, *p*, is false and those defeaters that attack the connection between *p* and the reason for believing that *p* (Pollock, 1974: 42).

To understand defeaters, however, we need to understand the notion of having a *reason to believe* something. Pollock and Cruz write that:

"A state *M* of a person *S* is a reason for *S* to believe *Q* if and only if it is logically possible for *S* to become justified in believing *Q* by believing it on the basis of being in the state *M*" (1999: 195).

Importantly, for Pollock and Cruz, a reason does not have to be a belief. A perceptual state could count as a reason, for example. So, my reason for believing that there is a red book on my desk is not a *belief* I have about how things appear to me, but rather, the state of it appearing to me that there is a red book on my desk. One sympathetic to the non-reductive position might forward an analogous state with respect to testimony. But instead of having a perceptual experience of p as a reason for believing that p , one would have a “testimonial experience that p ”, which could consist in something like a state of comprehending S 's utterance that p , as a reason for believing that p .

Pollock and Cruz state further that “[r]easons are always reasons for *beliefs*” (ibid.). So, insofar as one has a reason, one has a reason *for believing* some proposition, p . Now, for that reason to be a reason for believing, say, that the train to York is late, it must allow for the logical possibility of an agent becoming justified in believing *that the train to York is late* on the basis of that reason. If it fails to do that, it fails to be a reason for holding that belief.

Let us now turn to the concept of *defeaters*. We can begin by looking at the general definition of defeaters Pollock and Cruz provide:

“If M is a reason for S to believe Q , a state M^* is a defeater for this reason if and only if the combined state consisting of being in both the state M and the state M^* at the same time is not a reason for S to believe Q ” (ibid.).

The idea here is that a defeater is a state (doxastic or non-doxastic) such that, when taken together with a reason for believing that p will *not* yield a reason for believing that p . It is also important to note that, in talking of *defeaters*, we must focus on *defeasible reasons*. This is because, as Pollock and Cruz state, “[d]efeasible reasons are reasons for which there can be defeaters” (ibid. 37). Given that the current discussion is focused on defeaters, we can therefore ignore those reasons for which there *cannot* be defeaters.

As previously stated, following Pollock, we can make a distinction between two kinds of defeater (1974: 42). First there are *rebutting defeaters*. Pollock and Cruz write that:

“If M is a defeasible reason for S to believe Q , M^* is a *rebutting defeater* for this reason if and only if M^* is a defeater (for M as a reason for S to believe Q) and M^* is a reason for S to believe $\sim Q$ ” (1999: 196).

Rebutting defeaters defeat justified belief in the sense that they present us with reason to believe the *negation* of the proposition under consideration. For example, an audience, A , might believe *that the shop closes at 5pm*. Let’s say that A ’s reason for belief comes from a speaker, S , telling A that it closes at 5pm. However, A gains a rebutting defeater for her belief when she walks past that shop and sees a sign stating that they close at 4pm. This is a rebutting defeater insofar as it provides A with reason to believe that the negation of her original belief is true. That is, it gives A reason to believe *that the shop does not close at 5pm* is true.

The second class of defeaters identified by Pollock are *undercutting defeaters*. Pollock and Cruz give two definitions for this class of defeater. The first concerns *doxastic undercutting defeaters*:

“If believing P is a defeasible reason for S to believe Q , M^* is an undercutting defeater for this reason if and only if M^* is a defeater (for believing P as a reason for S to believe Q) and M^* is a reason for S to doubt or deny that P would not be true unless Q were true” (ibid.).

The second concerns *non-doxastic undercutting defeaters*:

“If M is a nondoxastic state that is a defeasible reason for S to believe Q , M^* is an undercutting defeater for this reason if and only if M^* is a defeater (for M as a reason for S to believe Q) and M^* is a reason for S to doubt or deny that he or she would not be in state M unless Q were true” (ibid. 197).

But these definitions are rather cumbersome. Fortunately, Pollock and Cruz suggest that “[t]his can be read more simply as ‘*P* does not guarantee *Q*’” (ibid.). An example will aid us here.

Consider again the case in which *A* believes *that the shop closes at 5pm* because *S* tells her so. Suppose that *A* tells a friend this and they inform her that *S* is unreliable on such matters. Let’s say that *S*’s memory isn’t great when it comes to these sorts of things. In this case, *A* gains an *undercutting defeater*. And this is because, whilst the defeater *A* gains does *not* give her reason to believe that the negation of the proposition *that the shop closes at 5pm* is true, it does give her reason to be suspicious of the connection between her reason for believing that *p*, which comes from *S*’s testimony, and the truth of *p*.

Lackey notes that we can make a further distinction between *psychological* and *normative* defeaters (2008: 44-45). Concerning the former, Lackey writes that:

“A psychological defeater is a doubt or belief that is had by *S* and that indicates that *S*’s belief that *p* is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being *had* by *S*, regardless of their truth-value or epistemic status” (ibid.).

To continue with the example we have been using, let us say that *A* receives testimony from *S that the shop closes at 5pm*. Let’s add that *A* takes *S* to be a chronic liar. In this scenario *A* has a belief about the reliability of *S* as a testifier. As such, *A* has a *psychological defeater* for her reason for believing that *p*. As a result, testimonial acceptance is not justified for *A*, as things stand. Importantly, what defeats the justification for acceptance here is *A*’s belief that *S* is a chronic liar, *irrespective of the truth of the claim that S is a chronic liar*.

A is not necessarily barred from coming to know that *p* because of the defeater she has. *A* might come to learn that *S* is the victim of a cruel rumour and, on the contrary, is a perfectly reliable testifier. In such a scenario, *A*’s psychological

defeater ceases to exist and is therefore no longer blocking justified testimonial acceptance.

Let us now consider the second kind of defeater identified by Lackey, viz., *normative defeaters*. Lackey writes that:

“A normative defeater is a doubt or belief that S ought to have and that indicates that S’s belief that *p* is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being doubts or beliefs that S should have (whether or not S does have them) given the presence of certain available evidence” (ibid. 45).

Let’s say that *A* receives testimony *that the shop closes at 5pm* from *S*. Let’s say further that, on every occasion of *S*’s reporting the closing hours of shops, he has been incorrect. Despite knowing this, *A* accepts *S*’s report that *p*, neglecting his awful track-record of reporting truths. In this case, *S*’s awful-track record of reporting truths is a consideration available to *A* and a consideration which defeats her entitlement to accept his report *by virtue of it being something that A should have, given the evidence available to her*²⁵. This is the case *irrespective of one’s psychological state*. But what is important about this sort of defeater is not simply that it is the sort of thing one *should* have, but the sort of thing *one should bring to bear on the relevant belief*. Whilst it is true that normative defeaters are reasons or doubts which, given the available evidence, one *should* have, it is also the case that normative defeaters are reasons or doubts which one *should* bring to bear. By this I mean that it is a reason or doubt that one ought to *have* and which ought to *affect* one’s belief.

Whilst Lackey omits mention of the truth status of normative defeaters, I take it that we should regard them similarly to psychological defeaters in this respect. Consider a case in which your available evidence suggests that *p* when $\sim p$ is the

²⁵ This is because the conjunction of the defeater with *A*’s reason for belief does not provide a reason for *A* to believe *S*’s report.

case. Let's say that you neglect your evidence and believe that p , nevertheless. You gain a true belief, but you only achieve that through demonstrating an ignorance of your available evidence. What is important for something to be a defeater is that, when taken together with one's reason for believing that p , it does not provide one with a reason for believing that p . As in the case of psychological defeaters, that reason or doubt might lead one away from the truth. To enforce a rule that normative defeaters *must* always lead one to the truth is to be committed to the view that, when an agent has available to them a wealth of misleading, but very convincing, evidence suggesting that p , when $\sim p$ obtains, then they are justified in believing that $\sim p$ despite demonstrating an utter neglect of that evidence in support of p .

Further, insofar as one *ought* to have something as a defeater, it must be the case that one *could* have it as a defeater, given the available evidence. This is in keeping with Kant's claim that "[t]he action to which the 'ought' applies must indeed be possible under natural conditions" (1781/2007: A548/B576). In short, that which one *ought* to do must be something one *can* do. So, if one cannot have something as a defeater then that thing cannot be a normative defeater insofar as it is not something one *ought* to have. What is important, for normative defeaters, is that you *could* have it and *could* respond to it. We might characterise it by saying that a normative defeater is the sort of thing that ought to be a psychological defeater, given the available evidence. If something cannot become a psychological defeater for your believing that p , then it cannot function as a normative defeater for your believing that p . For example, if you cannot understand French, and have no resources available to you at some time, t , to comprehend French, then a speaker's testimony that $\sim p$, delivered in French, cannot function as a psychological defeater or a normative defeater for your believing that p at t . Your inability to comprehend French prevents you from *having* a defeater here. And insofar as you cannot *have*

the speaker's testimony as a defeater, nor *could* have it given your available evidence, it is not the case that you *ought* to have it.

Let us now return to the non-reductive position. The current discussion of defeat allows for the possibility of conflicting non-reductive theories which disagree about the conditions under which the entitlement to accept what we are told is defeated. One might support a non-reductive position according to which any of the defeaters so far discussed can override that default position of justified acceptance. By contrast, one might argue that it is only one sort of defeater which has such a power. Burge might be seen as opting for this latter route.

We can look to Burge's comments from his paper, "Interlocution, Perception, and Memory", with the aim of clarifying his view:

"The [acceptance] principle says that the entitlement holds unless there are stronger reasons (available to the person) that override it. It does not say that the person must know there are no stronger reasons; the individual may lack the concept of a stronger reason, and in any case need not rule out the existence of defeaters in advance. It is enough for the individual's being warranted that there *are* no defeaters; defeaters of the entitlement must be available to him" (1997/2013: 305).

Here, Burge states explicitly that the entitlement to accept testimony is lost only when stronger reasons, *which are available to the agent in question*, override it. If we understand "reasons available to the person" as reasons which are *accessible* to an agent, then Burge can be read as defending the claim that it is only *psychological defeaters* which override the entitlement to accept what we're told.

But I concede that it is not entirely clear that Burge is arguing that psychological defeaters are the only kind of defeater which override the entitlement to accept what speakers report. The ambiguity lies in his use of the term "available". He might mean that the reasons in question are something an agent has and can bring to bear on the belief in question. It might instead mean, however, that the

reasons in question are such that, if an agent inquired well enough, she would acquire them and then be able to bring them to bear. My own position is that normative defeaters should be included in the set of conditions which can defeat the entitlement to accept what we are told. As such, I take it that the advocate of the non-reductive position should defend the claim that normative defeaters, as well as psychological defeaters, can override the default position of justified testimonial acceptance.

I argue that if one allows that *only* psychological defeaters override the entitlement to accept speakers' reports, one is committed to holding that intuitive cases of unjustified belief are not so. And if one permits that such cases are *not* cases of unjustified belief then one is committed to taking it that they are cases in which a subject could attain knowledge. I aim to show that, to overcome this issue, we must maintain that the defeating conditions for our default justification to accept testimony includes normative, as well as psychological, defeaters. To show this, I will revisit a case introduced in the previous chapter.

Consider again the case of *Continued Acceptance*. Recall that this is a revised version of Aesop's fable of *The Boy Who Cried Wolf*. In the original fable, the villagers accept the boy's first two deceptive cries of wolf but reject his third cry, judging that he is attempting to fool them again. In *Continued Acceptance*, the villagers accept 49 deceptive cries of wolf. In such a scenario, how are we to assess the justificatory status of the villagers' beliefs *that there is a wolf in the flock* resulting from a 50th cry of wolf? Surely, we are to say that the villagers do *not* attain justified beliefs here. Intuitively, this is because they are being *gullible* or *epistemically reckless*. They have available to them evidence which strongly supports the claim that the boy is not reporting the truth. The issue is that, in accepting the boy's testimony, the villagers neglect such evidence. They *should* be sensitive to the awful track-record of the boy's reporting truths and, consequently, reject what he says.

Now, let us start from the position that psychological defeaters are the only sort of defeaters which can override the entitlement to accept what a speaker says. According to this position, since the villagers do not *have* reason to believe that the boy is delivering false testimony nor do they *have* doubts about his sincerity or competence, their entitlement to accept what he says is intact. That is, they are justified in accepting what the boy says by virtue of the *a priori epistemic right* to accept what speakers say, and, consequently, they form justified testimonial beliefs. Intuitively, however, this is not right. Surely, the villagers cannot gain justified beliefs from the boy's 50th cry of wolf, be it deceptive or veridical. The current position, I maintain, fails to supply us with the tools we need to handle this problem.

As stated in the previous chapter, we can also take issue with the epistemic conduct of the villagers. We can take issue with them *believing* other than they should, but we can take issue with them *neglecting considerations which could be advanced in favour of the truth, or falsity, of the boy's testimony*. The first issue is that, rather than *believe* the boy's testimony, they should *disbelieve* it. In short, they go wrong by adopting a doxastic state which is at odds with their available evidence. The second issue concerns their neglect of the boy's track-record of reporting truths. Recall that, insofar as they demonstrate such a neglect, we can regard them as *doxastically irresponsible*. But it is difficult to make sense of this ascription unless we regard them as doing something other than they should.

I suggest that we can overcome this issue by maintaining that the reasons which defeat their entitlement to accept what the boy says are reasons *which the villagers ought to have*. In taking this route, we allow that *normative defeaters* can override the entitlement to accept testimony. This solves the issue insofar as it means that the villagers do not gain justified beliefs from accepting what the boy says. Further, it gives us a way to make sense of the intuition that they are being neglectful. They are not bringing to bear available considerations to which they *ought* to respond.

Non-reductionists must allow that both normative and psychological defeaters can override the entitlement to accept a speaker's testimony. However, I do not aim to provide any specific formulation here. No doubt, non-reductionists will disagree on the details here. As an approximation, I will simply state that the non-reductionist must enforce as an audience condition on justified testimonial acceptance that there must not be sufficiently strong counterevidence for accepting a speaker's testimony. This includes both *normative* and *psychological* defeaters. As such, an audience might *have* a reason or doubt which functions as a defeater (irrespective of its truth). Alternatively, there might be a reason or doubt *available* to an audience, given her evidence, to which she ought to respond (irrespective of its truth).

But this requirement by itself is not enough. To see this, imagine the individual in the rare position of having no defeaters for their beliefs. They accept what they are told by numerous speakers and, since their entitlement to accept testimony is not overridden, they gain many justified beliefs in doing so. According to the view so far discussed, all is well, epistemically speaking; this agent is doing what they ought to do. However, let us imagine that this individual is also incredibly negligent. They are so negligent that, even if they *did* have defeaters for their beliefs, they would pay no attention to them. As such, in a nearby possible world, this individual *does* have defeaters for their beliefs and, consequently, forms *unjustified* beliefs which fall short of knowledge. Now, if all that is required of the audience is a sensitivity to counterevidence, then this individual is doing nothing wrong. This is clearly an error and requires revision.

Fricker raises this concern when she places a demand on the audience such that "if there were signs of untrustworthiness, she would register them, and respond appropriately" (Fricker 1994: 150). Appreciating this concern, Goldberg and Henderson respond that the non-reductive position can account for this requirement for a "counterfactual sensitivity" to the presence of defeaters (2006:

610). Such a requirement might be familiar from the previous chapter, in which I discussed the modal element of doxastic responsibility, remarking that doxastic responsibility requires one to respond to considerations which could be advanced in favour of the truth or falsity of *p*, *and be disposed to do so*. In the case above, this counterfactual sensitivity is what is missing. As such, we should amend the non-reductive position to accommodate this requirement.

The need for this counterfactual sensitivity might equally be understood as a cognitive system for monitoring defeaters, or counterevidence, being “online”. Consider the following analogy: for a smoke alarm to be considered *properly functioning*, it must be such that *if there was a fire, the alarm would sound*. The broken smoke alarm which produces no sound when there is no fire generates the same output as the fully functioning smoke alarm under the same environmental conditions. Whilst the output is identical to that of the fully functioning smoke alarm when there is no fire to detect, it is obviously the case that the broken smoke alarm fails in some respect. The respect in which it fails is that, if there was a fire, the smoke alarm would *not* sound. In the nearby possible world in which there is a fire, it makes no sound. This thought should carry over to the testimonial domain. So, the non-reductive position must require agents to respond to defeaters. But an agent can only respond to defeaters if she is *disposed* to do so. That is, she must be disposed such that if there *were* defeaters, she would respond to them. Such responsiveness to defeaters requires a *capacity* to respond to defeaters though. Without such a capacity, an agent could not demonstrate any responsiveness. We see then that the non-reductive hypothesis position requires agents to be properly disposed to respond to defeaters, or *counterevidence*. Without any such capacity, or disposition, agents would fail to be sensitive to the presence of defeating conditions and thus, come to form many unjustified testimonial beliefs.

Having now discussed the concept of defeaters and argued that the non-reductionist should maintain that normative, as well as psychological, defeaters can

override an audience's entitlement to accept what she is told, we can turn our attention directly to the non-reductive hypothesis.

IV

The non-reductive hypothesis states *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to counterevidence*. I now aim to show that: 1) this is an accurate depiction of the non-reductive answer to the question of what explains the truth-directedness of our testimonial discriminations, and 2) that this fails to provide a distinct answer to this question, instead collapsing into the evidential hypothesis.

In the previous section, I argued that non-reductionists must include an audience condition on justified testimonial acceptance, viz., that there must not be sufficiently strong counterevidence for accepting a speaker's testimony. Given this, the non-reductionist position is concerned not with considerations in *support* of a speaker's testimony, but considerations *against* a speaker's testimony. One might say that, rather than *evidence*, the non-reductionist is interested in *counterevidence*.

The advocate of the non-reductive position defends the thought that we have some *a priori* epistemic right, or entitlement, to accept a speaker's testimony *without any supplementary premises*. When defeating conditions obtain, however, that entitlement is overridden. In ordinary cases, as Lackey notes, no "positive epistemic work" is required by the audience (2008: 156-157). At the level of testimonial acceptance, this means that, when all goes well in a testimonial exchange, the audience has no positive epistemic work to do. In the successful testimonial exchange in which a speaker, *S*, reports truthfully that *p*, the audience, *A*, accepts that *p* and comes to form a justified belief that *p*. *A*'s coming to have a justified belief that *p* is the result not of any work *A* has put in, but rather, the entitlement *A* has, by default, to accept testimony. When *A* has reason, or doubt, or *should* have reason, or doubt, against *S*'s testimony, then her entitlement is overridden, and she cannot

form a justified belief that p on that occasion. This implies that epistemic work is only necessary for the audience, at the level of testimonial acceptance, when the audience has, or should have, *counterevidence*.

We saw in the previous section, however, that agents must demonstrate a counterfactual sensitivity to counterevidence. That is, they must be disposed such that, if there *were* reasons against accepting the speaker's testimony available, given their evidence, they would respond to them. Failure to do so would result in the formation, or sustainment, of many unjustified beliefs. So, the non-reductive position that audiences have no epistemic work to do to gain justified testimonial belief and testimonial knowledge must accommodate the demand for a counterfactual sensitivity to counterevidence.

As the truth-directed testimonial discriminations condition is a condition on the audience, we must look to the conditions that the non-reductive theorist puts on the audience. Given the requirement that agents respond to defeaters for the sake of avoiding unjustified belief, we can try to explain the truth-directedness of our testimonial discriminations by pointing to a sensitivity to *counterevidence*. This is motivated by the thought that, since the default position, in the testimonial domain, is one of justified acceptance, a given audience does not need to carry out any positive epistemic work. This carries over to the level of testimonial discriminations which means that, discriminations are made on the basis of counterevidence. The thought is that when there is no counterevidence, testimonial acceptance is permissible and not otherwise.

Consider again the analogy with the smoke alarm. When a properly functioning smoke alarm sounds, it alerts you as to the presence of smoke. When it does not sound, however, one has no reason to believe that there is a fire in the building. Analogously, when one has counterevidence (or should have), this indicates that one must *not* accept the incoming testimony. As noted previously, however, one's responding to counterevidence necessitates one's being disposed

to respond to defeaters. The smoke alarm, for example, needs to operate such that, if the right conditions obtained, it would give the right output. This requires that agents are, like the properly functioning smoke alarm, prepared to deal with defeaters *should they arise*. There is a sense in which agents need to be on the lookout for counterevidence. Though I will state explicitly here, this makes no demand on the agent that they must demonstrate that sensitivity *consciously*.

Notice further that, with the current proposal, we still require something like *a principle of verisimilitude*. That is, we still need some link between speakers' reports and the truth of those reports which allows audiences to presume that veridical testimony is the norm. I will not delve into this matter but will rather assume for the sake of argument that we can satisfy such a condition unproblematically (see Burge (1993)).

Having articulated the non-reductive hypothesis. I now aim to show how it fails to offer a view distinct from the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to, and so are based on, evidence*. This stems from there being no difference, in principle, between evidence and counterevidence.

In the case of rebutting defeaters, a sensitivity to counterevidence necessitates a sensitivity to evidence. Consider the following case:

Closing Time: Katya asks her friend Fatima what time the shop closes and is told that it closes at 5pm. As a result, Katya forms the belief *that the shop closes at 5pm*. She gets to the shop at 3pm and chats with the owner, Audrey who mentions that the shop closes at 4pm today. As such, Katya revises her belief that the shop closes at 5pm and instead believes that it closes at 4pm. Whilst out for a stroll at 4:30pm, Katya walks past the same shop and notices that it is still open. She goes inside the shop and chats with Audrey, asking what time they close today. Audrey says that she thought that it was Sunday, and on Sunday they close at 4pm.

But it is Saturday today, and on Saturdays they close at 5pm. As a result, Katya revises her belief that the shop closes at 4pm and forms the belief that it closes at 5pm today.

In this case, Katya receives testimony from Fatima *that the shop closes at 5pm* (p), thus acquiring reason to believe that p . She then acquires reason to believe that $\sim p$ from Audrey's *first* report. Later, she has reason to believe that p (or that $\sim\sim p$) from Audrey's *second* report, in which Audrey corrects herself.

Initially, we can explain Katya's justified belief that p from Fatima's testimony that p by commenting that, since defeating conditions do not obtain, she is not moved out of the default position of justified acceptance. When she receives testimony from Audrey, however, she gains a *rebutting defeater* insofar as she acquires a reason to believe that the negation of Fatima's testimony is true. That is, she gains reason to believe that $\sim p$. Later, Katya gains a further rebutting defeater from Audrey's second report insofar as she acquires a reason to believe that the negation of Audrey's first report is true. In other words, she gains reason to believe that $\sim\sim p$, or put otherwise, p . In gaining this further rebutting defeater, Katya gains a defeater for the defeater she gained from Audrey's first report. In short, she gains a *defeater-defeater*. And, as we see here, the defeater-defeater she gains provides her with positive reason to believe that p . Moreover, without that defeater-defeater, Katya would not be able to attain the justified belief that p , given that her entitlement was overridden due to her original defeater. This is important because it has a bearing on the way we think about evidence and counterevidence.

The existence of the evidential hypothesis and the non-reductive hypothesis implies a distinction between evidence and counterevidence. On one side, we have evidence which consists in considerations *supporting* p , and on the other side, we have counterevidence which consists in considerations *against* p . As we have seen, the non-reductive hypothesis explains the truth-directedness of our testimonial discriminations by suggesting that we demonstrate a sensitivity to counterevidence.

The issue with this is that it is not so easy to demonstrate a sensitivity to counterevidence *alone* due to the contextual nature of evidence.

Given that the default position of justified acceptance is overridden when there is sufficiently strong counterevidence available to an audience (meaning that she either *has* that counterevidence, or *should* have it), we can rephrase the *Closing Time* case to use similar language. We can say that when Katya receives Audrey's first report, she receives *counterevidence* for *p*. When she receives Audrey's second report, however, she receives *counter-counterevidence* for *p*. But counter-counterevidence for *p* functions just the same as evidence for *p*, *in this case*. Consider, Katya's counter-counterevidence comes in the form of Audrey's testimony *that the shop closes at 5pm today*. But this is as much a positive reason to believe that the shop closes at 5pm as it is counterevidence for the claim "that the shop closes at 4pm today". Perhaps the purported distinction between evidence and counterevidence is not so clear then.

Whenever we speak of evidence, or counterevidence, we must specify the proposition for which something purports to be evidence or counterevidence. Audrey's second report is evidence in support of the claim *that the shop closes at 5pm today* and counterevidence for the claim *that the shop closes at 4pm today*. Notice that *both* of Audrey's reports are *evidence* for the claim *that the shop will close today*. And *both* of Audrey's reports are *counterevidence* for the claim *that the shop never closes*. What this tells us is that determining the status of something as a piece of evidence or counterevidence is a *contextual* matter. It depends on the proposition for which the candidate-evidence is being used to support. Put more concisely, that which is evidence can also be counterevidence, depending on the context. This is bad news for the non-reductive hypothesis because, if there is an ambiguity with respect to the term "counterevidence", then there is an ambiguity with respect to that thing to which agents demonstrate a sensitivity.

The problem becomes exacerbated when one appreciates that the status of something as evidence or counterevidence can also be affected by an agent's background beliefs. If, for example, one believes that there is a global conspiracy to protect government officials, all of whom belong to a clandestine criminal syndicate, then an instance of a government official being cleared of any wrongdoing by an independent report might act as evidence for the claim that there is a global conspiracy to protect government officials. However, for the individual who has no such background beliefs, reading an independent report clearing a government official of any wrongdoing would act as counterevidence for the claim that there is a global conspiracy to protect government officials. There wouldn't have been an independent report otherwise. This reinforces the idea that there is no difference, in principle, between evidence and counterevidence.

Notice that the argument here does not affect the non-reductive position with respect to justified testimonial acceptance. On the non-reductive view, one can still answer the question of what justifies testimonial acceptance by pointing to an *a priori* epistemic right, or entitlement, to accept testimony, all things considered. What I have argued, following Goldberg and Henderson (2006), is that this position necessitates a counterfactual sensitivity to counterevidence. Given the contextual nature of counterevidence, we can also regard this as a counterfactual sensitivity to *evidence*. It is this requirement on the audience which motivates the non-reductive hypothesis.

The non-reductive hypothesis states that *our testimonial discriminations are truth-directed because they demonstrate a sensitivity to counterevidence*. In short, it tries to account for the truth-directedness of our testimonial discriminations by suggesting that, in demonstrating a sensitivity, and being *disposed* to demonstrate a sensitivity, to counterevidence, the manner in which we discriminate between bits of testimony is *truth-directed*. But if there is no difference in principle between counterevidence and evidence, then the explanation offered by the non-reductive

hypothesis agrees entirely with the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. And insofar as they agree entirely, we see that the non-reductive hypothesis becomes redundant. Those who are sympathetic to the non-reductive position should therefore be sympathetic to the evidential hypothesis²⁶.

V

In this chapter, I have focused on the non-reductive hypothesis, arguing that it is redundant. First, I surveyed the work of Reid (1764), Burge (1993), and Fricker (1994) with the aim of establishing the conditions under which testimonial acceptance is justified, per non-reductionism. I found that, though there is of course disagreement between non-reductive theorists, there is agreement in there being some sort of *epistemic right or entitlement* to accept a speaker's testimony *in the absence of defeating conditions*. From here, I introduced the work of Pollock (1974) and Pollock and Cruz (1999), examining further the conditions under which an audience's entitlement to accept a speaker's testimony is defeated.

In detailing the defeating conditions for justified testimonial acceptance, I recognized that defeaters come in different shapes and sizes. Undercutting defeaters contrast with rebutting defeaters (Pollock and Cruz, 1999: 195-197), and normative defeaters contrast with psychological defeaters (Lackey, 2008: 44-45). After clarifying these terms, I argued that the non-reductive theorist must defend an account which allows *normative defeaters*, as well as *psychological defeaters*, to override an audience's entitlement to accept testimony. This means that, if there is a defeater available to an agent - which they have, or to which they ought to respond - then the entitlement to accept testimony is defeated. This is irrespective of their

²⁶ Here, I recognize that if one is sympathetic to a *reliabilist* non-reductive position, then one may be far from sympathetic to the evidential hypothesis. The reliabilist hypothesis will be the focus of my next chapter.

response to such a reason and of the truth-status of that reason. The audience might neglect such a reason, but in so doing, fail to attain justified belief.

Next, I introduced the work of Goldberg and Henderson, who stipulate that the non-reductive position requires agents to demonstrate a “counterfactual sensitivity” to the presence of defeaters (2006: 610). Without being disposed such that if there *were* defeaters, one would respond to them, an agent can be regarded as going wrong epistemically. Given that it is on the basis of counterevidence that one’s entitlement to accept what a speaker says is lost, a failure to demonstrate a sensitivity to it allows for the widespread formation, and sustainment, of unjustified belief.

Having established this, I gave a brief defence of my presentation of the non-reductive hypothesis and proceeded to show how the non-reductive hypothesis is redundant. The issue revolves around the contextual nature of evidence. That which is evidence can also be counterevidence, depending on the context. For example, the background beliefs of one individual, A_1 , may suggest a candidate for evidence supports some proposition, p , whilst that same candidate for evidence may, when taken in conjunction with the background beliefs of another individual, A_2 , do the very opposite. What this tells us is that drawing a clear line between evidence and counterevidence is not possible. The issue is that the non-reductive hypothesis, contrasted with the evidential hypothesis, implies a clear distinction between evidence and counterevidence; that demonstrating a sensitivity to counterevidence is something different to demonstrating a sensitivity to evidence. I have argued that this is not the case because there is no difference, in principle, between evidence and counterevidence.

Having now eliminated the non-reductive hypothesis, we are left with the following three hypotheses which attempt to account for the truth-directedness of our testimonial discriminations:

Reliabilist Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process.

Assurance Hypothesis: Our testimonial discriminations are truth-directed because, in trusting a speaker, we are sensitive to the truth of what is said.

Evidential Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to, and so are based on, evidence.

In the next chapter, I will focus on *the reliabilist hypothesis*.

3. The Reliabilist Hypothesis

I

In this chapter, I consider the reliabilist hypothesis *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process*. In doing so, I will first examine the process-reliabilist accounts offered by Goldberg (2007), (2010), (2012) and Graham (2010). I will then argue that the process of making truth-directed testimonial discriminations should be understood as a process of *monitoring evidence*. Through appealing to the empirical data gathered from studies into “deception detection”, I aim to bolster this claim, using the shortage of evidence in the experimental setting to explain the findings from the research. I then argue that the reliabilist must agree with the evidential hypothesis given the role they grant evidence. The upshot of this is that the reliabilist hypothesis collapses into the evidential hypothesis.

After my discussion of process-reliabilism, I will examine Lackey’s “dualist” position, arguing that the audience conditions she proposes are more in keeping with the evidential hypothesis than the reliabilist hypothesis (2008: 178). From here, I argue that the reliabilist hypothesis must be eliminated from the pool of explanatory hypotheses currently under examination.

In §II, I provide a brief exposition of the reliabilist position concerning epistemic justification. In doing so, I focus on the *process-reliabilism* of Goldman (1979). Next, I introduce the process-reliabilist accounts of testimony from Graham (2010) and Goldberg (2010), in §III and §IV. I find that both theorists can explain the truth-directedness of our testimonial discriminations by appealing to a reliable *filtering* process, which monitors for *coherence* and, as I will argue, *evidence*.

Next, in §V, I argue against Goldberg and Graham by introducing the empirical data obtained from studies into deception detection. These studies show that we perform poorly when it comes to detecting lies. Whilst this might seem to

provide a serious problem my proposal that our testimonial discriminations are truth-directed, by and large, I argue that this merely gives us a challenge: to ensure that the explanation we provide for the truth-directedness of our testimonial discriminations can accommodate the empirical data. In attempting to accommodate the data, I argue that the reliabilist is forced to give evidence a key role. Insofar as the reliabilist concedes this, they must agree with the evidential hypothesis.

But process-reliabilism is not the only type of reliabilism on offer. In §VI, I introduce Lackey's "dualist" account (2008). Lackey's position deviates dramatically from the accounts offered from Goldberg (2010) and Graham (2010). Rather than positing a reliable cognitive process, Lackey appeals to a *subjunctive reliability speaker condition* (2008: 74). In addition to this, she also posits an audience condition requiring *that the audience has appropriate positive reasons for accepting the speaker's testimony* (2008: 178). The issue I find is that, whilst her account of justified acceptance might allow for a *dualist* perspective, focusing on both the speaker and the audience, an explanation of the truth-directedness of our testimonial discriminations is solely focused with the agent making the discriminations, viz., *the audience*.

The audience conditions to which Lackey appeals in her account of justified acceptance and testimonial knowledge make no reference to reliability. They do, however, make appeal to *defeaters* and *positive reasons*. As such, I argue that Lackey's account supports the evidential hypothesis. Here, however, I make an important note: that the *kind* of evidence to which Lackey appeals is insufficiently strong to account for the *truth-directedness* of our testimonial discriminations. As such, either her account fails insofar as it is unable to accommodate the truth-directed testimonial discriminations condition, or it must be amended to employ a stronger sense of evidence. At any rate, Lackey's account lends no support to the reliabilist hypothesis.

I conclude this chapter by eliminating the reliabilist hypothesis from the pool of candidates, thus decreasing our total to just two: *the assurance hypothesis* and *the evidential hypothesis*.

II

In the previous chapter, I argued against the non-reductive hypothesis. I found that it fails to provide a distinct explanation of the truth-directedness of our testimonial discriminations, instead collapsing into the evidential hypothesis. It does not follow from this, however, that a supporter of the non-reductive position, with respect to justified testimonial belief and testimonial knowledge, must therefore be a supporter of the evidential hypothesis, with respect to truth-directed testimonial discriminations. Instead, one might support the reliabilist hypothesis *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process*. And this is because, it is possible for two non-reductive theorists to agree that testimonial acceptance is justified without the need for evidence in support of a speaker's report but *disagree* over the nature of justification. In the previous chapter, I forwent an explicit discussion of epistemic justification. It is now necessary to provide some detail on this matter.

Reliabilism, broadly speaking, is an approach to epistemology which focuses on *reliable processes / faculties*. A given process / faculty is deemed *reliable* when it has "a 'tendency' to produce beliefs that are true rather than false" (Goldman, 1979/2012: 38). Reliabilism about epistemic justification, which is my focus here, is the view that what justifies beliefs are such processes / faculties. As a result of this, reliabilism is an *externalist* position (in contrast to *internalism*). And this is because, in conceiving of reliability as a tendency to output more true beliefs than false beliefs, one is thereby forced to invoke a notion which is *external* to agents, viz., *truth / truth-conduciveness*. This is external to the agent in the sense that it is something to which an agent has no introspective access. Whilst one might *learn*

that one's perceptual belief-forming processes tend to output true beliefs, one has no access to such data introspectively.

Goldman introduced his *process reliabilism* in an attempt to provide an account of justified belief which avoids any appeal to *epistemic terms*. A definition of justified belief which made reference to, say, *warrant, justification, reasons to believe, knowledge*, would fail to be of much use, given that it would entail a further inquiry into the nature of the included epistemic term or else result in circularity (ibid. 30). After setting this constraint, and rejecting several proposals, Goldman suggests that:

"The justificational status of a belief is a function of the reliability of the process or processes that cause it, where (as a first approximation) reliability consists in the tendency of a process to produce beliefs that are true rather than false" (ibid. 37).

Consider the following example, involving perceptual belief: I see the red book on my desk and form the belief "that book on my desk is red". Why, according to Goldman, is this belief justified? It is justified because the *cognitive process* that caused me to form that belief tends to output true beliefs. Put otherwise:

"If S's believing p at t results from a reliable cognitive belief-forming process (or set of processes), then S's belief in p at t is justified" (ibid. 40)".

Since perceptual processes tend to output true beliefs - at least when there is no impairment to the individual's means of receiving and recording perceptual data - we are justified in believing what the perceptual belief-forming process outputs.

Here, it is important to appreciate that the tendency of a given process to output mostly true beliefs can be affected by a range of things, such as one's *environment* (Sosa, 1988: 174) (1991: 276). Take vision, for example. Your ability to form true beliefs on the basis of your vision necessitates, amongst other things, good lighting. If your environment is very badly illuminated, then this can have an impact on the reliability of your perceptual belief-forming process. As such, the

reliability of a given process must be *specified to an environment*. Human vision, for example, is specified to certain lighting conditions. In this sense, insofar as reliability is specified to an environment, “justification is relative to environment” (ibid. 144).

It is also worth noting that the reliability of a given process can be affected by an agent’s *perceptual and cognitive capacities*. For example, individuals with colour blindness will have difficulty forming true beliefs about the colour of objects. This is not to say that the perceptual belief-forming processes of such individuals is wholly unreliable. Individuals with colour blindness can tend to form true beliefs about the size and shape of objects. Instead, this is to say that “[d]ifferent possible creatures – indeed different actual humans – have different [perceptual and] cognitive capacities and this fact seems to have important epistemic implications” (Lyons, 2022: 202). So, insofar as humans have different perceptual and cognitive capacities, this can affect the reliability of our respective belief-producing processes.

For Goldman, a distinction can be drawn between *belief-independent processes* and *belief-dependent processes*. A belief-independent process, per Goldman, is a process which has no ‘belief-states’ as inputs for the process in question (1979/2012: 38). Perceptual processes would count therefore, as belief-independent processes, since the input includes not beliefs but either “states of the environment” or “events within or on the surface of the organism, e.g., receptor stimulations” (ibid. 39). Belief-dependent processes, on the other hand, must include at least *some* belief states. An example of this would include the process by which I come to believe *that the butler did it* whilst reading a murder mystery. This is because, the process in question, i.e., the *reasoning process*, must include some beliefs to output a subsequent belief. In this case, the process will take as its input various beliefs I have about the butler, the crime, the murder weapon, and so on. The output is the belief *that the butler did it*. So, some processes necessitate the

inclusion of beliefs as input (belief-dependent processes), whilst others do not (belief-independent processes).

In making this distinction between belief-dependent and belief-independent processes, Goldman makes a further clarification. A belief-dependent process might be functioning normally yet output *false beliefs* if given *false beliefs* as the input. As Goldman states, “[a] reasoning procedure cannot be expected to produce true belief if it is applied to false premises. And memory cannot be expected to yield a true belief if the original belief it attempts to retain is false” (ibid. 40). So, what is required of belief-dependent processes is that they be *conditionally reliable*. That is, that these processes reliably output true beliefs *given true beliefs as inputs*.

We see that, for Goldman, cognitive processes are at the very heart of his account of epistemic justification. But what exactly are they? He writes that a process is:

“[S]omething that generates a mapping from certain states – ‘inputs’ – into other states – ‘outputs’. The outputs in the present case are states of believing this or that proposition at a given moment. On this interpretation, a process is a type as opposed to a token. This is fully appropriate, since it is only types that have statistical properties such as producing truth 80% of the time; and it is precisely such statistical properties that determine the reliability of a process” (ibid. 38-39).

Classing processes as *types*, allows for us to bestow statistical properties onto them. And doing so is essential for process-reliabilism since a given process’s tendency to output true belief is what dictates a belief’s status as justified or otherwise. Consider, a *token* process cannot produce true beliefs 80% of the time. The particular case of me forming the belief *that the butler did it* is a particular instance of a belief formed by the reasoning process. On that particular occasion, the process might yield a true belief or a false belief. But it cannot *on that particular occasion* tend to produce true

beliefs 80% of the time. Rather, the reasoning process, *as a type of belief-forming process*, can be said to produce true beliefs 80% of the time.

But there is an issue here. Types are abstract entities that are not spatio-temporally located. As such, it is difficult to understand how a process, defined as a *type*, could produce beliefs. Tokens meanwhile fail to allow for the possession of statistical properties in the manner Goldman requires, given that they are merely singular events. In response to this concern, Goldman writes that:

“[W]hen we say that a belief is caused by a given process, understood as a functional procedure, we may interpret this to mean that it is caused by the particular inputs to the process (and by the intervening events ‘through which’ the functional procedure carries the inputs into the output) on the occasion in question” (ibid. 39).

Goldman argues that *processes* don’t cause beliefs, strictly speaking. Rather, the *inputs of processes* cause beliefs. Nevertheless, the reliability of processes is essential for justified belief insofar as it is *because of* the process’s tendency to output true beliefs which confers justification to the outputted beliefs.

In sum, process-reliabilism is characterised by the elegantly simple proposal that “[i]f *S*’s believing *p* at *t* results from a reliable cognitive belief-forming process (or set of processes), then *S*’s belief in *p* at *t* is justified” (Goldman, 1979: 40). As we shall see, the reliabilist accounts of Graham (2010) and Goldberg (2010) are in keeping with Goldman’s process-reliabilism. Admittedly, however, this is not the only form of reliabilism on offer. In §VI, I will introduce a subjunctive reliability condition when I discuss Lackey’s “dualist” account.

My aim now is to assess the strength of the reliabilist hypothesis - *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process* - by analysing the accounts offered by Graham (2010), Goldberg (2010), and Lackey (2008). Whilst the accounts of Graham and Goldberg can be grouped together under the heading “process-reliabilist

accounts”, this cannot be said of Lackey’s proposal. As such, my discussion will reflect this. I will first deal with the process-realist accounts offered by Graham and Goldberg, objecting to both on the same grounds. After this, I will turn to Lackey’s account, which appeals to a subjunctive reliability condition. In doing so, I will briefly discuss Nozick’s *sensitivity* condition (1981: 172).

III

Graham argues that we are *prima facie* entitled to form testimonial beliefs, or as he writes “comprehension-based beliefs” since they are formed through a *reliable* process, viz., *comprehension-with-filtering*. For Graham:

“A belief enjoys (a kind of) *prima facie pro tanto* entitlement when based on a normally functioning belief-forming process that has forming and sustaining true beliefs reliably as a function. And comprehension-with-filtering has that function. Beliefs based on comprehension-with-filtering thus enjoy (a kind of) entitlement when our capacity to comprehend and filter functions normally, for our capacity has forming and sustaining true beliefs reliably as a function” (2010: 149-150).

So, there is a reliable process, call it the *comprehension-with-filtering* process. When it is functioning normally, it *reliably* outputs true beliefs. And given the process-realist understanding of epistemic justification (that *S*’s belief that *p* is justified iff it is the product of a cognitive process / faculty which tends to output true beliefs (Goldman, 1979: 40)), agents are thereby justified in accepting testimony thanks to the operation of the reliable comprehension-with-filtering process (assuming its normal functioning).

Graham argues that the “comprehension-with-filtering” system has forming and sustaining true beliefs *reliably* as a function. As a result of this, beliefs produced by this system enjoy a kind of *entitlement*. As discussed in chapter two, if *S* has an *entitlement* to believe that *p*, then *S* has *positive epistemic support* for belief in *p*.

Importantly, entitlements differ to justifications insofar as *entitlements* do not necessitate the subject being able to cite some argument in favour of their belief that *p*. Nevertheless, entitlements can be thought of as a kind of warrant or, justification (Burge, 1993) (1997). Given Graham's appeal to a belief-forming process that forms true beliefs *reliably* as part of its normal functioning, it is understandable why he focuses instead on entitlements.

Notice too that Graham does not speak of *testimonial beliefs*, but *comprehension-based beliefs*. By this he means any belief which is "caused and sustained by taking comprehension states at face value" (2010: 151). And by *comprehension state*, he means the audience's representation of the speaker's utterance (ibid.). He writes that:

"In the ordinary case, a speaker (sender) asserts that *P*. The hearer (receiver) represents *P* as asserted. The representation of the speech act – the hearer's comprehension state – disposes the hearer to believe that *P*. In the default case, comprehension prompts belief in the propositional content of the assertion as represented. Being so moved constitutes taking comprehension at face value" (ibid).

The idea is that, in response to a speaker's reporting that *p*, the audience will represent the speaker's utterance. That representation – the audience's *comprehension state* – disposes the audience to believe that *p*²⁷.

I am interested in answering the question of whether the reliabilist hypothesis *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process / faculty* can be adequately supported. As such, it is not necessary to provide a full analysis of Graham's nuanced account. Before

²⁷ Graham allows that, in addition to belief in the propositional content of the speaker's assertion, the audience's comprehension state can also dispose an audience to believe "simple logical entailment or obvious evidential consequences from what was strictly speaking or meant" (ibid. 174). It may further include *conversational implicatures*. This might take the form of me believing the petrol station is not out of petrol when you tell me that there is a petrol station nearby.

proceeding, however, it is necessary to emphasize some important claims that he defends.

Graham maintains that: 1) “positive epistemic standings are goods, successes, fulfilments, or achievements understood in terms of promoting true belief and avoiding error” (2010: 157); 2) “function fulfilment is a success or good” (ibid. 158); 3) “for any belief-forming system or process whose etiological function is to form true beliefs reliably, *ceteris paribus*, it will form true beliefs reliably in normal conditions when functioning normally” (ibid.). From these key points, Graham concludes that “[a] normally functioning belief-forming process that has forming true beliefs reliably as a function confers entitlement on the beliefs it normally causes and sustains” (ibid. 159). And this is because, if the function of some belief-forming process is to output true beliefs *reliably*, then for that process to function normally is just for it to output true beliefs reliably. If it fulfils its function, then we can say that there is some “success or good”. And given that what success means here is the avoidance of error and the production of true belief (reliably), then the belief-forming process will provide positive epistemic standings *just by functioning normally*. In so doing, the agent with that normally functioning belief-forming process enjoys an entitlement, with respect to the beliefs formed by that process, just because, in functioning normally, it outputs true beliefs reliably and this confers positive epistemic status to those beliefs.

Graham’s thesis is that “comprehension-*with-filtering* has forming true beliefs reliably as a function, *not* comprehension *neat*, comprehension *taken alone*” (ibid. 170). As such, I will forego a detailed look at his discussion concerning *comprehension* and *assertion*. My focus is with his proposed filtering system as it is this, Graham proposes, which accounts for “producing a sufficiently high truth ratio” (ibid.). The reason for my focus on Graham’s proposed filtering system is due to the truth-directed testimonial discriminations condition being an *audience* condition. Given that *assertion* does not concern the audience, but the speaker, it is a

discussion I can avoid. Further, given that Graham himself turns to a proposed filtering system in order to account for a process which forms true beliefs *reliably*, I can likewise turn my attention to his proposed filtering system.

Now, it is no secret that people sometimes offer false reports, intentionally or not. How does a recipient of testimony combat this threat of accepting false testimony? Graham answers that we *filter*: “hearers do not sit idly by as speakers mislead or otherwise misinform. Hearers develop countermeasures. Hearers develop capacities for filtering out reliance on fake information providers” (ibid. 172). Additionally, Graham argues, the fact that hearers do this motivates speakers to avoid reporting falsities (ibid. 173). He writes that “filtering not only dampens the possibility of accepting a false report, it also provides an incentive for speakers not to cheat in the first place. For cheaters often get caught” (ibid.).

One filtering capacity developed as a countermeasure, comments Graham, is the detection of *cognitive conflicts* which manifests as a kind of *coherence checking* (ibid.). If I strongly believe that p and you tell me that $\sim p$, then I will (probably) not abandon my belief and form the belief that $\sim p$ (unless I have good reason to do so). Speaking more broadly of filtering, Graham writes that:

“[F]iltering involves sensitivity to counter-considerations: where there are counter-considerations of a certain sort, acceptance would be less likely. Different filters involve different sensitivities. We obviously do not believe whatever we are told; credulity as default does not imply unalterable gullibility” (ibid. 153).

He writes further that:

“Filtering need not involve explicit awareness – belief or judgement – that counter-considerations are absent. Filtering need not involve reasons and reasoning. The subject need only respond by dampening, if not entirely suspending, the normal force of comprehension” (ibid.).

Rather than gullibly accepting whatever we are told, we filter. That is, we demonstrate a sensitivity to *counter-considerations*. This need not be a conscious process of reasoning that, since such-and-such is the case, the speaker's testimony must be false. Such a process can be running in the background without "explicit awareness".

Graham also proposes that filtering can "block" the normal force of comprehension without ruling out the possibility of coming to believe what we have been told. He uses the following example to illustrate this:

"Consider Nick and Oscar. Nick correctly comprehends Oscar's assertion. But Nick is suspicious of Oscar's credibility. Perhaps Nick antecedently doubts that Oscar will tell the truth. Perhaps Oscar's story does not fit with other things Nick believes. Perhaps Oscar's report contains internal contradictions. Or perhaps Oscar is alarmingly fidgety. If this were the end of the matter, Nick would not believe what Oscar just said. But on reflection Nick thinks Oscar can be relied upon" (ibid.).

This is just to say the filtering system can flag relevant counter-considerations without entailing testimonial rejection. So, whilst testimonial rejection is one possible response from the filtering system identifying the presence of counter-considerations, it is not the *only* response.

Here, we see that Graham echoes Lipton, with respect to this "default-trigger" account (2007). According to Lipton:

"[I]t is natural to give a "default-trigger" model of testimony. According to a simple version of such a model, in most contexts the hearer simply accepts what she is told, without engaging in any conscious evaluation or inference ... There are, however, in the case of testimony diverse conditions that may trigger the hearer to switch from default into evaluative mode, where he pauses to consider whether he should believe what he has been told ... The obvious triggers include cases where what

is claimed contradicts firmly held beliefs, cases of contradictory testimony, and cases where there is reason to worry about incompetence or insincerity" (ibid. 240-241).

Appealing to such a model, Graham allows that the presence of counter-considerations can trigger the "evaluative mode". As such, the filtering system can be regarded as system which detects counter-considerations and can dispose one to reject testimony, to suspend belief on the matter, or it can trigger the evaluative mode, wherein an agent engages in a more reflective deliberation (ibid. 241) (Graham, 2010: 153)

In the case of coherence checking, an agent displays a sensitivity to counter-considerations by checking the extent to which the incoming testimony conflicts with her background beliefs. In doing so, the filtering system might also be viewed as checking for *defeaters*. If I strongly believe that there are two train stations in Canterbury and you tell me that there is only one, then I have reason to reject your testimony insofar as what you tell me conflicts with what I already believe. The idea here is that the filtering system will detect a conflict with a strongly held belief and dispose me towards testimonial rejection.

In addition to the coherence-checking filter, Graham also discusses filters which involve *track record information* and the *manner of presentation*. Appealing to the empirical work of Jaswal (2004), Graham states that we have developed a filter which focuses on the confidence of assertion (2010: 152). The thought here is that the filtering system checks for signs of deception or uncertainty by tracking the "manner of presentation". Things like fidgety behaviour, a less committal tone of voice can raise the filtering system's alarm, increasing the possibility of an outcome other than testimonial acceptance. By contrast, a confident delivery can increase the probability of testimonial acceptance.

The other filter Graham mentions involves track record information. He writes that:

“Young children acquire track record information, either first hand or from others. On the basis of this information, they tend to selectively accept assertions. When children have information that a speaker is reliable, that speaker tends to prompt belief more often than the speakers which children ‘know’ are less reliable” (ibid.).

Citing the work of Koenig, Clement, and Harris (2004), Graham defends the claim that the more information we have in support of the “reliability” of a given speaker, the more likely we are to accept what they say. That is, we are more likely to accept what a speaker says when we have evidence in support of the claim that they have a good track record of reporting truths. The reverse is also true. When we have evidence *against* the claim that a speaker has a good track record of reporting truths, testimonial acceptance is *less* likely. The filtering system will therefore use track record information in outputting a given response.

Graham states that filtering involves a sensitivity to “counter-considerations” (2010: 153). Following my work in the previous chapter, we can revise Graham’s description of filtering so that it involves a sensitivity to *evidence* rather than a sensitivity to *counter-considerations*. And this is because, as I argued, there is no difference, in principle, between evidence and counterevidence. So, a sensitivity to counterevidence (defeaters, counter-considerations) can also be regarded as a sensitivity to evidence, depending on the context. Filtering, I suggest, can therefore be seen as a matter of *monitoring evidence* then, not simply *counterevidence*. Admittedly, this does not grant evidence any *intrinsic* value on Graham’s account. As we have seen, beliefs are justified not due to their “fit” with the evidence, but because they are outputted by a reliable process. Nevertheless, it is important to note that, contrary to what Graham writes, the filtering process can be regarded instead as a process which *monitors evidence* insofar as it is sensitive to a range of considerations, such as track-record information, manner of presentation, and background beliefs that have a bearing on the truth of the received testimony.

For Graham, comprehension-with-filtering is a reliable process whose function is to *reliably* output true beliefs. When it functions normally, and fulfils its function, there is a success. Given that this success is the avoidance of error and the production of true beliefs, it confers a positive epistemic status (in the form of an entitlement to believe) on the outputted beliefs *just because the function of comprehension-with-filtering is to output true beliefs reliably*.

Graham's account most certainly affords us a way to satisfy the truth-directed testimonial discriminations condition. Our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process. In this case, that process is the *comprehension-with-filtering* system. As we have seen, this involves a "sensitivity to counter-considerations" (ibid. 153). So, our truth-directed testimonial discriminations are truth-directed because of this reliable process which performs, amongst other things, coherence-checking.

From here, we can ask further *how* the comprehension-with-filtering process works. I will be brief here, as the manner in which the process functions does not impact my argument. The process can function in one of two ways. It could be a *cognitively inaccessible* process, or it could be a *cognitively accessible* process. If the former, Graham must maintain that we do *not* have access to the comprehension-with-filtering system, and in the case of the latter, he must maintain that we do.

If the comprehension-with-filtering system is a *cognitively inaccessible process*, then, as the name suggests, I am not able to access it. This suggests that the truth-directed testimonial discriminations process must function in a fashion similar to our *facial-recognition system*. Consider, I recognise her face and know that the person in front of me is my mother. I know neither how the system works, nor do I have access to the process itself. I merely have access to the *output* which comes in the form of the recognition that this person in front of me is my mother. Filtering, if it is a cognitively inaccessible process, must work in a similar way. So,

whilst we do not have access to the workings of the truth-directed testimonial discriminations process, we do have access to the output.

If instead the comprehension-with-filtering system is cognitively accessible, then perhaps it functions more like processes of *learning*. Take the case of a novice birdwatcher learning the difference between buzzards and kites. Our novice, whilst interested in birds, can't tell the difference between buzzards and kites. She reads her birdwatching book and carefully notes the identifying features of buzzards and kites. She notes that buzzards are typically smaller than kites. That kites have hooked beaks, whilst buzzards do not. And that buzzards have fanned tails, whereas kites have forked tails. Our novice repeats these key differences to herself again and again. She looks through pictures of buzzards and kites, attempting to identify the features in question. All the while, she is very aware of what is happening. She knows that she is using the information she has just attained in an attempt to accurately discriminate between buzzards and kites. If the testimonial discriminations process is a cognitively accessible process, then it should be more like this than the facial-recognition system.

But here, we must make a note. Whilst a process might be *cognitively accessible*, it does not mean that the agent must be conscious of its operation. Take our novice birdwatcher. After years of correctly identifying the differences between buzzards and kites, she may be able to perform this cognitive task without any conscious awareness. This does not mean that, if asked *how* she was able to correctly identify that the bird she saw was a buzzard, she is no longer able to do so. Rather, it means that a process that may once have been conscious no longer is so. As such, it is important to recognise that a process's being cognitively accessible does not entail that the subject must be conscious of it. The suggestion, therefore, that the truth-directed testimonial discriminations process is *cognitively accessible* does not entail that it is a *conscious process*.

Let us consider an example. Imagine I ask you how much a coffee at the university café costs. You tell me that it costs £2000. On receiving and comprehending your testimony, the filtering system checks for “counter-considerations”. In doing so, it checks whether the received testimony coheres with my relevant background beliefs, it checks the manner of presentation and it checks the track record of the speaker. In doing so, we can regard the process as sensitive to evidence which could be advanced in favour of the truth or falsity of the incoming report. Given my knowledge that coffees cost nowhere near £2000, the filtering process alerts me of an incoherence and pushes me to reject the received testimony²⁸. This process might be seen as outputting the command “Reject!” in cases of the received testimony failing to cohere with the audience’s background beliefs to a strong degree. In other cases, it might be that the filtering process outputs the command “Consider!”. The result of this might be a conscious consideration of the evidence prior to accepting or rejecting the testimony. Or perhaps the process instead simply generates a belief *that the incoming testimony isn’t true / probably isn’t true*, or a feeling of *doubt*. I will not dwell on this point. The precise nature of the process is not my focus here.

The filtering process, so far described, is tasked with performing a scan of *evidential considerations*. We see that Graham’s filtering system involves a coherence-checking process, an assessment of the speaker’s track record and an examination of the “manner of the speaker’s presentation”, where things like fidgety behaviour and the tone of one’s voice are relevant pieces of data. There is a sense, therefore, in which the reliabilist hypothesis agrees with the evidential hypothesis

²⁸ Not all cases of detecting incoherence will result in rejection of the received testimony, however. Sperber et al. note that “[i]n some cases, coherence is more easily restored by distrusting the source, and in others by revisiting some of one’s own background beliefs. Unless one option dominates the competition to the point of inhibiting awareness of the alternatives, it takes a typically conscious decision to resolve the issue” (2010: 375-376). The process might alert us to an incoherence, but it may be that, in some cases, revising one’s background beliefs and accepting the offered testimony is the best option. Alternatively, the evaluative mode might be triggered which will require reflective deliberation by the agent.

that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence because, in being sensitive to such considerations, the filtering process can be seen as being sensitive to *evidence*. I recognise, however, that the advocate of the reliabilist hypothesis explains the truth-directedness of our testimonial discriminations by appealing to the operation of a reliable process, rather than a sensitivity to evidence. But if what explains the truth-directedness of our testimonial discriminations is the operation of a reliable process, which operates by way of a sensitivity to evidence, then the reliabilist hypothesis and evidential hypothesis do not look so different. And this is because both require a sensitivity to evidence for our testimonial discriminations to be truth-directed.

Nevertheless, we see from Graham's work how the reliabilist hypothesis receives support. The filtering system is a reliable process which, when operating as it should, outputs different responses to testimony in such a way as to allow for the making of truth-directed testimonial discriminations. My aim now is to develop this by looking at the work of Goldberg (2007) (2010) (2012).

IV

Goldberg writes that "testimonial belief-formation is a belief-dependent process whose input is the testimony itself" (2010: 58). Whereas the input for, say, the reasoning process involves only the beliefs of the agent in question, the testimony process requires input from a *testifier*. But there are two key differences between the process by which agents come to have testimony-based beliefs and the process by which they come to have reasoning-based beliefs. Though both are belief-dependent processes²⁹, the testimony process requires an input which is *not* a belief, rather, it requires *testimony*. Further, the testimony process requires an input from an agent *other* than the one who forms the belief.

²⁹ Admittedly, the testimony process cannot be belief-dependent insofar as it depends not on beliefs but on *testimony*. Goldberg states that the testimony process is a "quasi-belief-dependent process" (ibid. 73).

Goldberg recognises this and emphasizes the crucial role of the speaker in the testimony process. He argues that:

“[W]hether a given subject has acquired testimonial knowledge will typically turn on whether the testimony she consumed was itself reliable, which (given Process Reliabilism) will turn on whether the cognitive processes that produced the testimony – processes in the mind/brain of the subject's informant – were themselves reliable. In such cases, the hearer acquires knowledge only if certain cognitive processes *in her informant* were reliable” (ibid. 13).

A case can be made for this, he suggests, since we can construct pairs of cases which differ only in the reliability of the speakers' testimony (and subsequently, the cognitive processes responsible for the production of the testimony) and, as a result, yield differences in the audience's attainment of testimonial knowledge (ibid. 15). Consider a speaker who delivers a report as a result of careful inquiry and another who delivers a report as a result of wishful thinking. The thought is that an audience will attain knowledge from the former speaker, but not the latter, other things being equal. And the difference, per Goldberg, is *the reliability of the speakers' testimonies*.

Goldberg maintains that process-reliabilists classify beliefs as *reliable* when they are the product of a reliable belief-forming process (ibid.). So, my belief that there is a lamp on my desk is *reliable* since the cognitive belief-forming process which is responsible for me forming this belief is itself reliable. Reliable testimony is realised in the same way. So, what it is for a piece of testimony to be reliable is just for it to be produced by a “reliable type of cognitive process” (ibid. 17). If there is a requirement then that, in order to attain knowledge via testimony, one must receive reliable testimony, and that reliable testimony just is testimony produced by a reliable cognitive process, it follows that “whether a subject knows through testimony depends on the reliability of *cognitive processing in the mind/brain of the*

subject's informant" (ibid. 16). This *extendedness thesis* jettisons the notion of exclusively internal belief-forming processes and forwards that, the process by which *A* comes to believe that *p* from *S*'s testimony depends on the reliability of the cognitive processing of *S*. The testimony process is *interpersonal* then.

This is not to say that the cognitive processes *internal* to the audience have no epistemically important role. Goldberg appreciates this much (ibid. 61). He simply rejects the thesis that "the relevant cognitive processes are restricted to those "'internal to' the hearer-subject herself" (ibid. 61). Cognitive processes internal *and* external to the audience have a role in assessing the epistemic status of the testimonial-belief formed by the audience.

Say you tell me *that the shop is open* and that this report was formed on the basis of your seeing that the shop is open. Let's say further that you have no deceptive intentions when making this report. As such, your testimony is reliable insofar as it is the product of a reliable cognitive process. I receive your report and, provided I have no undefeated defeaters, am justified in accepting what you say, assuming that the belief I form is outputted by a reliable cognitive process. This is an example of the *testimony process* in action.

How does Goldberg's account afford us a way to satisfy the truth-directed testimonial discrimination condition? In outlining his account, he includes three necessary conditions on testimonial knowledge, one of which is, helpfully for the current project, a discrimination condition. The conditions are as follows:

1. "[T]he hearer must rely epistemically on what in fact was reliable testimony;
2. [T]he hearer must reliably comprehend the testimony; [...]
3. [T]he hearer's acceptance of the testimony must be grounded in her capacity for reliably discriminating reliable from unreliable testimony" (2007: 133).

Goldberg's inclusion of a discrimination condition proves incredibly helpful for my project. Given that he offers an account of testimonial knowledge, he is clearly no sceptic about our capacity to attain knowledge via testimony. So, we can present the following argument, given condition (3):

1. The capacity to reliably discriminate between reliable and unreliable testimony is necessary for the attainment of testimonial knowledge (ibid.).
2. We know things from testimony.
3. Therefore, we have a capacity to reliably discriminate between reliable and unreliable testimony.

But whilst he may recognise the need for a truth-directed testimonial discriminations condition, we lack an understanding of what *explains* that discriminatory capacity.

Given Goldberg's process-reliabilism, I suggest that he is most likely to give support to the reliabilist hypothesis. This should not be controversial given his claim that "what makes for doxastic justification and knowledge has to do with the reliability of the type of process (or method) through which a belief is formed and sustained" (2010: 11-12). Given that this will include the process by which testimonial-beliefs are formed, I submit that we can read Goldberg as supporting the claim *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process.*

But how would such a process work? Goldberg suggests that it requires a "'counterfactual sensitivity' to the presence of defeaters" (2007: 168). That is, it requires agents to demonstrate a sensitivity such that if there *were* defeaters, then one *would* respond to them, irrespective of whether there are defeaters in the current scenario. But how might an agent demonstrate such a sensitivity?

In his discussion of foundationalism and coherentism, Goldberg provides an answer to the question above. He writes that:

"[A]ll belief-forming and belief-sustaining processes incorporate at least the following two components: the source α , which produces mental representations as outputs; and the "filter" β , which is a subcognitive coherence-monitoring process that is elicited by α 's output, prior to the subject's endorsement of that output in full-fledged belief. β 's aim is to determine whether the content of the mental representation that α yields as output coheres with the subject's background information" (2012: 189).

Here we see Goldberg explicitly appeal to a coherence-monitoring system as necessary for belief formation. So, perhaps his response to the question of what the truth-directed testimonial discriminations process is doing is *it is monitoring for coherence*. And if that is so, then the proposals of Goldberg and Graham are not dissimilar. They both involve filters for coherence-checking.

There is reason to believe that, for Goldberg, the reliable truth-directed testimonial discriminations process is a process of coherence-monitoring. Consider the following, in which he provides an outline of how the proposed process would run:

[W]e can think of this process, schematically, as involving a "source" process that recovers a propositional content (and which represents that content as having been asserted by the speaker), followed by a coherence-monitoring filter whose aim is to assess whether there are any features of the attestation – implausibility of the attested proposition, signs of insincerity or incompetence in the hearer – which should prompt the subject not to endorse the proposition in question. (Imagine H observing T testify that p, where H comprehends this speech act as such, but where for some reason T's testimony just doesn't sit right with H, and so H does not form the belief that p – and then it dawns on H that the subject-matter is one on which T has vested interests.)" (ibid.).

Here we can see that the proposed filtering system is clearly involved in the making of truth-directed testimonial discriminations. This is because it is “the aim” of the coherence-monitoring system to monitor for “signs of insincerity or incompetence”. This filtering system will then “prompt” the audience not to endorse the testimony of the agent in question if a coherence-check issues a red flag, so to speak.

Goldberg’s proposal most certainly affords us an explanation of the hypothesised truth-directed testimonial discriminations process; it is a process of coherence-monitoring. This should be familiar from the discussion of Graham’s filtering system in the previous section. Given that it is a reliable process, the coherence-monitoring system can lead us to the truth because, since it outputs more true than false beliefs, it will push us to discriminate between bits of testimony in a truth-directed manner reliably.

Both Graham and Goldberg require that coherence-checking involves a sensitivity to counter-considerations (Goldberg, 2007: 168), (Graham, 2010: 153). In the previous chapter I argued that there is no difference in-principle between *evidence* and *counterevidence*. That which is evidence for p can also be counterevidence that $\sim p$. When you tell me that the shop closes at 5pm, this can be seen as evidence for the reported claim, or it can be seen as counterevidence for the claim that the shop *does not* close at 5pm. Moreover, depending on the background beliefs of the agent in question, a candidate piece of evidence can turn out to be evidence or counterevidence. As such, whilst Graham (2010) and Goldberg (2007) (2010) (2012) speak in terms of *counter-considerations* or *defeaters*, I maintain that the systems they have proposed can also be regarded as *evidence-monitoring systems* insofar as they are systems devoted to the detection of considerations which could be advanced in favour of the truth or falsity of a piece of testimony.

Having now shown how the reliabilist hypothesis receives support, and how we might understand the reliable process in operation, I will now highlight the issues facing it.

V

We see from Graham and Goldberg how the reliabilist hypothesis receives support. For both figures, testimonial belief is justified on the basis of it being outputted by a reliable process. For Graham, that process is the *filtering process* (2010). And for Goldberg, that process involves a “counterfactual sensitivity to defeaters” (2007: 168) which involves a “coherence-monitoring filter” (2012: 189). From here, generating the reliabilist hypothesis is simple. The reliabilist hypothesis maintains *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process*. That reliable process can be understood as a matter of filtering, which itself can be understood, as I have argued, as a matter of monitoring *evidence*.

For both Graham and Goldberg, the filtering process will include *at least* a coherence-checking filter. After this, there is room for manoeuvring, but it is important to recognize that both include filters which involve a sensitivity to other “signs of insincerity” (Goldberg, 2012: 189). This includes things like speaker behaviour (in the form of movement, perspiration, and tone of voice) and the speaker’s track record of reporting truths (Graham, 2010: 152). For all of this to work, however, it is vital that the process responsible for performing this task is *reliable*; it needs to *tend* to get things right.

In chapter one, I argued that we must understand our ability to make truth-directed testimonial discriminations as being *reliable*. This means that, when we make discriminations between bits of testimony in a truth-directed manner, we must tend to get things right; it isn’t just about *trying* to get to the truth, it’s about trying to get to the truth and *tending to succeed in doing so*. At the level of discriminations,

this will amount to accepting testimony when it true or likely true and rejecting testimony when it is false or likely false, *by and large*.

The reliabilist hypothesis makes explicit appeal to reliability. As such, advocates of the reliabilist hypothesis will maintain that the filtering system tends to get things right. That is, that the filtering system tends to push us to reject that which is false, or likely false, and accept that which is true, or likely true. The issue in positing that we have such a capacity is that it conflicts with the current empirical data.

The data to which I appeal here is obtained from studies into *deception detection*. These studies assess our ability to detect deceptions. Vrij outlines the typical format of such a study, involving the deception detection ability of laypersons, writing that:

“Laypersons’ ability to detect lies in strangers has been examined extensively in the last few decades. In a typical lie detection study, observers (normally undergraduate students) are given short video fragments of people they do not know who are either telling the truth or lying. They are asked to indicate after each fragment whether the person (often called the sender) was telling the truth or lying. Typically, half of the senders are truth tellers, and half are liars” (2008: 147).

If we have some reliable capacity to discriminate between reliable and unreliable testimony, then these studies should reflect this. After all, a lie surely counts as a classic example of unreliable or false testimony.

As stated by Vrij, in these studies, “typically, half of the senders are truth tellers, and half are liars” (ibid.). As such, he continues, “[i]n such a study, simply guessing whether the sender spoke the truth or lied would result in correctly classifying 50% of the truths (truth accuracy rate) and 50% of the lies (lie accuracy rate), resulting in a total accuracy rate of 50%” (ibid.). Additionally, classifying each report as ‘truthful’ would yield a truth accuracy rate of 100%, a lie accuracy rate of

0% and a total accuracy rate of 50%. Whereas classifying each report as 'deceitful' would yield a truth accuracy rate of 0%, a lie accuracy rate of 100% and a total accuracy rate of 50%. We would expect to achieve a total accuracy rate considerably higher than 50% if we are to defend the claim that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process.

The findings from these studies say otherwise, however. Bond and DePaulo write that:

"Despite decades of research effort to maximize the accuracy of deception judgments, detection rates rarely budge. Professionals' judgments, interactants' judgments, judgments of high-stakes lies, judgments of unsanctioned lies, judgments made by long-term acquaintances - all reveal detection rates within a few points of 50%. We wonder if it is premature to abort the quest for 90% lie detection and accept the conclusion implied by the first 384 research samples - that to people who must judge deception in real time with no special aids, many lies are undetectable" (2006: 231).

And so, we are presented here with a serious challenge to the idea that we have a reliable capacity to discriminate between veridical and false testimony. An accuracy rate of around 50% is hardly indicative a *reliable* capacity.

Similar comments can be found elsewhere. Aamodt and Custer conducted a meta-analysis into deception detection, concluding that "local and federal law enforcement agencies have levels of accuracy in detecting deception similar to students. The accuracy rate for students in this meta-analysis (54.22%) is similar to, but a bit lower than, the 57% reported in an earlier and much smaller meta-analysis by Kraut (1980)" (2006: 9). And Vrij has argued that "laypersons and professionals perform very slightly above the level of chance in lie detection when observing someone's behaviour or listening to their speech" (2014: 321).

Any advocate of the reliabilist hypothesis must appeal to the operation of a *reliable* process to account for the truth-directedness of our testimonial discriminations. The problem is that we have reason to think that such processes are *not reliable*. And if they are not reliable, then a defence cannot be provided for the reliabilist hypothesis. Decades of research into deception detection have yielded very similar results, indicating that our ability to discriminate truth from lies is only slightly better than that which we would expect with chance. Given that both Graham and Goldberg advocate process-reliabilism, they maintain that the reliability of a process is constituted by its tendency to output true beliefs. And so, given that our capacity to discriminate between reliable and unreliable testimony does not have such a tendency, it cannot be deemed reliable.

Note that the challenge raised here proves problematic not just for the reliabilist but for the truth-directed testimonial discriminations condition more generally. If we lack an ability to discriminate between veridical and false testimony reliably, then we just can't make truth-directed testimonial discriminations. And if that is so, then either we don't know anything from testimony, or the truth-directed testimonial discriminations condition must be wrong. I appreciate this challenge but will not respond to this immediately. First, I aim to focus on how the advocate of the reliabilist hypothesis might respond to the challenge raised by the empirical data.

Firstly, the reliabilist might suggest that the deception detection studies only test *some* filters. Perhaps these studies can give us data about our ability to detect deceptions by observing speaker behaviour, but they *cannot* give us data about our ability to detect deceptions by filtering for coherence or by considering track record information. To see this, imagine that you are conversing with a total stranger and have no information about the track record of that speaker. In such a situation, it is not clear how a filter which outputs responses on the basis of the speaker's track record can be expected to do its job. If I don't have any information about your track record of reporting truths, then the filtering process which requires such information

cannot function. Moreover, it is not clear how the coherence-checking filter is tested in these studies. If participants received testimony which failed to cohere with their background beliefs and likewise failed to respond to this information, we might be able to make that claim. As it stands, however, we have no reason to doubt the reliability of this particular filtering process. We are left to conclude, therefore, that the deception detection studies focus on one sort of filter: the filter which checks for observable signs of insincerity or incompetence.

The current challenge is that the poor results from the deception detection studies indicate that the speaker-behaviour filtering process for discriminating between veridical and false testimony is unreliable. We can ask *why* is that the case? In answering that, let us first ask what participants are trying to do in the experiment. It seems that they are trying to be sensitive to the physical movements of the speakers, their eye contact, their manner of delivery, their tone of voice and so on. The problem is that these are not reliable cues to deception at all (ibid. 4-5) (Kassin & Fong, 1999: 511). And what's more, some testifiers develop *countermeasures* to combat these classic signs of deception (Vrij, 2008: 375). That is, some testifiers are aware of classic "tells" and act in such a way as to purposely avoid those behaviours. For example, a speaker might make an effort to maintain eye contact with her audience because she believes that averting one's gaze is a classic sign of deception.

Vrij suggests that one of the reasons we fare so poorly in the deception detection tests is that lie detection is a difficult task. There are no reliable cues as to when a deception occurs. There is no 'Pinocchio Effect'; a liar's nose does not grow when they tell a lie (2008: 374). And if there are no reliable cues, there surely cannot be any reliable ways to detect a lie by observing speaker behaviour. This might help to explain just why we perform so poorly: there is one filtering process in operation, and it is an unreliable one. But this raises a new issue: if the reliable truth-directed testimonial discriminations process consists in the operation of more than one

process and at least one of the processes is unreliable, then how is the reliability of the process as a whole to be understood?

But one might suggest that I have been too quick to call the speaker-behaviour filtering process unreliable. In suggesting this, one might forward that the deception detection studies present us with *experimental conditions* which contrast with the everyday conditions we typically experience. Normally, we have an understanding of the social norms in play in a given testimonial exchange, we are dealing with people we know, and we have background beliefs on which we can rely. Under experimental conditions, this is all affected. Consequently, one might argue, the reliability of the truth-directed testimonial discriminations process must be understood as reliable *under certain conditions*. The point here, is that the truth-directed testimonial discriminations may well be unreliable in experimental settings. Under non-experimental settings, however, there is no cause for concern.

The line of argument introduced above echoes Sosa's view that "our faculties or virtues give us knowledge only if they work properly in an appropriate environment" (1991/2010: 276). Put otherwise, the reliability of a given belief-forming process depends, in some sense, on certain factors within the environment in which the process operates. The reliability of our visual belief-forming processes, for example, is indexed to environments in which we can see without hindrance, where the lighting is not too dark, where we have not been administered hallucinogenic drugs, and so on. This makes sense since our ancestors would have typically lived in environments similar to ours, visually. As such, the reliability of human vision is indexed to a certain kind of lighting, amongst other things. So, whilst we might tend towards false belief in very dimly lit environments, this says nothing about the reliability of the visual belief-forming process *per se*. Rather, it says something about the reliability of the visual belief-forming process *in a given environment*. Perhaps we can say the same of testimony: the reliability of our truth-directed testimonial discriminations is indexed to *non-experimental* environments.

By stipulating that the reliability of a given process is dependent to some extent on the environment in which it operates, the advocate of the reliabilist hypothesis can overcome the challenge from the empirical data. The thought here is that the truth-directed testimonial discriminations process is reliable in non-experimental environments but unreliable in the experimental setting. We can use this same response to overcome the more general problem facing the truth-directed testimonial discriminations condition. This is precisely how I respond to it. Whilst the empirical data might show that we don't perform very well in the deception detection studies, this should not concern us. Our ability to make truth-directed testimonial discriminations can be thought of as an evolved ability. As such, it evolved to perform well in the sorts of environment our ancestors inhabited, i.e., not in the experimental domain.

The advocate of the reliabilist hypothesis can be seen as maintaining that the truth-directed testimonial discriminations process works fine in "everyday" conditions but not in the experimental domain. We can ask: *why* is that the case? There are a few suggestions here: 1) the process doesn't operate in such conditions; 2) the process is not *reliable* under such conditions. Now, if the former suggestion is accurate, then it would entail that all reports are judged as veridical. Recall, filtering is developed as a countermeasure to combat false testimony. If there is no filtering, then we simply lack the cognitive tools to reject testimony. So, the former route cannot work. Participants in the studies accept and reject testimony.

The second route available to us suggests that the process isn't reliable in the experimental setting. This says nothing of its *ability* to operate. Indeed, the process does operate in experimental settings, but it does so badly. This explains the poor results from the empirical studies. But we want to know *why* this is so.

Why is it the case that the truth-directed testimonial discriminations process works fine outside of the experimental domain but is negatively affected within it? In short, what *hinders* the process? I see that there are two major suggestions here:

i) that the ratio of false testimony to veridical testimony is much higher in the experimental domain than in everyday life; ii) that, in the experimental domain, there is a lack of available evidence the kind of which we need to monitor in making truth-directed testimonial discriminations. Let us deal with the former suggestion first.

One might try to account for the poor performance of participants in the deception detection studies by suggesting that there is far more false testimony in the experimental setting than there is outside of it. Due to this, the response goes, the process cannot function as it normally does. The process has evolved to deal with a particular ratio of true to false testimony, one might say. In the experimental setting, that ratio is dramatically changed and so, of course the process fails to run as it should.

But take as a contrast the ability to discriminate between the English language and the French language. If you have an ability to reliably discriminate between such languages, then it should not matter whether you are reading a book with 20% English and 80% French text or 50% English and 50% French text. The fact that you have a reliable ability to make those discriminations entails that you will, most of the time, correctly pick out which language is which. So, if we have a reliable capacity to make truth-directed testimonial discriminations, it cannot be that a mere increase in the ratio of false to veridical testimony can affect the reliability of that process for the same reason as the language example shows.

Further, in explaining the significant difference between experimental and non-experimental environments by pointing to the ratio of false to veridical testimony, there is a worry that such a suggestion jettisons the need for any reliable truth-directed testimonial discriminations capacity at all. The worry here is that reliance on such an explanation gives the fact that there is a predominance of veridical testimony in non-experimental settings too much explanatory power. If what explains the poor performance in the experiments is simply that the ratio of

false to veridical testimony has increased, then it starts looking as though the ability to make truth-directed testimonial discriminations is unimportant. Instead, all that need be the case is that we are disposed to accept what we are told and since we are mostly offered veridical testimony, we mostly form true testimonial beliefs. This, of course, changes in the experimental setting³⁰.

Consider an analogy: if you live in an area in which 90% of the birds are buzzards, then your poor ability to discriminate between buzzards and kites will not hinder you so much, *provided you are disposed to believe that the bird you see is a buzzard*. If you move to an area in which 50% of the birds are buzzards, then your belief-forming method no longer serves you well. Suddenly, you will start forming far more false beliefs than you were before. This can be applied to the matter at hand. Let's say that we lack a capacity to make truth-directed testimonial discriminations reliably. If we inhabit an environment in which most of the testimony offered is veridical, then lacking such a capacity will not affect us *so long as we are disposed to accept what we are told*. Lacking such a capacity would negatively affect us in the experimental domain however, as the ratio of veridical to false testimony would be altered. So, we can explain the poor performance in the deception detection studies if we are happy to concede that we lack a capacity to make truth-directed testimonial discriminations *and* are disposed to accept what we are told.

The explanation currently under examination demands a huge concession because it forwards that we have no reliable capacity *whatsoever* for making truth-directed testimonial discriminations. It is not the case that there is a reliable truth-directed testimonial discriminations process operating in "everyday" settings, rather, there is no such process at all. But the arguments I presented in chapter one, give us very good reason to believe that we have a cognitive ability to make truth-directed testimonial discriminations. My arguments from doxastic responsibility and

³⁰ Such a hypothesis allows for testing. We could devise a study in which participants deal with mostly true testimony and examine the manner in which they respond to it.

from epistemic luck both show that truth-directed testimonial discriminations are necessary for knowledge. So, for the explanation under examination to work, it requires the falsification of the arguments I presented in the first chapter. One might take the sceptical route, arguing that we do not know so much from testimony, but dealing with such claims is beyond the scope of this project. The point here is that opting for this first route might commit one to dispensing of our ability to make truth-directed testimonial discriminations. It will, at the very least, certainly give their role less importance.

Let us now consider the second route; that whilst there is a reliable truth-directed testimonial discriminations process, it is not reliable in *experimental conditions*. As we have seen, both Graham (2010) and Goldberg (2010) (2012) propose that we have some kind of 'filtering system'. This filtering system just is the truth-directed testimonial discriminations system, or process, given the reliabilist hypothesis framework. As already stated, this system consists of a range of filters. There are filters that run checks for hesitancy in speaker delivery, some run checks on the track record of the speaker, some perform a coherence check on any received testimony, and so on. In doing so, the truth-directed testimonial discriminations process is demonstrating a sensitivity to defeaters (Goldberg, 2007: 168) or counter-considerations (Graham, 2010: 173). But, as previously argued, there is no difference in principle between evidence and counterevidence. That which is evidence for p can, depending on the context, be counterevidence that $\sim p$. As such, instead of regarding the truth-directed testimonial discriminations process as involving a sensitivity to counter-considerations, I have argued that we can regard it as involving a *sensitivity to evidence*. And it is here that we see the significance of evidence for the reliabilist position.

I suggest that we can make sense of the difference between the experimental and non-experimental settings by looking at the availability of *evidence*. This is relevant because it is *evidence* that the filtering process is trying to monitor. If it is

available in one situation but not the other, then we can explain why the process can function reliably in one domain but not the other.

When we consider the format of the deception detection studies, it is clear that there is a paucity of the kind of evidence we tend to have available to us in our everyday exchanges with people. We are usually dealing with interlocutors familiar to us, we tend to understand the sort of social norms in play, we have relevant background beliefs to which we can appeal, and so on. In the deception detection studies, those considerations are unavailable. If the filtering process is a process which monitors evidence, then we can easily explain why it fails to operate reliably in the deception detection studies. Yes, it receives an input, viz., testimony, but it is not able to perform the sort of work that it usually would, *given the lack of available evidence*. The evidence is all-important here because without that, the process cannot perform as it otherwise would.

We see then that appealing to evidence here is of great explanatory value. In short, the reason that we perform badly in the deception detection studies is that there is not nearly as much evidence available to us. Given that the truth-directed testimonial discriminations process is a process of monitoring evidence, a shortage of evidence negatively impacts us insofar as the truth-directed testimonial discriminations process lacks what it needs to carry out its operations.

To appreciate the implications of this, let us now view the reliabilist hypothesis and the evidential hypothesis side-by-side:

Evidential Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence.

Reliabilist Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process.

The implication of adopting the approach above, I propose, is that the reliabilist hypothesis must agree with the evidential hypothesis.

To see why the reliabilist hypothesis agrees with the evidential hypothesis, we must consider what it is that the reliable truth-directed testimonial discriminations process is doing. I have argued that the truth-directed testimonial discriminations process demonstrates a sensitivity to *evidence*. The reliabilist hypothesis states that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process. But, given that that process involves a sensitivity to evidence, this leads us to the conclusion that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process *which operates by way of a sensitivity to evidence*. After all, it is precisely because the process is sensitive to evidence, that it is able to carry out the operations required of it. Without that, the process cannot operate.

In arguing that the reliabilist hypothesis agrees with the evidential hypothesis, one might note I am committed to viewing evidence sensitivity as involving coherence-checking. I am happy to concede this. The evidential hypothesis states that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence. It is entirely consistent with the evidential hypothesis to maintain that processes such as coherence-checking are background processes which operate by checking incoming testimony against the background beliefs of the audience³¹. In checking incoming testimony against background beliefs of the audience, the audience is thereby demonstrating a sensitivity to evidence. And, importantly, this need not be something about which the audience is conscious.

The reliabilist may very well say that it is *because* of the filtering process that we can make truth-directed testimonial discriminations; without that process, we can't make them whatsoever. The advocate of the evidential hypothesis can accept

³¹ In chapter five, I will defend Williamson's E=K conception of evidence which equates one's evidence with one's knowledge (2000: 185). The effect of this is that a process of coherence-checking will be a process of checking incoming testimony against one's evidence, or, given E=K, one's knowledge.

that. And this is because, when we elucidate the nature of the filtering process, we see that it depends on evidence insofar as the filtering process monitors it. Allowing that our testimonial discriminations are truth-directed because of a reliable evidence-monitoring system requires one to support the claim that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence. So, the reliabilist hypothesis must agree with the evidential hypothesis. In short, if one defends the reliabilist hypothesis in the way so far discussed, then one is forced to defend the evidential hypothesis. The reliabilist hypothesis necessitates agreement with the evidential hypothesis. This is not to say they are one and the same. One can discuss an agent's sensitivity to evidence without appealing to reliable processes³².

VI

My discussion so far has focused exclusively on *process-reliabilism*. The proposals of both Graham (2010) and Goldberg (2010) are straightforwardly process-reliabilist. But this is not the only kind of reliabilism on offer. I will now focus on the work of Jennifer Lackey, who has advocated a 'dualist' account of testimonial knowledge and justified acceptance (2008). As we shall see, rather than supporting a process-reliabilist approach, Lackey takes a different reliabilist route, arguably appealing to the work of Nozick (1981).

Lackey's account is motivated by what she regards as, the failings of reductionism and non-reductionism. Reductionism, she claims, neglects the contribution of the *speaker* and focuses only on the *hearer*, requiring the hearer to have sufficiently strong *positive reasons* to accept the speaker's testimony (2008: 176). Non-reductionism, so says Lackey, recognizes the contribution of the speaker but includes too weak a hearer condition, permitting justified acceptance so long as the recipient of testimony is not in possession of any relevant undefeated

³² I will discuss the evidential hypothesis more in chapter five.

defeaters (2008: 176-177). The problem is that “it takes two to tango” (ibid. 177). A testimonial exchange involves *two* parties, and the aforementioned accounts fail to appreciate this, per Lackey. With this in mind, she writes that:

“The positive epistemic work of testimonial beliefs can be shouldered neither exclusively by the hearer nor by the speaker. To put this point somewhat crudely, the speaker-condition ensures reliability while the hearer-condition ensures rationality for testimonial knowledge (justification / warrant). Thus, we need to look toward a view of testimonial knowledge (justification/warrant) that gives proper credence to its dual nature, one that includes the need for the reliability of the speaker (from non-reductionism) and the necessity of positive reasons (from reductionism)” (ibid.).

In response, Lackey forwards an account which includes both *speaker* conditions and *audience* conditions.

Lackey provides the following set of conditions in forwarding her dualist account:

“D: For every speaker, A, and hearer, B, B knows (believes with justification / warrant) that p on the basis of A’s testimony only if:

(D1) B believes that p on the basis of the content of A’s testimony,

(D2) A’s testimony is reliable or otherwise truth-conducive,

(D3) B is a reliable or properly functioning recipient of testimony,

(D4) the environment in which B receives A’s testimony is suitable for the reception of reliable testimony,

(D5) B has no undefeated (psychological or normative) defeaters for A’s testimony has no undefeated (psychological or normative) defeaters for A’s testimony, and

(D6) B has appropriate positive reasons for accepting A’s testimony”

(ibid. 177-178).

As a note, Lackey writes that dualism includes *at least* these presented conditions. This account, therefore, is a set of necessary conditions rather than necessary and sufficient. She writes the following:

“Because dualism specifies only necessary conditions, there may be other conditions that need to be added for a complete epistemology of testimony. What is of import, here, however, is that testimonial knowledge (justification/warrant) requires positive epistemic contributions from both the speaker and the hearer” (ibid. 178).

So, Lackey’s major claim is that any account of justified acceptance or testimonial knowledge must include speaker conditions *and* hearer conditions. Importantly, at least *some* of those conditions make appeal to *reliability*. We can therefore regard her proposal as a kind of reliabilism.

I will focus more attention on conditions D2 and D6 since these are the key speaker and audience conditions respectively. Before doing so, however, it will be helpful to briefly examine Lackey’s other conditions. I will begin with D1 which I take to be uncontentious and D4 which is an *environment* condition.

D1 stipulates simply that *A*’s belief that *p* be based on the content on *S*’s report that *p*. This ensures that the justified belief, or knowledge, that *p* is *testimonial*, rather than *inferential* or something else entirely. Consider, one might receive testimony *that the shop is open*, or one might see this, or one might *infer* this inductively given the previous opening hours of the shop. Including this condition distinguishes testimonial belief from other kinds of belief.

D4 stipulates that the environment in which the audience receives the speaker’s testimony be suitable for the reception of reliable testimony. This, Lackey comments further, is a condition guarding against *environmental luck* (ibid. 165). I might be very unlucky in that I just so happen to inhabit an environment populated overwhelmingly by deceptive testifiers. Whenever these testifiers offer a report, they lie. Let’s say, however, that I stumble upon one of the few sincere and

competent testifiers in the environment who consistently reports truths. Given the state of the environment which I inhabit, I experience good fortune when I accept this testifier's report and believe that p truly. And this is because, it is far more likely that I speak with a deceptive testifier, and if I had done so, I would come to accept their report *but believe falsely* (ibid. 164-165). In such an environment, any true belief that I might attain from a speaker's report is a matter of luck, given the risk of error resulting from the insincerity of most of my interlocuters. In order to overcome this problem, Lackey employs D4, requiring that the environment be suitable for the reception of reliable testimony.

D3 requires that the audience be a reliable or proper functioning recipient of testimony. This is connected to D5, which requires that there be no relevant normative or psychological undefeated defeaters for S 's testimony. Put otherwise, D5 requires that there are no sufficiently strong counter-considerations against S 's report that p , where such considerations are either possessed by A (in the case of psychological defeaters) or are the sort of considerations to which A ought to respond (in the case of normative defeaters). If, for example, S reports to A that it is 2pm on a summer's day in Paris when there is no daylight, A has an undefeated defeater for S 's report. And this is because A knows that in the summertime in Paris, the sunset is around 9pm. As such, A has reason to reject S 's report. And even if A does not attend to such a consideration, it is the kind of thing to which A ought to respond. This too fails to permit justified testimonial acceptance. But for A to respond to such a consideration she must have the *capacity* to respond to defeaters. This is Lackey's D3 condition. If it is the case that there are some considerations to which I *ought to respond*, then it follows that I must have the *capacity* to respond to it. So, we can view D5 as a condition regarding the audience's *responsiveness* to defeaters, whilst D3 focuses on the audience's *capacity* to respond.

One might ask why D3 is included as its own condition and not simply an extension of D5. Expanding on this, Lackey states that "[i]n order for satisfaction of

the no-defeater condition to properly reflect on the epistemic agent, the agent in question must have the *general capacity* to recognize and process counterevidence; otherwise put, the agent must *substantively* satisfy the no-defeater condition" (ibid. 161). This might also be understood as a matter of being appropriately disposed to respond to defeaters should they arise. Or, to use Goldberg's terminology, as a "counterfactual sensitivity to the presence of defeaters" (2007: 168). Though, it is worth noting here that Lackey does not make a direct appeal to the operation of some reliable process as Goldberg does.

We are now left with D2 - that *S*'s testimony be reliable or otherwise truth-conducive - and D6 - that *A* have appropriate positive reasons for accepting *S*'s testimony. In D2 (as well as D4), we see the term *reliable* being used to describe a speaker's testimony rather than a process. There is an immediate concern with this. For Goldman, as we have seen, something is *reliable* if it has a tendency to output the truth. Per Goldman, *processes* are reliable, and reliability is a statistical property which can only belong to *types*. A given piece of testimony cannot be a type, since it occurs at a particular time and is offered by a particular agent. As such, a given occurrence of some event cannot have the property of, say, yielding the truth 80% of the time (Goldman, 1979: 38-39). Consider, if I report to you that *p*, how are we to make sense of the question: *does that report have the tendency to produce truth and avoid falsity?* It is very difficult to make sense of such a question. Lackey must mean something else when she talks about reliability. But, as we shall see, Lackey does not use the term 'reliable testimony' in the same way that Goldberg does (2010: 17). So, we cannot simply make sense of reliable testimony as the product of a reliable testimony-producing cognitive process.

Whilst defending her "statement view of testimony", Lackey argues for the reliability of statement-necessity thesis:

“RS-N: For every speaker, A, and hearer, B, B knows (believes with justification/warrant) that p on the basis of A’s testimony that p only if A’s statement that p is reliable or otherwise truthconducive” (2008: 74).

She then writes the following:

“The reliability of the statement in question can, in turn, be fleshed out in any number of ways. For instance, it may be necessary that the speaker’s statement be [1] sensitive, [2] safe, [3] properly or [4] virtuously formed, and so on. Accordingly, it may be necessary that [1] a speaker would not state that p if p were false, or that [2] she would not state that p without it being so that p, or [3] that her statement be offered by testimonial faculties functioning properly or [4] virtuously, and so on. What is of import for our purposes, however, is that being a competent believer and a sincere testifier are epistemically relevant only insofar as they bear on the speaker’s capacity to be a competent testifier” (ibid.).

Unfortunately, we lack here a precise definition as to what Lackey means by reliable testimony, as she allows for it to be realised in different ways. However, we have now at least some candidates as to what she might mean by the term ‘reliable testimony’.

One suggestion Lackey includes is that the speaker’s statement be *sensitive*. But the *sensitivity condition*, first introduced by Nozick, concerns *belief* and not *testimony*. It is defined thusly:

“If p weren’t true, S wouldn’t believe that p” (Nozick, 1981: 172).

Applied to testimony, we can reformulate it to require that:

If p weren’t true, S wouldn’t have reported that p.

Realised in this way, we can view D2 straightforwardly as a sensitivity condition on speakers’ reports. If the speaker satisfies the sensitivity condition, then their testimony is reliable.

What I have presented here is one route Lackey might take in talking about reliable testimony. Whilst I risk presenting something with which she may disagree, I do so with the intention of attaining greater clarity. I recognize that I may be in danger of presenting a distorted account, but I hope that the range of possible understandings of Lackey's work allows my interpretation to present at least one possibility which is not uncharitable. In keeping with this aim of attaining clarity, let us make D2 clearer still by putting it in terms of *sensitive* rather than *reliable* testimony. In doing so, we specify Lackey's speaker condition for testimonial knowledge, requiring that:

(D2) The speaker's testimony is sensitive, such that if p were false, she wouldn't have reported it.

This is helpful in making clearer Lackey's theory, but it does nothing to advance our understanding of how Lackey's theory might support the reliabilist hypothesis. So, let us now deal explicitly with the question of how, if at all, Lackey's theory can rescue the reliabilist hypothesis.

I will answer this question now: it cannot. The issue is that D2 is a speaker condition, and the truth-directed testimonial discriminations condition *must be an audience condition*. We want to explain how our testimonial discriminations are truth-directed and, to do that, we need to focus on the *audience* as it is the audience who makes discriminations between bits of testimony. So, we must look at the audience conditions Lackey provides to see if her theory supports the reliabilist hypothesis in a different way.

In response to this, one might ask whether Nozick's sensitivity condition could be used as an audience condition to ensure that Lackey's account supports the reliabilist hypothesis. In taking this route, one would place a condition on testimonial acceptance such that if p were false, the audience, A , wouldn't accept that p . This alone is inadequate, however. Indeed, so too thought Nozick, who introduced a second subjunctive condition: If p were true, S would accept that p

(ibid. 176). Together, this yields Nozick's "truth-tracking theory". So, can this support the reliabilist hypothesis that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable process / faculty? The answer is no.

We want to explain the *truth-directedness* of our testimonial discriminations. Here, the truth-tracking theory is ill-equipped to aid us. This is because Nozick's subjunctive conditions focus on *particular beliefs* rather than capacities or abilities to make discriminations. Rather than giving an account of what justifies our testimonial beliefs, sensitivity is instead a condition employed to combat epistemic luck, looking at specific beliefs in specific scenarios. At best, we could yield a theory which attempts to explain the truth-directedness of our testimonial discriminations by saying that they are truth-directed because they demonstrate our capacity to track the truth. But this is not explanatory at all. Now we need an explanation of our capacity to track the truth. This, however, is the very project with which we are currently engaged; we are trying to explain the truth-directedness of our testimonial discriminations. As such, sensitivity fails to provide a viable route for us here.

Lackey's D5 condition, as we have seen, requires that the audience have no undefeated (psychological or normative) defeaters for the speaker's testimony (2008: 178). But it is not the case, for Lackey, that this is the *only* epistemic work required of the audience. She requires further that the audience have *positive reasons*. Given that the current aim is to explain the truth-directedness of our testimonial discriminations, we need to look at what Lackey requires of the audience epistemically. This is because it is the epistemic work of the audience which is considered in producing an explanatory hypothesis for the truth-directedness of our testimonial discriminations. So, let us examine Lackey's other audience condition (D6), requiring that the audience have positive reasons.

Lackey writes, "no matter how reliable a speaker's testimony is, this cannot by itself make it rational for a hearer to accept her report" (ibid.). Instead, we need a

condition requiring that the audience have *appropriate positive reasons* for accepting the speaker's testimony (Lackey, 2008: 178). But this requirement is not a *strong* requirement, she says. All that is necessary is that "the positive reasons possessed by a hearer need to be such that they render it, at the very least, *not irrational for her to accept the testimony in question*" (ibid. 181). Of course, this prompts the question of what *kind* of reasons can render testimonial acceptance "not irrational". Additionally, it prompts the question of what it is for something to be "not irrational".

In answering the question of what kind of reasons can render testimonial acceptance "not irrational", let us consider the following example from Lackey:

"Even if I do not have specific beliefs about British newspapers, I have all sorts of beliefs about England, the people who live there, their government, their social and political values, and so on. Surely, this information is enough to make it *not irrational* to form beliefs on the basis of British newspapers, even if it is not itself fully sufficient for *justifying* such beliefs or for rendering the formation of such beliefs positively *rational*" (ibid.).

In this example, Lackey cites a number of reasons which make it *not irrational* to form beliefs on the basis of reading a British newspaper. These reasons can be understood as weaker than *justifying* reasons, which would allow for justified belief. Instead, they can be thought of as *rationalising* reasons, which are weaker and do not necessarily entail justified belief. So, the reasons which make it *not irrational* to believe that *p* can be thought of as rationalising reasons that *p*. The thought here is that reasons which make it not irrational to believe that *p* just are reasons which make it rational to believe that *p*.

Presumably, if Lackey believed that British newspapers were typically misleading, then such a belief would act as a reason to render beliefs formed on the basis of reading a British newspaper *irrational*. One suggestion as to why this is the

case is that she would have a background belief which fails to *cohere* with the formation of beliefs on the basis of reading British newspapers. This seems to rest on the idea that rationality consists in a kind of coherence between one's various doxastic states. Such an idea can be found in the work of Sosa, who writes that reason's fundamental objective is "to maximize the coherence and comprehensiveness of our body of beliefs" (1985: 21-22). To be irrational, on this picture, would be to adopt some doxastic state which fails to fit, or cohere, with one's background beliefs. Likewise, to be rational is to adopt some doxastic state which fits, or coheres, with one's background beliefs. In the example above, testimonial acceptance fits with Lackey's background beliefs which, presumably, are along the lines that British newspapers are a reliable source of information. And this is because accepting what the British newspaper says coheres with the reasons that Lackey has for forming beliefs on the basis of what the British newspaper says.

Commenting more specifically on the kind of positive reasons required of the audience, Lackey states that "there seem to be at least three classes of inductively based positive reasons that are available to epistemic agents for distinguishing between reliable and unreliable testimony" (2008: 191). Those three classes are as follows:

- 1) "[C]riteria for individuating epistemically reliable *contexts* and *contextual features*" (ibid. 182).
- 2) "[C]riteria for distinguishing between different kinds of *reports*" (ibid.).
- 3) "[C]riteria for identifying epistemically reliable *speakers*" (ibid. 183).

Given that these reasons are discussed in relation to assisting agents in discriminating between reliable and unreliable testimony, these reasons can be regarded as concerned with the *truth* or *reliability* (of the testimony). If they were not, how could they possibly assist agents in discriminating between reliable and unreliable testimony? Let us examine each class in turn.

The first class involves reasons which concern the *context* in which a report is delivered. Lackey notes that one's stance might be more critical in an astrology lecture relative to an astronomy lecture (ibid. 182).

The second class involves reasons which concern the *kind* of report delivered. We can make intuitive distinctions between different kinds of reports. For example, we can distinguish between *testimony concerning oneself*, that is, testimony that you offer about yourself (e.g., your name, age, date of birth), and *testimony concerning alien encounters* (ibid). Again, one's stance might be more critical in receiving the latter kind of report.

The third class involves reasons concerning *reliable speakers*. That is, it concerns those reasons which help to identify a speaker who tends to offer reliable testimony. For example, you may have excellent reason to believe your friend's testimony, given the evidence you have amassed concerning their track record of reporting truths and not reporting falsities. Lackey notes, however, that such reasons need not concern *token* speakers, but *types* of speakers. She writes that "one may have accumulated inductive evidence for believing that accountants tend to be reliable sources of information about taxes, while politicians in the middle of their campaigns tend to be unreliable sources of information about the characters of their political opponents" (ibid. 183). So, the third class of reasons involve reasons which help to identify reliable *types* of speakers (though, this can include reliable *token* speakers also).

But the picture emerging here is a far cry from the reliabilist accounts so far examined. Another look at Lackey's conditions for testimonial knowledge clearly illustrate as much. For Lackey, 'reliability' features only on the side of the *speaker*³³. But it is the *audience* who has the job of discriminating between bits of testimony

³³ Whilst D3 requires that the audience is a *reliable* or proper functioning recipient of testimony this just equates to the requirement that the audience has the *capacity* to respond appropriately to defeaters, or that the audience is disposed such that if there *were* relevant defeaters, she would respond to them appropriately.

and does this through attending to their *evidence*. Consider again the audience conditions that Lackey employs³⁴:

(D3) The audience must have the capacity to respond to defeaters / be disposed such that they *would* respond appropriately to defeaters;

(D5) The audience must have no undefeated defeaters;

(D6) The audience must possess inductively based positive reasons which render testimonial acceptance, at the very least not irrational.

Unless Lackey proposes that some cognitive system is responsible for carrying out these operations, it looks like Lackey is more closely aligned to the position offered by the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*.

As argued in chapter two, there is no difference in principle between evidence and counterevidence. That which is evidence for p can, at least sometimes, be counterevidence that $\sim p$. So, in requiring that agents be disposed such that if there were defeaters, they would respond to them, Lackey requires agents to demonstrate a kind of counterfactual monitoring of *evidence*. This does some work in pushing her account closer towards the evidential hypothesis, but when we look to her positive reasons requirement (D6), there is an interesting discussion to be had.

Lackey's D6 audience condition requires that the reasons possessed by an audience need to render testimonial acceptance, at the very least, "not irrational". I suggested that Lackey might be adopting here the position that rationality consists in a kind of coherence between doxastic states here. We might reword this to yield an explanatory hypothesis stating that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to reasons which maximise the coherence of our beliefs. But consider that one might have a bunch of beliefs which are false *but coherent*. When one receives a piece of testimony which coheres with

³⁴ I have reworded the conditions to remove reference to reliability.

that set of false beliefs, one has positive reasons to accept that piece of testimony. But in what sense are those beliefs truth-directed? The positive reasons condition Lackey forwards is satisfied not on the basis of *truth* or *likely truth*, but on the basis of *coherence*. Coherence takes priority over truth, but coherence is not *necessarily* truth directed; the two can come apart.

To show that coherence maximisation can allow for systematic error, it is enough to provide cases which show that increasing coherence fails to push one towards the truth or probable truth. Generating such examples is not difficult. Take a conspiracy theory, say, that the government is tracking its citizens by injecting them with 5G chips. The individual who believes this conspiracy theory will have beliefs related to the proposition *that the government is tracking its citizens by injecting them with 5G chips*. For example, they might believe that the government is spying on us by way of the 5G chips, that the government is engaged in a large-scale coverup of their 5G spying operation, that the NHS is conspiring with the government, and so on. These beliefs hang together quite nicely, despite being obviously false. Now imagine that this individual hears a friend complain of a sore arm after a recent injection. "Well..." this conspiracy theorist thinks, "being injected with a 5G chip probably hurts. Aha! My friend has been injected with a 5G chip!". Forming such a belief maximises coherence insofar as it allows this individual's beliefs related to the conspiracy theory to hang together even better. What explains their friend's sore arm is that they have been injected with a 5G chip. No beliefs need to be revised here; they all cohere beautifully. In short, adopting such a belief maximises coherence.

The issue with the case above is that the individual is, very clearly, far away from the truth despite having a coherent body of beliefs. And if one can go wrong in such a way by maximising coherence, then this shows that coherence and truth can, at least sometimes, come apart. That is, maximising coherence can allow for *systematic error*. You can maximise coherence but get nowhere near the truth. And

if that is the case, then attempting to explain the truth-directedness of our testimonial discriminations by looking at coherence *alone* will fail to provide a tenable explanation. And this is because one can be sensitive to reasons which maximise coherence, but such reasons need not get you to the truth.

Notice that this criticism strikes a blow against the process-realist accounts discussed earlier. Recall, those accounts rely on a filtering system which is composed of various filters: a filter for tracking observable speaker-behaviour, a filter focusing on the speaker's track record, and a filter for *checking coherence*. We had reason to be suspicious of the reliability of the process focusing on observable speaker-behaviour, but this was not too problematic as it left us with the processes for checking coherence and speaker track record. But if coherence and truth come apart, then the filter here is affected too. What can we say about this?

As we have already seen, for Goldberg (2010) (2012) and Graham (2010), checking for coherence is a filtering process which has been developed as a countermeasure against the possibility of accepting false testimony. Through this process operating normally, we check any incoming testimony for coherence and then accept it depending on the result of that check. But the example presented above shows that coherence can come apart from the truth. The coherence checking process allows for *systematic error* given that agents can have coherent bodies of beliefs which are entirely false, and which direct them to form more false beliefs for the sake of coherence. Again, this is not to say that coherence will never get one to the truth. It may be that maximising coherence can tend towards true belief when we have mostly true beliefs. Or it may be that maximising coherence on the basis of particular reasons can tend towards true belief. I will not dwell on this point. What I wish to emphasize here is that, insofar as coherence can come apart from the truth, we must concede that the coherence-checking filter can tend towards failure in the form of pushing us to accept falsities.

Let us now return to Lackey's position. With the argument presented above, we have reason to believe that Lackey's account fails to accommodate the truth-directed testimonial discriminations condition. The problem is, as we have seen, the D6 audience condition is too weak. A condition which focuses on reasons which make it "not irrational" to accept a piece of testimony do not allow for the sort of connection to the truth, or likely truth, that truth-directed testimonial discriminations require. I take it that only one route is left for one sympathetic to Lackey's position: appeal to a stronger version of the D6 condition.

I take it that Lackey's account, as it stands, employs a *weak sense of evidence*. By this, I mean that Lackey's account supports a sense of evidence which does not require that the reasons an audience has increase the objective probability that the speaker's report is true or false. Rather, the reasons an audience has must simply allow for a coherent body of beliefs when combined with the received testimony. But this says nothing of the *probable truth* of the received testimony. If that is the case, then explaining how our testimonial discriminations are truth-directed becomes quite the task, as the focus becomes *coherence* rather than truth, or probable truth.

Following this, I argue that one sympathetic to Lackey's account must appeal to a *strong sense of evidence* if they are to accommodate the truth-directed testimonial discriminations condition. By a "strong sense of evidence", I mean a sense of evidence which sees it as increasing the objective probability of the thing it supports. The Bayesian understanding of evidence provides us with one such example, as this conception sees something as evidence for a hypothesis when the probability of the hypothesis obtaining, given the evidence, is greater than the probability of the hypothesis alone (or $P(h | e) > P(h)$) (Williamson, 2000: 186)).

In supporting a strong sense of evidence, we appeal to a conception of evidence which sees the reasons an audience possesses for accepting what a speaker says as increasing the objective probability that what the speaker says is

true. Now, if one does this, then one will see very obviously that we have moved far away from the reliabilist hypothesis and that we are now comfortably in agreement with the evidential hypothesis. This is the dilemma facing Lackey's position: either the evidence condition is too weak, in which case, we cannot accommodate the truth-directed testimonial discriminations condition, or is it strong enough to accommodate it, but is simply in agreement with the evidential hypothesis. Note that on either route, Lackey's account gives no support to the reliabilist hypothesis, despite the reliabilist conditions she includes in her proposal.

VII

In this chapter, I have argued against the reliabilist hypothesis. After introducing Goldman's process-reliabilism in §II, I discussed the reliabilist accounts of Graham (2010) and Goldberg (2010) in §III and §IV. My aim was to show how one might support the reliabilist hypothesis *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process*. As the proposals from both Graham and Goldberg are in keeping with the process-reliabilism framework set out by Goldman (1979), it was not difficult to show how those accounts could be seen as supporting the reliabilist hypothesis. Both accounts maintain that we perform some kind of filtering which combats the threat of accepting false testimony. I suggested that, for these theories, it is the operation of a reliable filtering process (or set thereof) which explains our truth-directed testimonial discriminations.

Inquiring into the role of the filtering process, I found that it must monitor evidence. This is because, in performing a coherence check, something which both Graham (2010: 152) and Goldberg (2010: 75) take the filtering process to do, there must be a monitoring of *counter-considerations* (Graham, 2010: 153). But, as I argued in the previous chapter, there is no difference in principle between

counterevidence and evidence. Consequently, we can also regard the filtering process as a process of monitoring *evidence*.

After this, in §V, I introduced the empirical data amassed from the studies into deception detection. This presents a challenge to the reliabilist hypothesis which attempts to explain the truth-directedness of our testimonial discriminations by appealing to a *reliable* process. If there is reason to believe that the sort of process in question is *not* reliable, then the reliabilist hypothesis fails. And this is what the data can be interpreted as indicating. However, the empirical data is problematic not just for the reliabilist hypothesis, but also for the very notion that we have an ability to make truth-directed testimonial discriminations. In response to this, I remarked that there is a stark difference between experimental and non-experimental domains, viz., the availability of evidence. In the experimental setting, there is a distinct lack of evidence.

My next task was to assess the response that the reliabilist might give to the challenge issued from the deception detection studies. The best response, I found, came in the form of arguing that, though there is a reliable truth-directed testimonial discriminations process which is in operation in both experimental and non-experimental environments, that process is reliable *only* in non-experimental environments. But *why* doesn't the process work reliably in the experimental conditions? What *hinders* it?

After examining some possible replies, I found that the most explanatorily powerful response appeals to the paucity of evidence in experimental domains. This makes sense, I argued, if the filtering system performs *evidence monitoring*. The thought here is that, if the filtering system involves evidence monitoring, then it cannot operate as it should in the experimental setting where evidence is restricted. This is what explains our poor performance in the experiments. In "everyday settings", however, the truth-directed testimonial discriminations process functions just fine, given that evidence is not restricted in the same way. The upshot of this is

notable. In arguing that the truth-directed testimonial discriminations process must be carrying out evidence monitoring, the reliabilist hypothesis *must* agree with the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence.*

Next, I looked at Lackey's dualist account (2008). Lackey's account presents a contrast to the process-reliabilist accounts of Graham (2010) and Goldberg (2010) insofar as Lackey makes no explicit reference to the reliability of *processes*. Instead, she appeals to *reliable testimony*. Unlike Goldberg, Lackey does not realise this simply as testimony which is the product of a reliable testimony-producing faculty / process (2010: 17). Rather, Lackey provides a list of candidates for reliable testimony. One such candidate, the one on which I focused, appeals to Nozick's sensitivity condition. Understood in this manner, reliable testimony is testimony such that *if p were false, S wouldn't have reported it.*

After this, I asked how, if at all possible, Lackey's account might rescue the reliabilist hypothesis. I found that it was unable to do so. The issue is that Lackey's reliability condition focuses on the *speaker*. Since the truth-directed testimonial discriminations condition is an *audience* condition, Lackey's account fails to lend support to the reliabilist hypothesis.

In looking to see which hypothesis Lackey's account might support, I focused on her D6 audience condition, which requires that an audience have reasons which make it *not irrational* for one to accept the speaker's testimony. In attempting to make sense of this, I forwarded that Lackey might be appealing to Sosa's view that the objective of reason is "to maximize the coherence and comprehensiveness of our body of beliefs" (1985: 21-22). Following this, the D6 condition might be read as requiring the audience to have reasons which allow for a coherent body of beliefs.

Next, I argued that coherence and truth can come apart. This is evidenced by the existence of cases wherein one has a body of beliefs which hang together

elegantly but are very far away from the truth. This demonstrates, I maintain, that maximising coherence will not always get one to the truth. This is not to deny that there is epistemic value in maximising coherence. However, insofar as one can maximise coherence and be led away from the truth, we see that purely coherence-based reasons can allow for *systematic error*. It allows for us to get things wrong more often than we things right.

I then argued that Lackey's account employs a *weak* sense of evidence insofar as it relies on a sense of evidence which does not require an increase in the objective probability of the hypothesis it supports. In opposition to this, we can think of a *strong* sense of evidence which requires an increase in the objective probability of the hypothesis it supports. Insofar as Lackey's account appeals to a weak sense of evidence, it fails to accommodate the truth-directed testimonial discriminations condition and, therefore, cannot help to explain it. What is necessary, I maintain, is a sense of evidence which has a clear connection to the truth. Appealing to such an understanding of evidence, I argue, is the only route available to one sympathetic to Lackey's view. The upshot, however, is clear agreement with the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*.

Given my work in this chapter, we can now eliminate the reliabilist hypothesis as a candidate explanation of the truth-directedness of our testimonial discriminations. This leaves us with the following two hypotheses:

Evidential Hypothesis: Our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence.

Assurance Hypothesis: Our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said.

In the next chapter, I will discuss the assurance hypothesis.

4. The Assurance Hypothesis

I

Having now eliminated the non-reductive hypothesis and the reliabilist hypothesis from the pool of candidates currently under examination, my focus shifts to the assurance hypothesis *that our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said*. In this chapter, I argue that the assurance hypothesis can account for the truth-directedness of our testimonial discriminations *only* through an appeal to *evidence*. Without such an appeal, the assurance position is not equipped to deal with the question of what explains the truth-directedness of our testimonial discriminations. The upshot is that the assurance hypothesis collapses into the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*.

In §II, I introduce Moran's assurance view (2006), which I take to be a natural supporter of the assurance hypothesis. For Moran, testimonial acceptance is justified through the audience recognizing that the speaker, in reporting that p , is acting as 'guarantor' for the truth of p . In so doing, the audience is provided with a "characteristic reason to believe, different in kind from anything provided by evidence alone" (Moran, 2006: 279). Thereafter, I appeal to the work of Faulkner, who argues that, for the audience, A to come to believe that p from a speaker's, S 's, report that p , it is necessary that A , "takes S 's intentions at face value, believes S , or trusts S to tell the truth" (2011: 143). Next, in §III, I show how the assurance position motivates the assurance hypothesis *that our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said*. In doing so, I introduce the *truth-connection requirement* which demands that trust is connected to the truth such that it can function as a guide to truth, or probable truth. If the assurance hypothesis cannot meet this requirement, then it is

entirely unclear how it can account for the truth-directedness of our testimonial discriminations.

In §IV, I make clear my understanding of trust, given its central role in the assurance hypothesis. Here, I make the argument that trust is connected to the truth insofar as it is connected to *trustworthiness*, which is connected to the truth. However, this connection does not allow us to explain the truth-directedness of our testimonial discriminations. This is because trust might utterly fail to capture trustworthiness. For the assurance hypothesis to provide a tenable theory, we need trust to tend towards success. To do this requires that it can pick up the distinction between trustworthiness and untrustworthiness. If it cannot capture that distinction, then we allow that trust can, by and large, get things wrong.

In §V, I introduce the first available route to defending the claim that trust functions as a guide to truth, or probable truth: that trust *causes* trustworthiness. One might defend this proposal by suggesting that, in trusting a speaker, we thereby give her a reason to prove trustworthy. On such a view, we would see a positive correlation between trusted individuals and trustworthiness. Appealing to Lackey (2008), I argue that this suggestion fails to account for truth-directedness. The issue is that trust will move a speaker to be trustworthy only if she is *trust-responsive*. This raises a serious problem which sees trust as having to be sensitive to trust-responsiveness. This is a problem because trust-responsiveness is dependent on a range of various factors, and an audience will either have reason to believe that a speaker is appropriately trust-responsive or she won't. If she doesn't, then it will be a matter of luck whenever she trusts a trust-responsive agent. If, however, an audience has good reason for thinking a speaker is trust-responsive, then the epistemic work is done here by the audience's evidence in support of the speaker's trust-responsiveness rather than the audience's trust.

In §VI, I introduce the second route to defending the claim that trust functions as a guide to truth, or probable truth: that trust *responds* to trustworthiness. Here, I

appeal to the work of McMyler (2011) who is sensitive to the tension between evidence and trust. Through applying McMyler's work to the level of testimonial discriminations, we can yield an account which maintains that, when we consider whether to trust someone, we weigh up evidence of their trustworthiness. The result of this will issue a response of trust or fail to do so. In this sense, trust is issued by evidence of speaker-trustworthiness. The problem with this position, I argue, is that it collapses into the evidential hypothesis. Our ability to explain the truth-directedness of our testimonial discriminations on this model has nothing to do with trust and everything to do with evidence. The upshot is that we must eliminate the assurance hypothesis from the pool of candidates under consideration.

II

In proposing his assurance view, Moran takes aim at, what he calls, the *evidential* views of testimony. In so doing, he takes issue with both *reductionism* and *non-reductionism*. The issue, so says Moran, is that these accounts neglect the "basic relationship between people" (2006: 273). In an ordinary testimonial exchange, we are *told* some proposition, *p*, by some speaker *S*, and we *believe* it (ibid.). Importantly, in such cases, we are *believing the speaker* rather than *believing what the speaker says*. The direct object of belief, as Moran puts it, is the *speaker* and not the *proposition* (or the speaker's report) (ibid.). For Moran, evidential views cannot account for this as they regard testimony purely as "*evidence* for the truth of various claims about the world" (ibid. 275).

Reductionists demand that an audience possesses sufficiently strong evidence in support of the speaker's report to attain justified testimonial belief. Hume requires evidence concerning the speaker's track-record of reporting truths (1748), while the likes of Lipton (1998) and Fricker (2017) require evidence which allows for a *testimonial inference to the best explanation* (e.g., the knowledge that the speaker knows that they stand to lose prestige through reporting a falsehood).

But this evidence simply permits the audience to *infer* that what the speaker is saying is true. This is notably different to believing the speaker *because she said it*. The object of belief here is the *speaker's testimony* rather than the *speaker*.

Non-reductionists, on the other hand, maintain that an audience is provided with a defeasible *a priori* reason to accept a speaker's report by virtue of some *a priori* connection between speaker's reports and facts. But again, this *a priori positive correlation* merely provides us with *a priori evidence* that *what the speaker says is true*. On the non-reductionist picture, the audience's testimony is still treated as evidence that what she says is true, it is simply that the audience has a "non-empirical right" rather than an "empirical right" to accept what she says (Moran, 2006: 275). But that there exists such a connection between speakers' reports and facts says nothing about the distinctiveness of coming to know that *p* from being *told* that *p*, per Moran. Insofar as non-reductionism likewise treats testimony purely as evidence in support of various claims about the world, it falls into the same category as reductionism, *viz.*, it is an *evidential view*.

If testimony just is evidence, Moran argues, it is difficult to see how acquiring beliefs through testimony is epistemically different to acquiring beliefs through *observing behaviour* (2006: 275). Consider, I might come to believe *that it is cold outside* from your testimony. I might also come to believe *that it is cold outside* from seeing you looking chilly despite wearing a warm coat. But if your testimony in this case just is evidence that it is cold outside, and my observation of your behaviour is evidence of the same thing, then wherein lies the difference? Both cases result in my belief that *it is cold outside* on the basis of evidence. And if testimony is only valuable as evidence, then we might simply view it as belonging to the more general scheme of *interpreting behaviour* (*ibid.* 275-276). On such a picture, per Moran, there is "no particular role for the notions either of a speaker *telling* someone something, or of *believing* that speaker" (*ibid.* 275).

Here, one might ask what is so problematic about a theory which treats testimonial beliefs as belonging to the more general scheme of interpreting behaviour. Perhaps a supporter of one of the evidential views might concede that acquiring beliefs through testimony just is a form of behaviour interpretation. If so, we must ask how exactly this presents a problem.

For Moran, the evidential views go wrong very fundamentally about testimony³⁵. In viewing the epistemic value of testimony as exhausted by its status as evidence, the evidentialist is unable to account for the *intentional* character of testimony. Testimony is something that we *voluntarily* and *purposefully* offer. Moreover, *we all recognize this*. But if the epistemic value of testimony is simply in its being evidence of the truth of some proposition, then the fact that the report is voluntarily offered should prove troublesome.

Moran writes that “if we are considering speech as evidence, we will have eventually to face the question of how recognition of its intentional character could ever *enhance* rather than detract from its epistemic value for an audience” (2006: 277). To put it otherwise, if testimony is just evidence of the truth of some claim, surely the fact that I *know* you are trying to induce in me some belief devalues it.

Consider the case of a criminal investigator tracking a suspect. The investigator sees the suspect drop a cigar and forms the belief that this person smokes cigars. Now suppose that the suspect was aware that they were being tracked and intentionally dropped a cigar with the intention of inducing in the investigator the belief that they smoke cigars. The evidence looks different when we suppose this. If the suspect dropped the cigar intentionally, the evidence is “contaminated by its aspect of performance”, according to Moran (ibid. 288). If, however, the suspect had dropped the cigar *unintentionally*, then it seems that the

³⁵ Though Moran does not mention the reliabilist approach to the epistemology of testimony, it seems that it would fail in the same respect. If testimonial beliefs are justified by virtue of being produced by a reliable belief-forming process, then the focus of the belief is the *testimony* rather than the speaker. As such, the reliabilist will also face difficulties in making sense of the distinction between *believing a speaker* and *believing what a speaker says*.

evidence is good. The thought here is that the suspect's intentions influence how we view the evidence. Immediately, this presents us with a problem.

Whenever we receive testimony, we know that the speaker wants us to believe the content of their report. As your audience, I know that you want me to believe that p when you report that p . But if we are viewing testimony as evidence, the inclusion of intentions makes things more problematic. This is because we are forced to deal with evidence which is *intentionally offered*. So, how could I be moved to believe that p given the fact that I know you intend for me to believe that p ? Is this not analogous to the person who drops the cigar intentionally? If so, perhaps it would be easier to go behind people's backs and gather data without worrying about their intentions to induce in us particular beliefs. In adopting such a practice, we would not be at the mercy of the speaker's sincerity and competence.

Consider again the example of coming to believe *that it is cold outside*. If I infer this from observing your behaviour, I don't have to worry about your intentions so much. You might be engaged in a performance to trick me, of course. But the likelihood of such an occurrence is low when we compare this to the ease and frequency of telling lies (or giving false testimony by an honest mistake). If, however, I come to believe *that it is cold outside* on the basis of your say-so, then I am fully aware that you are trying to get me to believe that claim. Moreover, we are both aware that I depend on you epistemically, and this dependence is not present when I infer *that it is cold outside* by observing your behaviour. Coming to believe *that it is cold outside* from your testimony therefore seems like a riskier way of forming beliefs, assuming that testimony is valuable only as evidence. But if that is the case, then we should abandon testimony as a source of knowledge, opting instead for safer methods, like interpreting behaviour.

Given this, Moran argues that "[i]f speech is seen as a form of *evidence*, then once its intentional character is recognized (that is, not just as intentional behavior, but intentional with respect to inducing a particular belief) we need an account of

how it could count as anything more than *doctored* evidence" (ibid. 277). From here, Moran concludes that the view that testimony is evidence is inherently problematic; we need an account of justified testimonial belief which recognizes the *intentional* character of testimony.

In attempting to provide such an account, Moran appeals to Grice's distinction between "letting someone know" and "telling" (1957: 382). Let's say that I want to induce in you the belief *that I have visited Niagara Falls*. Consider the following two scenarios:

- (i) I show you an unedited photograph of me by Niagara Falls.
- (ii) I draw a picture of me by Niagara Falls and show this to you.

Forming the belief that I have visited Niagara Falls in (i), does not require you to reflect on my intentions. In fact, my intention to induce in you the belief *that I have visited Niagara Falls* need play no role at all. Merely observing the photograph suffices to induce in you that belief irrespective of my presenting it to you (provided that you do not take it that the photograph has been doctored). This is not so in (ii). The drawing *by itself* will fail to induce in you the belief *that I have visited Niagara Falls*. Instead, you must recognize my intention to induce in you the belief *that I have visited Niagara Falls* and "not to be just doodling or trying to produce a work of art" (ibid. 383). In short, if you fail to recognize my intention to induce in you the belief *that I have visited Niagara Falls* then you will *not* form the corresponding belief. If, for example, you take me to be trying to produce a work of art, then you would not form the belief *that I have visited Niagara Falls*. What we can conclude from this is that your recognition of my intention to induce in you the belief *that I have visited Niagara Falls* is *necessary* for you to form the corresponding belief.

But why should your recognition of my intention to induce in you the belief that *p* result in you forming the belief that *p*? Moran appreciates this worry, writing that "recognition of the speaker's intention may seem inadequate to induce belief. It may also seem pointless, adding nothing of epistemic value to what the audience

already has" (2006: 287). To overcome this concern, Moran emphasizes the significance of the speaker's *presentation* of their testimony. He argues that "it is not, in fact, the audience's mere *awareness* of the speaker's intention that is to provide a motivation for belief. If I simply *discovered* on my own that this person had the intention that I believe P, this need not count for me as a reason for belief at all" (ibid. 288). Whilst an awareness of the speaker's intention to induce the belief that *p* is necessary, it is not *sufficient* for inducing belief. You might discover that I want you to believe that *p* after overhearing me having a conversation with a third party. Rather, what is required, is a *reason for belief* which is manifested in the act of *telling* an audience something.

On this point, Moran proposes that, when a speaker, *S*, tells his audience, *A*, that *p*, he "intends not just that the recognition of his intention play a role in producing belief that P, but that the particular role this recognition should play is that of showing the speaker to be assuming responsibility for the status of his utterance as a reason to believe P" (ibid. 290). Formulated in this way, we see that the speaker is doing something epistemically interesting when she *presents* her testimony. In delivering testimony that *p* to her audience, *A*, the speaker, *S*, thereby presents herself as *responsible* for *A* coming to form the belief that *p*.

Per Moran, when *S* offers testimony that *p* she intends to induce in *A* the belief that *p*, and further intends for *A* to recognize this intention. *A*'s recognition of that intention acts a reason for belief insofar as it demonstrates that, in reporting that *p* to *A*, *S* freely assumes responsibility for *A* forming the belief that *p*. For Moran, that assumption of responsibility is manifest when *S tells A* that *p*, thus giving *A* an *assurance*, which functions similarly to a *promise*. If I promise to ϕ then you are entitled, other things being equal, to take issue with me if I fail to ϕ . Further, in promising to ϕ , you have reason to believe that, other things being equal, I will ϕ , given that I freely assume responsibility for ϕ -ing.

On the assurance view, when I *tell* you that *p*, I give you an *assurance* for the truth of *p*. In so doing, I present myself as a “guarantor of the truth of *p*” (ibid. 279). That is, when I *tell* you that *p*, I assure you that *p* is true and present myself as accountable to you should *p* be false. According to Moran:

“When all goes well, in testimony a speaker gives his audience a reason to believe something, but unlike other ways of influencing the beliefs of others, in this case the reason the audience is provided with is seen by both parties as dependent on the speaker’s making himself accountable, conferring a right of complaint on his audience should his claim be false” (ibid. 295).

Construed in this way, we see that the *intentional character* of testimony is epistemically vital. The value of my report is not exhausted by the evidential support it gives to some proposition. If that were so, we would have to tackle the question of how my intention to induce in you some belief does not *detract* from the act of giving testimony. Rather, in recognizing my intention to induce in you the belief that *p*, you recognize that, in *telling* you that *p*, I thereby present myself as accountable to you, or as responsible for you coming to believe that *p*.

Moran maintains that “dependence on someone’s freely assuming responsibility for the truth of *P*, presenting himself as a kind of guarantor, provides [the audience with] a characteristic reason to believe, different in kind from anything provided by evidence alone” (ibid. 279). This is because, in assuming this epistemic responsibility, the reason for accepting a speaker’s report is not reducible to evidence, rather it comes from recognizing the speaker as freely accepting accountability in offering their report. As such, on the assurance view, “going behind agents’ backs” is not a better way of coming to form beliefs via testimony (ibid.); it is a characteristically different way of doing so.

But there is an issue that arises here. This issue is introduced by Faulkner, who contests that Moran’s assurance view faces a ‘dilemma’:

“The dilemma starts with the observation that a speaker can have the aim of being believed – the speaker can *tell* an audience something – even if the speaker does not intend to convey knowledge [...] Moreover, in lying a speaker equally intends an audience accept what is told because the audience recognizes what the speaker intends. Consequently, that an audience *A* recognizes that a speaker *S* intends *A* to believe that *p* will not by itself suffice to move *A* to believe that *p*.” (2011: 143).

The problem is this: a speaker can intend to *tell* an audience something, thereby intending the audience to recognize her intentions *whilst simultaneously intending to deceive her*. A deceiver presents herself as trustworthy and purports to give assurance when she attempts to deceive her audience.

Here Faulkner notes that, “*A* gains the kind of reason for belief the assurance theory is at pain to describe *only if A* takes *S*’s intentions at face value, believes *S*, or trusts *S* to tell the truth” (ibid.). In other words, for *S*’s assurance to play the role of inducing in *A* the intended belief, *A* must first trust *S* to tell the truth. After all, if I don’t take you to be sincere, how can recognition of your intentions move me to believe you? I might recognize your intentions whilst also taking it that you are trying to deceive me.

We now see the role of both speaker and audience on the assurance view. *S*, in telling *A* that *p*, intends for *A* to recognize that *S* intends to induce in *A* the belief that *p*, and that this recognition should move *A* to believe that *p* because it demonstrates that *S* freely assumes responsibility for *A*’s believing that *p*. In other words, in telling *A* that *p*, *S* gives *assurance* that *p* is true. As we have just seen, however, *S* might *purport* to offer assurance that *p* whilst intending to deceive *A*. If *A* takes *S* to be insincere whilst recognizing *S*’s intentions, *A* will not be moved to believe that *p*. As such, what is required, per Faulkner, is that “*A* takes *S*’s intentions at face value, believes *S*, or trusts *S* to tell the truth” (ibid.). My focus will be with this

last suggestion, viz., of *A* trusting *S* to tell the truth³⁶. In trusting *S* to tell the truth, *A* can thereby position herself such that she can respond to the reason for belief which comes in the form of *S*'s assurance.

In the typical case of believing that *p* from *S*'s testimony that *p*, viz., those cases whereby "one person tells a second person something, and the second person believes him" (Moran, 2006: 273), being moved to believe that *p* necessitates *A*'s trusting *S*. Without trust, *A* will not be moved to accept what she is told by *S* on the basis of *S*'s telling. This is not to imply that trust is only *psychologically* important insofar as it allows the audience to adopt the right sort of mental state such that testimonial acceptance can occur. Trust has *epistemic* importance too insofar as *A*'s trust in *S* can allow *A* to "see" an act of telling as an "assumption of responsibility" (Faulkner, 2020: 332). The precise nature of *A*'s trust is not relevant for the present discussion. What is noteworthy, however, is that trust has an epistemically significant role to play in the attainment of testimonial knowledge and justified testimonial belief, for assurance theorists.

It is also worth noting that, whilst the typical case of *A* coming to believe that *p* from *S*'s testimony that *p* might consist in *S* telling *A* that *p* and *A* believing *S*, there are atypical cases too. *A* might fail to trust *S*, but nevertheless come to believe that *p* because of evidential considerations. An example will help us here. Consider that, during the UK's negotiations to leave the EU, Boris Johnson was the foreign secretary for the UK government (albeit for a short time). In this position, he was required to visit Belgium frequently for the sake of negotiating. Despite his awful reputation for reporting truths, one might still come to believe that he was in Belgium at such-and-such time from his saying so *despite failing to trust him*. In such a case, one might reason "that's the sort of thing that the foreign secretary must do",

³⁶ I focus on trusting over the other two options included by Faulkner on the grounds that I am interested in the route to testimonial knowledge. The issue is that taking a speaker's intentions at face value implies gullibility, or at the least, a lack of attentiveness. This is the same with simply believing the speaker. As such, I find trusting the speaker to be the most promising option.

thereby relying on evidential considerations. In such a case, one does *not* believe the speaker, but merely believes what he says. The assurance position thus allows for two routes to testimonial belief: *believing the speaker* and *believing what the speaker says* (Moran, 2006: 273) (Faulkner, 2011: 137).

With this in place, we can now turn to the matter of truth-directed testimonial discriminations.

III

So far, I have been discussing the “assurance position”; a position concerned with justified testimonial belief and testimonial knowledge. According to Moran’s assurance view, when *S* tells *A* that *p*, *S* freely assumes responsibility for the truth of her report, thereby *assuring* *A* that the contents of the offered report is veridical (2006: 279). In so doing, *A* receives a “characteristic reason to believe, different in kind from anything provided by evidence alone” (ibid.). Faulkner comments that “[w]here an audience *A* has a speaker *S*’s assurance that *p*, *A* has a distinctive reason for believing that *p*, which defines testimony as a distinctive route to knowledge and warranted belief” (2011: 141). As a view concerning justified testimonial acceptance and testimonial knowledge, the assurance position can therefore be regarded as a position demanding that the *characteristic reason* for believing a speaker’s report is generated through the speaker’s assurance which contrasts to any reason for belief generated by evidence. To reject the claim that the speaker’s assurance has an epistemically significant role in the attainment of justified testimonial belief and testimonial knowledge is therefore to reject the assurance view. This motivates a unique view at the level testimonial discriminations.

However, Moran’s assurance view focuses primarily on the *speaker*. Given that I am interested in accounting for the truth-directedness of our testimonial discriminations, my focus is primarily with the audience, since it is only when one is

acting as an audience that one is making testimonial discriminations. What is needed then, is an audience condition rooted in the assurance position.

As we have seen, Faulkner maintains that “A gains the kind of reason for belief the assurance theory is at pain to describe *only if A takes S’s intentions at face value, believes S, or trusts S to tell the truth*” (2011: 143). This is because it is possible for A to regard S as untrustworthy and consequently reject what S says, despite simultaneously recognizing that S intends for A to believe that *p*. In short, without trusting S, A will not be moved by S’s assurance to form the belief that *p*.

We see that the audience, A, *must* trust the speaker, S, if she is to position herself such that she can be respond to S’s assurance. Given this, we can forward that, when A trusts S, she can respond to S’s assurance with testimonial acceptance. When A fails to trust S, however, A will not respond with testimonial acceptance. On such a position, it is trust which manifests a discriminatory capacity. Through trusting speakers, and failing to do so, we discriminate between bits of testimony.

But we are trying to explain the *truth-directedness* of our testimonial discriminations. We therefore need to add something else to this account before we can understand how trust might play a role in explaining the truth-directedness of our testimonial discriminations. We can do this, I suggest, by proposing that there is something about trust that guides us to the truth. It is connected to the truth such that, when we trust speakers, we are sensitive to the truth of what is said. Call this the *Truth-Connection Requirement*:

(TCR) Trust is connected to the truth such that it can function as a guide to truth, or probable truth.

TCR demands that trust be connected to the truth. If we can satisfy TCR, then we can generate the assurance hypothesis *that our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said*. The underlying thought here is that, if trust is connected to the truth, it can act as a sort of guide or bridge to it. Insofar as trust can perform that function,

discriminating between bits of testimony through trusting speakers would see our discriminations as truth-directed.

It is vital for the success of the assurance hypothesis that trust act as guide to the truth. If it is not connected to the truth, then it is entirely unclear how, through trusting a speaker, an audience could acquire knowledge through accepting what the speaker says. So, how might one accommodate TCR whilst utilising the assurance framework? I suggest that trust can function as a guide to truth or probable truth if it captures *trustworthiness*.

IV

If the assurance hypothesis is right, then it is through trusting speakers that our testimonial discriminations are truth-directed. As such, it is important that we have some understanding of what trust is and what it involves. The first thing to appreciate is that trusting involves, as Faulkner notes, a “three-place relation”: A trusts S to ϕ (2020: 332). For example, A might trust S to drive her to the airport by 10am. My focus is with the testimonial domain, in which A’s trusting S can be regarded as A trusting S *for the truth of S’s report* (ibid.), (Faulkner, 2007: 880), (Hieronymi, 2008: 219), (McMyler, 2011: 136), (Hawley, 2014: 16), (Keren, 2014: 2594). This brings us to the question of what an audience is doing when she trusts a speaker.

One suggestion is that, when A trusts S for the truth of her report, A is *relying* on S, in some sense (Baier, 1986) (Holton, 1994), (Jones, 1996), (Faulkner, 2011), (Hawley, 2014). As Faulkner highlights, however, the “trust-relation”, in which A trusts S to ϕ , contrasts with the mere “reliance-relation”, in which A relies on S to ϕ , since “trusting is necessarily willing” (2020: 332). Whilst reliance can be voluntary, it can also be *forced*. One can be forced to rely on another agent (e.g., one might, despite one’s best interests, rely on their nocturnal housemate to accept a parcel due to arrive at 8am). Trust differs insofar as it can be seen as an “attitude towards

reliance that explains why reliance is willing" (ibid.). In the testimonial scenario, *A* relies on *S* to tell the truth when *A* trusts *S* for the truth of her report that *p*. But what *explains* *A*'s reliance on *S* is that *A* trusts her. The willing attitude that *A* has towards her reliance on *S*'s word is what differentiates this from a simple case of *A* relying on *S* for the truth of *p*. Such cases of trusting a speaker for the truth of her report might be contrasted with cases of relying on a calculator to be operating as it should. Though both involve reliance on some agent or instrument offering the truth, one case is characterised by an attitude towards that reliance, viz., trust.

Another feature of trust involves *vulnerability*. That vulnerability, argues Hieronymi, is the vulnerability of *betrayal* (2008: 222). Insofar as *A* trusts *S* for the truth of *p*, *A* is vulnerable to having that trust betrayed. This can result in *A* accepting a falsehood, but it need not. *S* might, for example, report that *p* when *p* obtains, but when she has no good grounds for believing that *p*. So, in trusting *S*, *A* can be seen as taking a risk. An effort to eliminate that risk, or as Keren writes, to "take precautions" is incompatible with trusting (2014: 2605). Keren's example of the shop owner highlights this:

"[A] shop owner might leave her employee alone in the shop with a significant amount of money in the till while she goes out on some important errand. But if, before she leaves, she turns on the CCTV camera to monitor the employee's movements, then she does not really trust him. Even if she believes that he is trustworthy, or is optimistic that he will not steal, and turns on the cameras just as a precaution, it would not be correct to say of her that she trusts him not to steal" (ibid.).

This example suggests that the taking of precautions with the aim of diminishing the risk involved in trusting someone is incompatible with trusting. To trust *is* to accept risk; it is to be vulnerable.

In addition to asking what is involved when an audience trusts a speaker, we can also ask the broader question of what trust is. Here, there is a great deal of

disagreement. One thing on which various theorists agree, however, is that trusting an agent to φ is not synonymous with having evidence that they will φ :

“If I trust you, I will, for example, believe that you are innocent of the hideous crime with which you are charged, and I will suppose that the apparently mounting evidence of your guilt can be explained in some way compatible with your innocence” (Jones, 1996: 16).

“[T]rust need not satisfy either a positive or a negative evidence condition: it need not be based on evidence and can demonstrate a wilful insensitivity to the evidence. Indeed there is a tension between acting on trust and acting on evidence that is illustrated in the idea that one does not actually trust someone to do something if one only believes they will do it when one has evidence that they will” (Faulkner, 2007: 876).

“Having epistemic trust in a source of information does not amount to having evidence that the trusted source is trustworthy, and neither does trusting require having such evidence” (Kappel, 2014: 2024).

“[H]aving adequate evidence to believe something precludes the need for trust. In particular, if you have adequate evidence to believe that someone will do something, then there is no need to trust her to do it” (Marušić, 2015: 180).

Whatever trust is, it is different to having evidence that some agent will φ . So, when I trust you for the truth of your testimony, I cannot be understood *simply* as having evidence that your testimony is veridical. To maintain such a position is to neglect another important feature of trust: that we can, at least sometimes, trust in the face of evidence to the contrary (Baier, 1986), (Jones, 1996), (Faulkner, 2018), (Keren, 2020). This is not to reject entirely the idea that evidence has some part to play when it comes to trusting a speaker but is instead to say that our understanding of trust is not exhausted by this evidential component.

Lastly, trust can be understood as taking one of two forms: *doxastic* or *non-doxastic*. The theorist who takes trust to be *doxastic* maintains that, when *A* trusts *S*, *A* believes that *S* is *trustworthy* (Hieronymi, 2008), (McMyler, 2011) (Keren, 2014). On such a position, *A*'s trust in *S* can be viewed as involving the belief that *S* is trustworthy. This contrasts with *non-doxastic* accounts of trust, such as those proposed by Jones (1996) and Faulkner (2007) (2011) (2018). On Jones's account, for example, *A*'s trust in *S* amounts to "an attitude of optimism" (1996: 4). For Faulkner, *A*'s trust in *S* amounts to a *presumption* of trustworthiness (Faulkner, 2007: 883). One thing on which most theorists agree, however, is that trust involves some sort of *representation of trustworthiness*. That representation might take the form of a belief that the trusted party is trustworthy, but it might also take the form of something like a *presumption*. Going forward, I aim to be neutral between these two broad positions. In doing so, I understand trust as involving a *representation of trustworthiness*. Consequently, when I talk about *A* trusting *S* for the truth of her report that *p*, this should be understood as *A representing S* as trustworthy. One might take that representation to involve the belief that the speaker is trustworthy, but this needn't be the case.

According to the view of trust to which I appeal, *A*'s trusting *S* for the truth of *S*'s report amounts to *A representing S* as trustworthy. This raises the question of what it is to be *trustworthy*. One might argue that the trustworthy individual is the individual with the right sort of motives (Baier, 1986), (Jones, 1999) (Jones, 2012). Jones, for example, writes that "someone who is trustworthy [...] takes the fact that they are being counted on to be a reason for acting as counted on in their motivationally efficacious practical deliberation" (ibid. 66). Alternatively, one might follow Hawley in arguing that trustworthiness is not about motives, but about behaving in accordance with your commitments (2019: 76). She writes that, "a trustworthy person is good at managing her commitments: she will take care not to over-commit, to ensure that she can fulfil the commitments she does undertake"

(ibid. 77). For our purposes, it is not necessary to delve into this debate. Let us say, as an approximation, that the trustworthy testifier will be a testifier who speaks the truth.

The above definition is an approximation, but it is in keeping with the intuitive conception of trustworthiness. The trustworthy individual is the individual deserving of trust; the individual who can be trusted to make good on what they are being trusted to do. In the testimonial domain, a speaker, *S*, is being trusted by the audience, *A*, for the truth of their report. So, insofar as one can be expected to make good on what they are being trusted to do, the trustworthy speaker is the speaker who can be expected to report that *p* when *p* obtains.

But it is important to recognize that *S*'s being trustworthy in the testimonial domain is not synonymous with *S*'s reporting that *p* when *p* obtains. To see this, consider the scenario in which *S* intends to deceive *A* that *p*. She reports to *A* that *p*, believing that *p* fails to obtain. Unbeknownst to *S*, however, *p* does obtain. As such, when *S* reports to *A* that *p* with deceptive intentions, *S* unknowingly offers veridical testimony. Surely, we are not to say that *S*, in such a case, is trustworthy. Despite offering veridical testimony, *S* falls short of such an ascription. But if being trustworthy is simply a matter of speaking the truth, then wherein lies the problem?

The view that the trustworthy individual is simply the individual who offers veridical testimony clearly has limitations. As such, our understanding of trustworthiness, at least in the testimonial domain, should be seen as involving two aspects: an *internal* aspect and an *external* aspect.

The internal aspect of trustworthiness requires the trustworthy individual to *aim* at telling the truth. The trustworthy testifier doesn't stumble into the truth; it is exactly what they aim to do. As such, the trustworthy testifier must be *sincere*. This means that when *A* trusts *S* for the truth of her report, *A* represents *S* as trustworthy, and this representation contains within it the idea of *sincerity*.

The external aspect of trustworthiness requires that the trustworthy individual *succeed* in telling the truth, by and large. As such, the trustworthy testifier will, by and large, be the testifier who reports that p when p obtains. We can capture this thought by regarding the trustworthy testifier as *competent*. This means that when A trusts S for the truth of her report, A represents S as trustworthy, and this representation contains within it the idea of *competence*.

The thought here is that the trustworthy individual will not only *try* to tell the truth, but they will also *succeed* in doing so, by and large. Appealing to the assurance position, we can understand this in different language. We can also understand being trustworthy, in the testimonial domain, as a matter of *giving assurance*. Consider, when one makes an assertion, one can do so with the intention of transmitting knowledge or not. Put otherwise, when one makes an assertion, one can give assurance or merely purport to give it. One can genuinely act as a guarantor for the truth of their report, or just purport to do so. But *merely purporting* to offer assurance is not in keeping with the idea of trustworthiness. The individual who merely purports to give assurance is not an individual on whom we would want to rely. Our idea of trustworthiness is instead in keeping with the speaker who offers *genuine assurance* rather than *merely purported assurance*. The individual who offers genuine assurance intends to induce in their audience the belief that p by way of a genuine and free assumption of responsibility for the truth of what they say. In doing so, this individual can be considered *trustworthy*.

Through understanding trust and trustworthiness in such a way, we can make the argument that trust is connected to the truth. Consider again, what it is to trust a speaker is to represent her as trustworthy. And what it is to be a trustworthy speaker is to be a competent and sincere speaker who, by and large, tells the truth. As such, trust is connected to trustworthiness which is connected to the truth. Moreover, the trustworthy speaker is the speaker who gives genuine, rather than merely purported, assurance. And since what it is to give genuine assurance is to

freely assume responsibility for the truth of one's testimony, we see here another connection between trust and truth. In short, trust is connected to trustworthiness insofar as trusting an agent involves a representation of that agent as trustworthy. Since trustworthiness contains within it the idea of telling the truth, or giving genuine assurance, trustworthiness is connected to the truth. Trust is therefore connected to the truth *by way of trustworthiness*.

As previously stated, for the assurance hypothesis to account for the *truth-directedness* of our testimonial discriminations, it must satisfy the truth-connection requirement (TCR) *that trust is connected to the truth such that it can function as a guide to truth, or probable truth*. Whilst we have a connection between trust and truth by way of trustworthiness, the matter of whether trust can function as a guide to truth is not yet settled. Even if trust is connected to the truth, it might not function as a guide to the truth because, whilst trusting a speaker might involve the representation of them as trustworthy, they might be utterly untrustworthy. In such cases, trust fails to act as to guide to the truth, or probable truth given that it does not capture trustworthiness reliably. In such cases, we can say that trust is *unsuccessful* or *inaccurate* given that it fails to capture trustworthiness. For trust to function as a guide to the truth then, it must be *successful* or *accurate*, by and large. That is, we need trust to be extended to the trustworthy individual and not the untrustworthy individual, by and large. Insofar as we require this of trust, we require that trust can capture the distinction between trustworthiness and untrustworthiness.

If trust cannot pick up the distinction between trustworthiness and untrustworthiness, then it is unclear how trust can function as a guide to the truth. Consider an analogy: eyesight clearly cannot capture the distinction between *bitterness* and *sweetness*. If one were to use eyesight in attempt to establish whether a given food is bitter or sweet, then one's epistemic practice would have a high risk of error insofar as one could very easily tend towards forming *false beliefs*.

Accordingly, eyesight cannot function as a guide to bitterness insofar as it cannot discriminate between bitterness and sweetness. The same applies to the matter at hand. If trust cannot pick up the distinction between trustworthiness and untrustworthiness, then it cannot function as a guide to the truth because it may very well lead us to trust untrustworthy agents, by and large. And given that, on the current view, it is through trustworthiness that trust explains truth-directedness, a failure to be trustworthiness-directed will entail a failure to be truth-directed. As such, a successful defence of TCR will not leave open the possibility of *systematic error* in trusting trustworthy speakers.

V

Clearly, trust does not track the truth with 100% accuracy. We have all experienced cases in which our trust has been misplaced. We have all experienced cases in which we trust a speaker, only to have been the victim of deception. But we do not need 100% accuracy to give an account of the truth-directedness of our testimonial discriminations. Instead, trust need only be reliable in that it gets things right *by and large*. This will mean that, by and large, when we trust a speaker for the truth of their report, we are trusting a trustworthy speaker who is, by and large, offering veridical testimony. It will also mean that, by and large, when we withhold trust from a speaker, we are dealing with an untrustworthy speaker who is, by and large, offering false testimony. The question with which we must now deal is how trust can get things right reliably.

I see that there are two ways that trust might get things right reliably: 1) trust *causes* trustworthiness; 2) trust *responds* to trustworthiness. Let us deal with the former proposal first.

How might trust *cause* trustworthiness? The thought here is that, through placing trust in a speaker, the audience thereby generates a reason for a speaker to be trustworthy which then increases the probability that she will be trustworthy. This,

in turn, means that, in trusting a speaker, an audience can be sensitive to trustworthiness in the sense that trusted individuals generally turn out to be trustworthy individuals. If this is right, then the set of agents who are trusted should positively correlate with the set of agents who are trustworthy.

Pettit proposes the following argument in support of the claim that putting trust in a speaker can thereby generate a reason for her to be trustworthy:

“1. There are situations where an act of trust will signal to a trustee, and to witnesses, that the trustor believes in or presumes on the trustworthiness of the trustee—believes in or presumes on his loyalty or virtue or prudence—and so thinks well of him to that extent.

2. The trustee is likely to have a desire, intrinsic or instrumental, for the good opinion of the trustor and of witnesses to the act of trust.

3. The desire for that good opinion will tend to give the trustee reason to act in the way in which the trustor relies on him to act.

Conclusion. And so the trustor, recognizing these facts, may have a reason to trust someone, even when he actually has no reason to believe in the other’s pre-existing trustworthiness” (1995: 216).

The idea is that, through trusting you, I make it clear to you that I think well of you to the extent that I presume you to be trustworthy. Not wishing to upset that presumption, you are *pushed* to be trustworthy. But you are also *pulled* by the desire for my good opinion to be trustworthy. Trust can therefore be seen as having a *causal* power insofar as it *pushes* and *pulls* the trustee to be trustworthy. Through causing trustworthiness, trust can act as a guide to the truth, or probable truth, in the sense that the set of individuals who are trusted will consist mostly of individuals who are trustworthy given the power of trust to generate reasons to be trustworthy.

However, I might put my trust in somebody who doesn’t care about my good opinion in the slightest, or somebody who has no issue with failing to make good on that trust. In other words, trust might move a speaker to be trustworthy, and

therefore tell the truth on *some* but not *all* occasions. And this is because, *sometimes* a speaker might respond to the reason to be trustworthy generated by trust but sometimes they won't. In short, sometimes a speaker is *trust-responsive* and sometimes they are not.

Here, Pettit may well disagree. But it is worth noting that he expressly words his argument in such a way as to allow that some agents will lack the desire for the good opinion of others, in his second premise. One sympathetic to the position he forwards might argue that trust *necessarily* presents a reason for the trustee to be trustworthy and I am happy to concede as much. What I reject, however, is that the reason to be trustworthy is necessarily sufficient to *move* the trustee to be trustworthy. One might be trusted to ϕ and therefore given a reason to ϕ , but one might have more powerful reasons not to ϕ .

The issue that arises out of this is that trust can, at least sometimes, fail to move a speaker to be trustworthy. In response to this, Lackey revises conditions 2 and 3:

"2*. Some trustees are likely in some contexts to have a desire, intrinsic or instrumental, for the good opinion of some trustors and of some witnesses to the act of trust" (2008: 244).

"3*. The desire for that good opinion will tend to give some trustees in some contexts reason to act in some of the ways in which the trustor relies on him to act" (ibid. 245).

And if we employ conditions 2* and 3*, we are unable to reach Pettit's conclusion. The scope of the argument has been so narrowed that it refers only to *some* trustees and *some* contexts. This tells us only that *in some contexts* a speaker will respond to the reason to be trustworthy generated by trust. At the very least, this entails that trust *by itself* can't do the heavy lifting required of it. If it could, then the very act of trusting should evoke a response of trustworthy testimonial conduct *in all cases*, not just *some*. And even if we allow that the audience's trust in the speaker can play a

role in moving the speaker to prove trustworthy, this only tells us part of the story. For trust to prove effective in moving the trustee to be trustworthy, other conditions must be satisfied, such as the trustee's desire for the trustor's good opinion.

This introduces a serious problem. For trust to do the job of moving a speaker to be trustworthy, the agent in whom one puts their trust must be *trust-responsive*. So, to ensure that trust can function as a reliable guide to the truth given the causal explanation, we must ensure that there is systematic success in trusting trust-responsive agents. That is, we must ensure that the set of agents trusted, by and large, is *not* a set which contains within it a large number of agents who fail to respond appropriately to the reason to prove trustworthy generated by the audience's trust in them. If we cannot ensure this, that is, if we cannot rule out systematic error in trusting speakers who are *not* trust-responsive, then it is surely a matter of *luck* whenever we trust an agent who is appropriately trust-responsive.

We see here that, on this proposal, trust can capture trustworthiness only to the extent that it captures *trust-responsiveness*. This is because, if trustworthiness is caused by trust *only when an agent is trust-responsive*, then a failure to capture the distinction between trust-responsiveness and trust-unresponsiveness equates to a failure to capture the distinction between trustworthiness and untrustworthiness. Consider, on this account, we explain trust's ability to guide us to truth, or probable truth, by holding that it causes trustworthiness. However, trust has this effect on agents only when they are trust-responsive. As such, if trust fails to capture trust-responsiveness, by and large, then we allow that trusting may not tend towards producing trustworthiness.

In a related argument, Lackey notes that an agent's trust-responsiveness is the result of a variety of factors:

"[W]hen I decide to trust S with respect to x, my trust moving S with respect to x depends, at the very least, on S's character, S's regard for others, S's regard for me, S's attitude toward x, and S's current situation

or context. When I know absolutely nothing about S, then, my succeeding in choosing a trustee who satisfies all of these relevant factors—i.e., my success in choosing a trustee who is moved by my trust to trustworthy behavior in the context at hand—is simply a matter of *good luck*” (ibid.).

When we consider that a speaker’s trust-responsiveness is conditional on various factors to which we mightn’t have access, that an audience’s trust moves a speaker to be trustworthy is the result of good luck *whenever an audience lacks good reason to believe that the trustee will prove trustworthy*. This is because a variety of conditions must obtain for the audience’s trust in the speaker to move the speaker to be trustworthy. We see then that successfully engaging with a speaker who will respond to the reason to be trustworthy generated by an audience’s trust is lucky insofar as it would indicate that the various factors mentioned above by Lackey would be satisfied *unbeknownst to the audience*.

Here, one might point to the fact that Lackey explicitly refers to testifiers about whom we know “absolutely nothing” (ibid.). One might respond that, by and large, we are dealing with people about whom we know a good deal. Oftentimes, our interlocuters are familiar to us and we have some grounds for believing them to be trust-responsive or not. Put otherwise, we often have some *evidence* regarding the probability of a speaker’s being trust-responsive or not.

Whilst the above proposal might seem appealing, it does little to help. The problem is appreciated by Lackey (ibid.). Focusing on Pettit’s argument, which she refers to as the Reasons Generated by Trust Argument (RGTA), Lackey issues a dilemma. She writes that:

“[E]ither the relevant hearer has independent reasons for believing the speaker in question to be trust-responsive in which case beliefs acquired on the basis of her testimony may be epistemically justified, but not via an awareness of trust’s generative nature, or the hearer does not possess independent reasons for believing the speaker to be trust-responsive, in

which case beliefs acquired on the basis of her testimony may be grounded in an awareness of trust's generative nature, but they are not epistemically justified" (ibid. 242-243).

Applied to matter at hand, the issue is as follows: either the audience has independent reasons for believing the speaker is trust-responsive or she does not. If she does not, then an audience is lucky whenever she trusts a trust-responsive speaker given that she would lack grounds for believing that the speaker would be moved by her trust to be trustworthy.

Alternatively, if the audience has independent reasons, or *evidence*, for taking the speaker to be trust-responsive, then it seems to be those reasons which guide one to the truth, not trust. Consider, if my reason for thinking that you will respond to the reason to be trustworthy generated by my trust is that you've done so in the past, or I believe that you are motivated to do so, or that I think that you might lose prestige in not doing so, then it is *this evidence* which explains how I am sensitive to trust-responsiveness rather than trust itself.

If one attempts to get around the issue of trusting trust-responsive speakers through luck by requiring audiences to have evidence in support of a speaker's trust-responsiveness, evidence must play an important epistemic role for the assurance hypothesis. This is because, it would be only through the possession of evidence that an audience would have reason to regard a speaker as trust-responsive or not. And so, it would be on the basis of evidence that an audience would trust an agent or not. Consider, if I have evidence that you are not trust-responsive I will not be disposed to trust you. The problem is that now what seems to be determining my representations of trustworthiness, and so, my testimonial discriminations, is the *evidence* that I have. But if that is so, then we move away from the assurance hypothesis and simply defend the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity*

to evidence. Trust loses its explanatory power here insofar as the story we tell necessitates the inclusion of evidence³⁷.

Given all of this, the position that trust causes trustworthiness must be rejected. Not all speakers will respond to an audience's trust with trustworthy conduct. As such, the audience must have grounds for regarding an agent to be trust-responsive or not. But, as Lackey identifies, we see a dilemma here: if the audience has such grounds, then they can only be evidential. But if that is the case, then it is evidence which is doing the epistemic heavy lifting. If, on the other hand, the audience has no such grounds then, while evidence is not introducing difficulties on the account, we are forced to conclude that any instance of an audience trusting a speaker is the result of good fortune, given that the audience lacks any reason to believe that the audience will be trust-responsive and therefore moved by her trust to be trustworthy. The result here is an inability to explain the truth-directedness of our testimonial discriminations, given that this allows for systematic error.

Put otherwise, if we explain the truth-directedness of our testimonial discriminations by holding we are sensitive to trustworthiness insofar as, when we

³⁷ There is a further, separate concern raised by the notion that trust generates a reason to be trustworthy. Consider again the first premise of Pettit's RGTA: "There are situations where an act of trust will signal to a trustee, and to witnesses, that the trustor believes in or presumes on the trustworthiness of the trustee—believes in or presumes on his loyalty or virtue or prudence—and so thinks well of him to that extent" (1995: 216). My concern here relates to temporal order. We tend to think of the paradigmatic testimonial scenario as starting with a speaker, *S*, giving testimony that *p*, to *A*. After this, *A* will respond to *S*'s testimony in whichever way is compatible with the theory one defends. The issue is that the position currently under examination, that trust generates a reason to be trustworthy requires us to place something temporally prior to *S*'s giving testimony, since, for *S* to be appropriately affected by *A*'s trust in *S*, *A*'s trust in *S* must precede *S*'s testimony. The paradigmatic testimonial scenario would then be something like this: *A* regards *S* as appropriately trust-responsive. *A* trusts *S*. *S* recognizes that *A* trusts *S* and responds to that trust. *S* offers testimony that *p*. *A* responds to *S*'s testimony that *p*. Perhaps one is unmoved by this change to the paradigmatic testimonial scenario, but it is worth noting that the position currently under examination demands such a change. And whilst this testimonial scenario might not seem at all atypical when one is conversing with a friend, it does not look so typical when one considers those cases in which one's interlocuter is a stranger. In those cases, we must ask on what grounds *S* takes *A* to be signalling her trust and how *A* might go about it in the first place. Whilst I do not take such problems to be insurmountable, I do take them to require responses.

trust speakers, we generate reasons for them to be trustworthy, we must hold one of two positions: i) demand that audiences have grounds for taking a speaker to be trust-responsive; ii) allow that audiences need no grounds for taking a speaker to be trust-responsive. As we have seen, allowing that audiences need no grounds for taking a speaker to be trust-responsive fails to explain the truth-directedness of our testimonial discriminations in that it allows for systematic error in trusting trust-responsive agents. Alternatively, the assurance theorist can propose that audiences *must* have evidence in support of speaker trust-responsiveness. The issue with this is that, in giving evidence such a key role, it sees the assurance hypothesis collapse into the evidential hypothesis.

VI

Let us now consider the proposal that trust *responds* to trustworthiness. For this to accommodate the claim *that trust functions as a guide to truth, or probable truth*, it must be the case that trust is issued in response to a speaker being trustworthy or untrustworthy. In so doing, trust can capture the distinction between trustworthiness and untrustworthiness. So, how might one defend such a position?

McMyler attempts to do this by giving trust an irreducibly significant epistemic role yet allowing evidence to play a part too. He does this by distinguishing the *attitude* of trusting a person to φ from the *belief* that a person will φ , which he regards as entailed by trusting a person to φ (2011: 136). Whilst the belief that the trusted person will φ is justified on the basis of *evidence*, per McMyler, the audience's belief that p from a speaker's testimony that p is justified on the basis of *trust* (ibid. 136-137).

Trust, as Faulkner notes, can be both an *attitude* and an *act* (2011: 143). He writes "trust as an action involves depending on someone or something doing something, the attitude presupposes this dependence and is then characterized as an attitude towards this dependence" (ibid. 143-144). An act of trust might consist

in depending on your word, where this is realised as *accepting your testimony* and not trying to verify it independently (Keren, 2014: 2605). The attitude of trust, understood in the manner above, is the attitude *towards* this dependence. Throughout this chapter, I have aligned the term “trust” with the “attitude of trust” such that, it is through adopting the attitude of trust towards a speaker that an agent can be sensitive to the truth of what is said, thereby discriminating between bits of testimony in a truth-directed manner. When McMyler talks about the “attitude of trust”, this should be read as aligning with my use of the term “trust”.

Before looking at McMyler’s account in detail, it is important to recognize that he subscribes to a doxastic theory of trust. According to him:

“Trusting a person is an attitude that involves the truster's locating the trusted in an interpersonal logical space, but it is also an attitude that essentially involves taking something to be true with the aim of getting it right. In this sense, trust is very much like belief. In fact, it might as well be construed as a species of belief” (ibid. 131).

In the testimonial context, as previously stated, we can understand *A* trusting *S* as *A* trusting *S* for the truth of her report. Given McMyler’s position, *A* trusting *S* for the truth of her report will involve *A* *believing* that *S* is trustworthy³⁸.

Let’s say that you trust me and that you accept what I say. The subsequent testimonial belief you form is justified, so says McMyler, “on the basis of a particular kind of genuinely epistemic reason for belief” (ibid. 134). This, he continues, is “a distinctively second-personal reason that serves to justify the audience's belief in virtue of interpersonal relations of authority and responsibility between speaker and audience (ibid. 136). For our purposes, we can regard this second-personal reason much in the same way that Moran thinks of *assurance*. Whilst I recognize that

³⁸ McMyler also takes it that trusting a speaker seems to require believing the speaker (2011: 132-133). It is unclear whether such a strong claim is necessary, however. We might pull back from this slightly, simply understanding belief as being the “most natural response to trust”, rather than being positively required by trust.

McMyler's view differs slightly to Moran's, I align them simply for the sake of allowing for the continued use of the term "assurance".

Given his position regarding justified testimonial belief, McMyler would surely be an advocate of the assurance hypothesis. But consider his position concerning trust:

"When we deliberate about whether to trust someone, we do so by weighing the evidence of a person's trustworthiness. This evidence is assessed with respect to the degree to which it counts in favor of our adopting the attitude of trusting someone" (2011: 137).

Here we see the role of evidence on McMyler's account: evidence is what justifies my trust with respect to you ϕ -ing.

But if it is on the basis of evidence that our trust in a speaker is justified, why not think that, albeit indirectly, it is on the basis of evidence that a given testimonial acceptance is justified? McMyler appreciates this concern:

"If we see the belief that the person will ϕ involved in trusting a person to ϕ as justified simply by considerations of the person's trustworthiness, then it becomes hard to see how there is anything irreducibly interpersonal about the attitude of trust" (ibid. 137-138).

Clearly, McMyler recognizes the risk that the inclusion of evidence brings to an assurance view.

It is crucial to McMyler's account that what justifies the attitude of trust is not the same as what justifies trust-based beliefs. He writes that:

"[E]vidence of the trustworthiness of a person counts in favor of adopting beliefs on the basis of the truster's interpersonal relationship with the trusted, but it doesn't directly count in favor of the beliefs so adopted. Evidence of Mary's trustworthiness counts in favor of trusting Mary to pick up the kids today, but it doesn't directly count in favor of believing that Mary will pick up the kids today" (ibid. 137).

So, when I have evidence that you are trustworthy, I am justified in adopting the attitude of trust towards you. Importantly, however, such evidence does not simultaneously act as evidence in support of beliefs I might adopt on the basis of that trust.

Consider an analogy: when you have evidence that the lighting conditions are normal, you can rely on the visual data you receive. But the fact that you know the lighting conditions are normal doesn't justify your belief "that there is a lamp on the table". Intuitively, what justifies your belief "that there is a lamp on the table" is instead your perceptual experience. So, evidence that the lighting conditions are normal simply allows your perceptual experiences to function as justifying reasons for visual belief where they would not obviously have that same power were the lighting conditions abnormal (cf. McMyler, 2011: 137).

Applied to the testimonial domain we can say that evidence of a speaker's trustworthiness puts us in the position to respond to the speaker's assurance insofar as it allows us to trust the speaker. Without trusting the speaker, we cannot respond to her assurance and, consequently, cannot come to have justified testimonial beliefs in the characteristic manner espoused by assurance theorists. The point to emphasize here is that what justifies one's *trust* in the speaker is different to what justifies one's testimonial *beliefs* attained from the speaker. In the former case, it is evidence in support of trustworthiness which does the justifying. In the latter case, it is *assurance* which does the justifying.

McMyler's account is dealing with *testimonial acceptance* and *testimonial beliefs*, however, not *testimonial discriminations*. His argument is that the attitude of trust is justified by evidence, but that testimonial acceptance is justified by reasons from assurance. To yield an argument at the level of testimonial discriminations, we will need to capture the spirit of this position.

Let us begin by establishing whether McMyler's work be used to support the claim that trust responds to trustworthiness. I suggest this can be done by holding

that, when we are considering whether to trust a speaker, we weigh our evidence with respect to the speaker's trustworthiness. The result of this will see us either trust the speaker or not. By and large, the response issued by our considering the evidence of a speaker's trustworthiness must tend towards success. That is, by and large, when we are pushed to trust a speaker given our supporting evidence, the speaker will be trustworthy. This is necessary because if trust is issued on the basis of considering the evidence such that it often fails to capture trustworthiness, then we cannot satisfy the truth-connection requirement that trust functions as a guide to truth, or probable truth. As such, we can support the claim that trust responds to trustworthiness, or at least, the *features of trustworthiness*, by holding that a trust-response is issued on the basis of weighing evidence in support of a speaker's trustworthiness. In doing so, we can satisfy the truth-connection requirement and trust can guide us to truth, or probable truth.

If the above argument is right, then we explain the truth-directedness of our testimonial discriminations like so: in trusting a speaker, we are sensitive to the truth of what is said because what it is to trust a speaker is to represent them as trustworthy. Trustworthiness contains within it the idea of giving genuine assurance, which contains within it the idea of speaking truthfully. So, if we can reliably capture trustworthiness, we will reliably capture truth. Otherwise put, if our trust reliably captures the trustworthy speaker then, by and large, trusting speakers will allow us to be sensitive to the truth of what is said. Appealing to McMyler, one can maintain that trust reliably captures trustworthiness in virtue of it being issued by evidence supporting speaker-trustworthiness. And so, we now have an account which holds that our testimonial discriminations are truth-directed because, in trusting a speaker, we are sensitive to the truth of what is said.

But if McMyler is right, then we are committed to maintaining that trust responds to trustworthiness *in virtue of trust being issued by evidence*. This suggests

that trust is not sensitive to the features of trustworthiness, rather *evidence* is. It is here that we see a major problem with this proposal.

On McMyler's account, the attitude of trust depends on evidence given that it is issued by evidence. As such, evidence can be viewed as *causally* responsible for trust because, without the appropriate evidence in support of *S*'s trustworthiness, *A* will not adopt the attitude of trust towards *S*. If *A* does not adopt the attitude of trust towards *S*, then *A* will not be positioned to respond to her assurance and, consequently, will not come to believe that *p* from her testimony that *p*. This means that testimonial discriminations will be affected directly by the evidence one has because, if one has evidence in support of *S*'s trustworthiness, then one will trust *S*, and consequently, will accept what *S* says. If one fails to have the right sort of evidence, one will not trust *S*, and consequently, will not accept what *S* says. We see then that trust manifests a discriminatory capacity *only* insofar as it is the product of a process of weighing evidence. But if that is so, then trust just seems to indicate that one has, or doesn't have, evidence supporting speaker-trustworthiness. And if that is so, then it is hard to see how the assurance hypothesis offers an explanation which differs from the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. Instead, what we see is evidence in trust's clothing given that, on this account, evidence is what explains the ability of trust to capture trustworthiness.

For the assurance hypothesis to provide a tenable explanation of the truth-directedness of our testimonial discriminations, trust must reliably capture trustworthiness. Without this, trust fails to act as a guide to the truth. On the current picture, trust acts as a guide to the truth by virtue of it being issued by evidence. This means that, if evidence falls out of the picture, trust cannot capture trustworthiness. But if trust cannot capture trustworthiness, then we cannot explain the truth-directedness of our testimonial discriminations. The problem is that, through giving evidence the function of capturing trustworthiness, it is evidence

which has the explanatory power on this model. The consequence is that the assurance hypothesis collapses into the evidential hypothesis insofar as it is evidence that explains the truth-directedness of our testimonial discriminations.

VII

In this chapter, I focused on the assurance position. This represents a distinct position within the epistemology of testimony insofar as it claims that testimonial acceptance is justified by way of a speaker's *assurance*. From this, I worked to generate the assurance hypothesis *that our testimonial discriminations are truth-directed because, in trusting the speaker, an audience is sensitive to the truth of what is said*.

In generating the assurance hypothesis, I dealt with the question of how trust might aid us in getting to the truth. I argued that trust is connected to truth by way of *trustworthiness*. When we trust a speaker, we represent her as trustworthy. Given that trustworthiness contains within it the idea of giving *genuine* and not merely *purported assurance*, trust is connected to the truth insofar as it is connected to trustworthiness. The issue, however, is that showing trust to be connected to truth is not sufficient to explain the truth-directedness of our testimonial discriminations. This much is clear when one recognizes that, despite the connection between trust and truth, one might trust mostly *untrustworthy* speakers. What the assurance hypothesis requires for its success is for trust to be connected to the truth such that it can function as a guide to truth, or probable truth. This is the *truth-connection requirement* (TCR).

If we are to satisfy TCR, I argued, trust must be able to capture the distinction between *trustworthiness* and *untrustworthiness*. This is because, if trust is to act as a guide to the truth by way of capturing trustworthiness, it must be sensitive to cases where speakers are *not* trustworthy too. If it is not, then we cannot rule out the

possibility that, by and large, when we trust speakers, we are trusting *untrustworthy speakers*. This is not to require a faultless tracking of trustworthiness, but it is to require that trust tends towards success; it is to require *reliability*.

After this, I looked at two routes one might take in defending the claim that trust can function as a guide to truth, or possible truth: 1) trust *causes* trustworthiness; 2) trust *responds* to trustworthiness. I argued against (1) through an appeal to Lackey (2008). The issue is that trust can only move a speaker to be trustworthy if one is *trust-responsive*, and whether one is trust-responsive is contingent on myriad factors which are not obviously accessible to the audience. The problem that Lackey recognizes is that, whenever we lack information pertaining to the probability of a speaker's being trust-responsive, it turns out to be a matter of good luck that we put our trust in a trust-responsive speaker (ibid. 242-243). If, however, we have independent reasons for thinking that a speaker will be trust-responsive, then we might capture trust-responsiveness, but we only do so by way of *evidence*. So, while this position might afford us an explanation of the truth-directedness of our testimonial discriminations, it does so by agreeing with the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*.

After rejecting the claim that trust *causes* trustworthiness, I turned my attention to the claim that trust *responds* to trustworthiness. Here, I focused on the work of McMyler (2011). For McMyler, testimonial beliefs are justified by a non-evidential, interpersonal reason, while the attitude of trust is justified by evidence in support of speaker-trustworthiness (ibid. 279). Applied to the level of testimonial discriminations, we can motivate the position that, while evidence is responsible for issuing the trust we extend to speakers, it is trust itself that explains the truth-directedness of our testimonial discriminations. I argued that this view fails to answer the question of what explains the truth-directedness of our testimonial discriminations. The problem is that, on this model, an audience will trust the

speaker when she has adequate supporting evidence of speaker-trustworthiness, but not otherwise. If that is so, then, while trust might play the role of capturing trustworthiness, which thereby allows trust to function as a guide to the truth, it can only perform this role insofar as it is based on evidence in support of trustworthiness. If one removes evidence on this picture, then one is no longer able to explain how trust might capture trustworthiness and, without this, the assurance hypothesis is not explanatory.

The problem the assurance hypothesis cannot overcome is that of allowing trust to function as a reliable guide to truth, or probable truth, whilst simultaneously ensuring that it does not depend on evidence. As a result, I conclude that we must eliminate the assurance hypothesis.

5. The Evidential Hypothesis

I

In this final chapter, I argue in support of the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. A defence of this claim is necessary because, in previous chapters, I criticised other views for failing to differentiate themselves from the evidential hypothesis. As such, in this chapter I will make clear how the evidential hypothesis should be understood and show how it explains of the truth-directedness of our testimonial discriminations.

In §II, I show how the evidential hypothesis receives support, by introducing the evidentialist position concerning epistemic justification (Feldman and Conee, 1985) (2004). In doing so, I aim to show how evidentialism generates the evidential hypothesis. In §III, I focus on the question of what it is for evidence to *support* a given hypothesis. Here, I make a crucial distinction between two different, and incompatible, conceptions of the evidential hypothesis: the *internalist evidential hypothesis* and the *externalist evidential hypothesis*. I argue against the internalist evidential hypothesis. The issue, I maintain, is that if what explains the truth-directedness of our testimonial discriminations is our sensitivity to evidence, then evidence must be *truth-conducive*; it must act as a bridge to truth or probable truth. Given that the internalist evidential hypothesis permits evidence to support a hypothesis despite not increasing the objective probability of the hypothesis, it is entirely unclear how demonstrating a sensitivity to evidence so construed could explain why our testimonial discriminations are truth-directed.

Next, in §IV, I continue my discussion of evidential support, turning my attention to the *externalist evidential hypothesis*. In doing so, I introduce the work of Williamson (2000), who supports the claim that a piece of evidence, *e*, supports

some hypothesis, h , when the probability of h obtaining, given e , is greater than the probability of h alone.

In §V, I tackle the question of what counts as evidence for an agent. In doing so, I recognize two distinct ways of talking about evidence. We can talk about something, e , *being* evidence, but we can also talk about e being evidence for some agent, S . I bring out this contrast by introducing the work of Achinstein (1978)(2001) who deals with the former notion of evidence, before focusing on Williamson, who is primarily concerned with the question of what counts as an agent's evidence. This leads me to discuss Williamson's E=K thesis, which proposes that "knowledge, and only knowledge, constitutes evidence" (2000: 185).

After this, in §VI, I aim to provide a sketch of how I understand truth-directed testimonial discriminations given the externalist evidential hypothesis. This is done for the sake of providing a more detailed picture. It is important to recognize, however, that I do not commit myself to this position, nor does the view I defend entail the picture I sketch.

I will then respond to three major objections, in §VII. First, I revisit the empirical data gained from studies into deception detection, asking, if we are so good at demonstrating a sensitivity to evidence, how do we fare so poorly in the experiments? Second, I respond to the position that we must be *aware* of our evidence. How can we be sensitive to evidence if we are not aware of it? The final objection to which I respond concerns the evidential support condition. Through appealing to Achinstein (1978), I revise the evidential support condition thereby overcoming some issues Achinstein raises.

After having responded to these objections, I will conclude that the externalist evidential hypothesis is the strongest explanatory hypothesis examined. As such, I defend the view that *our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence.*

II

Evidentialism is a position concerning epistemic justification holding (roughly) that an agent's belief that p is justified when they have evidence which, on balance, supports p (Conee and Feldman, 2008: 83). According to this view, what justifies my belief, say, *that there is a red book on my desk* is my supporting evidence that there is a red book on my desk which I gain from my visual experience of there being a red book on my desk. If I went into my study and had no such experience, then I would lack supporting evidence for that claim and, supposing I nevertheless believed *that there is a red book on my desk*, would have an *unjustified* belief.

The evidentialist requires that, for an agent's belief that p to be epistemically justified, she must *have* evidence which supports p . But, as Feldman and Conee note, merely *having* evidence isn't sufficient for knowledge; we must also *use* the evidence we have (1985: 24). To understand why this is so, take the case of Sonya. Sonya is a police detective investigating a murder. She believes that the butler did it, and she has plenty of good evidence for this belief ranging from fingerprint analysis to the butler's motive. But let us imagine that, despite having good evidence for her belief *that the butler did it*, she does not hold that belief on the basis of that evidence. Let us imagine instead that Sonya believes *that the butler did it* because he has "an evil-sounding name". In this case, whilst Sonya *has* good evidence which, on balance, supports the claim *that the butler did it*, she does not *use* this evidence insofar as her belief is not based on that evidence.

The distinction just mentioned, between *having* and *using* evidence, captures the distinction between, what Feldman and Conee call, *justified belief* and *well-founded belief*. As we have seen, they regard an agent's belief that p as justified when that agent has evidence which, on balance, supports that p (2008: 83). Of well-founded belief, they write that:

"S's doxastic attitude D at t toward proposition p is well-founded if and only if

- (i) having *D* toward *p* is justified for *S* at *t*; and
- (ii) *S* has *D* toward *p* on the basis of some body of evidence *e*, such that
 - (a) *S* has *e* as evidence at *t*;
 - (b) having *D* toward *p* fits *e*; and
 - (c) there is no more inclusive body of evidence *e'* had by *S* at *t* such that having *D* toward *p* does not fit *e'*" (1985: 24).

To have a well-founded belief, one's doxastic attitude (for example, one's state of *belief* towards some proposition *p*), must be justified *and* be held on the basis of supporting evidence. As such, whilst Feldman and Conee would regard Sonya as having a *justified belief*, they would not regard her as having a well-founded belief. And given that they require well-founded belief for knowledge³⁹, Sonya's belief cannot amount to knowledge even supposing it is true. As Turri puts it "[b]elieving for a good reason is a valuable state, more valuable than merely having a good reason" (2011: 383).

One might disagree, arguing that Sonya clearly fails to have a justified belief. After all, she does go wrong in a sense. In taking such a stance, one might prefer to use the terminology of Firth, who distinguishes between *propositional* and *doxastic* justification (1978: 219-220). Applying this terminology to the case of Sonya, we might say that her belief *that the butler did it* is *propositionally justified* but not *doxastically justified*. And this is because, she *has* evidence which supports the proposition *that the butler did it*. The issue, however, is that her belief is not held on the basis of supporting evidence; she does not *use* her evidence. Throughout this thesis, I have been talking about *doxastic justification* when discussing justified belief. Rather than switch to talk of *justified* and *well-founded* belief now, I will continue to talk of *justification* with doxastic justification in mind. Continuing in this fashion, I take it that Sonya fails to have a justified belief insofar as her belief is not *doxastically justified*. Whilst her belief is *justifiable* in the sense that it is

³⁹ Conee and Feldman write that "[k]nowledge requires well-founded belief" (2008: 83).

propositionally justified, it is not *doxastically* justified as Sonya does not base her belief on the evidence she has. In other words, she *has* supporting evidence, but does not *use* it.

As a theory concerning *doxastic justification*, evidentialism can be seen as maintaining that an agent's belief that *p* is justified when: 1) that agent has evidence which, on balance, supports that *p* and 2) that belief is held on the basis of such evidence. Evidentialism, therefore, gives evidence a key role in conferring justification onto belief. However, to base one's belief on the evidence, one must first be *sensitive* to it. Through appreciating this, we can generate the evidential hypothesis.

For the evidentialist, a belief is justified when it is held on the basis of supporting evidence. Let's say you form the belief *that it is raining outside* because you hear what sounds like rain against your window. In this case, your belief is justified because you base it on your evidence which, on balance, supports the proposition *that it is raining*. The basis of your belief here is the auditory evidence you have when you experience the sound of the rain hitting your window. In other words, the reason *for which* you believe *that it is raining* is the evidence you acquire when you have the auditory experience of the rain hitting your window (Kelly, 2002: 179), (Turri, 2011: 385), (Rinard, 2019: 763), (Titus & Carter, 2023: 3). In this sense, we can think of evidence as having the power to *influence* or *affect* us insofar as we can revise our doxastic attitudes in accordance with it. When you have the experience of the rain hitting your window, your doxastic attitude towards the proposition *that it is raining* is revised in accordance with the auditory evidence you acquire. You go from a state of *disbelief* towards the proposition *that it is raining*, to a state of *belief*. *Other things being equal*, if you failed to have the auditory experience of rain hitting your window, you would likewise fail to form the belief *that it is raining*. If you were to form the belief *that it is raining* without any such basis, your belief would be *unjustified*, per evidentialism.

Now, let's say that a person with profound deafness is in the same room as you when you gain supporting evidence *that it is raining*. A person with profound deafness hears little to nothing. As such, this agent would fail to form the belief *that it is raining* on the basis of auditory evidence given that *they are not sensitive to auditory evidence*. The issue here is not that there is no supporting evidence that it is raining. Provided your hearing is not impaired, you have such evidence. The issue is instead that the person with profound deafness is not *sensitive* to such evidence. Due to their insensitivity to auditory evidence, this person cannot respond to such evidence in the sense that they cannot revise their doxastic attitudes in accordance with it, nor can the auditory evidence serve as a basis for their beliefs. To do so requires a *sensitivity* to evidence.

In short, to use a piece of evidence, *e*, as the basis of your belief that *p*, you must have the capacity to use *e* as the basis of your belief that *p*. Further, for *e* to influence you in the sense that your doxastic attitude towards *p* can be revised in accordance with *e*, you must have the capacity to be influenced by *e*. This is all to say that you must be *sensitive* to *e*. Given that the evidentialist gives evidence a key role in explaining the justificatory status of our beliefs, the evidentialist is committed to demanding that a necessary condition on the attainment of justified belief (and knowledge) is that we demonstrate a sensitivity to evidence. If we are not sensitive to evidence, then it is entirely unclear how we could ever respond to it in the manner the evidentialist requires.

From this, it becomes clear how the evidentialist position concerning epistemic justification motivates the evidential hypothesis. For the evidentialist, an agent's belief that *p* is justified when it is based on evidence which, on balance, supports that *p*. As we have seen, this necessitates a sensitivity to evidence. The evidential hypothesis seeks to explain the truth-directedness of our testimonial discriminations by appealing to this sensitivity to evidence. The task ahead of us

now is to detail how exactly the evidential hypothesis accounts for the truth-directedness of our testimonial discriminations.

III

The evidentialist subscribes to the thesis that an agent's belief that p is justified at some time, t , when that agent believes that p on the basis of evidence which, on balance, supports that p at t . Given its key role, the manner in which one understands evidence will greatly impact the kind of theory one defends. One might defend the claim that evidence supports a hypothesis only when it *entails* the hypothesis. On such a view, one is committed to ruling cases in which an agent believes that p because she mistakenly believes that her evidence entails that p as cases of *unjustified belief*. Alternatively, one might opt for a more permissive view, according to which evidence supports a hypothesis simply when an agent takes it that the hypothesis is made more likely given the evidence. On this view, by contrast, believing that p because you mistakenly believe that your evidence entails that p will allow for justified belief. Given the differing conceptions of evidence available, it is critical to our understanding of the evidential hypothesis that we specify what evidence must be in order to play a role in explaining the truth-directedness of our testimonial discriminations.

If our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence, it must be the case that evidence is, in some way, connected to the truth. It must act as a sort of bridge to the truth. This claim should be familiar from the previous chapter, in which I argued that trust must be connected to the truth. The reason behind this claim is that, if one is to account for the truth-directedness of our testimonial discriminations by appealing to something like trust, or in this case, evidence, then it must be that the thing to which we are appealing can succeed, by and large, in getting us to the truth. If evidence is not connected to the truth, then it is hard to understand how demonstrating a sensitivity

to it could possibly assist us in getting to the truth. To explain the truth-directedness of our testimonial discriminations by way of *evidence* demands that evidence is connected to the truth in some way. The next task is to evaluate the conceptions of evidence which fail in this respect and to understand why that is the case. In doing so, I will proceed by first examining the broadly *internalist* conception of evidence. We can use this to generate, what I will call, the *internalist evidential hypothesis*.

Conee and Feldman support an *internalist* evidentialist position (1985) (2004) (2008). In doing so, they defend the following two theses:

S “The justificatory status of a person’s doxastic attitudes strongly supervenes on the person’s occurrent and dispositional mental states, events, and conditions” (2004: 56).

M “If any two possible individuals are exactly alike mentally, then they are alike justificationaly, e.g., the same beliefs are justified to the same extent” (ibid.).

In supporting these two theses, Conee and Feldman are committed to the view that the justificatory status of an agent’s belief is determined by *internal factors* only (given (S)). Otherwise put, they maintain that the status of a belief as justified or otherwise is determined purely by an agent’s *mental states, events, and conditions*; no external factors play a role in determining the justificatory status of an agent’s belief. As Conee and Feldman are evidentialists, they maintain that “a person’s doxastic justification is a function of the evidence that the person has” (2008: 88). So, their subscription to evidentialism in conjunction with their defence of (S) commits them to the view that evidence consists in an agent’s mental states, events, and conditions. That is, evidence exists *internally only*. Whatever it is that justifies belief, it is not something like a fact about the world. Rather, it is something *mental*.

Here, we can ask for further detail; we can ask “what sort of thing counts as evidence?”. One suggestion, to which Conee and Feldman appear sympathetic, is that evidence consists in one’s *experiences* (ibid.). But experiences come in many different shapes and sizes. Contrast the experience of smelling a rose with the

experience of guilt for forgetting your friend's birthday. Though both qualify as experiences, they clearly differ a great deal. So, we can ask the question of *which* experiences count as evidence. In doing so, however, we set a challenge of sorts to any theorist who wishes to exclude a subset of experiences. Say you want to defend the position that only *perceptual experiences* count as evidence. If you aim to do this, you must give an argument to that end. You must say what is special about perceptual experiences such that only that type of experience can qualify as evidence.

And there is the further concern of whether one's evidence must be *conscious*. When you have an experience of smelling a rose, you are fully aware of it. You smell the rose, and you are conscious of this experience which supports the proposition *that this flower is a rose*. One might, after reflecting on examples such as this, feel pushed to hold that evidence *must* consist in conscious experiences; that the kind of evidence which supports our beliefs, thereby justifying them, is *conscious*. Indeed, this is the view Goldman has in mind when he introduces the problem of *stored beliefs* (1999: 278). The issue is that we have a wealth of beliefs stored in our memories. Whilst those beliefs are not currently at the forefront of our minds, we are able bring them into focus. Nevertheless, a great many of those beliefs equate to knowledge and so, must be justified. Goldman's challenge is that, if beliefs are justified only by *conscious states*, then those stored beliefs cannot be deemed as justified (*ibid.*). The evidentialist is not without response, however. One can hold, as suggested by Conee and Feldman, "that other non-occurrent internal states can contribute to the justification of non-occurrent beliefs" (2004: 67). Otherwise put, one need not defend the claim that it is *only* conscious states which justify beliefs (*ibid.*). I will not delve into this discussion any further, however. Whilst I recognize that there is a debate here, the outcome of it has no bearing on the arguments I present.

Given their defence of (M), we see that, for Conee and Feldman, it is *only* one's mental life which dictates the extent to which one is justified in believing that p ⁴⁰. Defence of (M) commits one to holding that, two different agents in different possible worlds who are "mental duplicates" will have the same evidence *given that evidence consists in one's mental life*. As such, the doxastic attitudes of these agents will be justified to the same degree given that evidence justifies belief.

Through supporting (M), Conee and Feldman defend a version of internalism known as *mentalism*. This is the view that the factors which confer justification onto an agent's belief that p must be part of the agent's mental life. Mentalism contrasts with *accessibilism* which instead holds that the factors which confer justification onto an agent's belief that p must be *accessible* to some extent (ibid. 55-56). Though there is disagreement between advocates of mentalism and advocates of accessibilism, both positions are internalist insofar as they are committed to the thesis that no *external* factors have a role to play in determining the justificatory status of an agent's belief that p . Rather, the factors responsible for conferring justification onto an agent's belief that p must be *internal*. Call this the *internalist thesis*.

Here, I will note that there are complexities contained within this characterisation of the internalist thesis. To see this, consider that McDowell (1995) and Williamson (2000) defend views according to which evidence can be viewed as consisting in agents' internal states. However, those *internal* states are, in a sense, governed by *external* factors. McDowell, for example, takes it that an agent's perceptual belief, say, "that there is a red book on the desk" is justified by the *factive* state of *seeing-that* (1995: 880). But notice that, for an agent to *see that p*, p must be the case. The state of *seeing that p* contrasts with the state of *it appearing that p* in the sense that the former depends on the world being a certain way, whereas the

⁴⁰ Note here that, in using the term "mental life", Conee and Feldman do not commit themselves to defending the claim that only *conscious* states have the power to justify beliefs.

latter does not. (ibid.). For Williamson, beliefs are justified by an agent's *evidence*, which he equates with an agent's knowledge (2000: 184). Insofar as one has evidence, for Williamson, one has *knowledge*. But notice that an agent might *believe* they have knowledge when they do not. As such, the state of *knowing* that p depends on external factors in a way that the state of *believing that p* doesn't. The internalist thesis should be viewed as ruling out such positions insofar as these accounts give external factors an epistemically significant role in the attainment of justification-conferring mental states.

We now have a better understanding of what counts as evidence, according to the internalist. The internalist must take evidence to be something *mental*, like an *experience* or a *belief*. We have seen that, for Conee and Feldman, the set of one's evidence includes only those items which are a part of one's mental life, whether they are accessible or not. But whilst this might help us in answering the question of what counts as evidence, it does not help us in determining how evidence can get us to the truth. To answer that, we need an understanding how a piece of evidence supports a given hypothesis.

There are a range of views one might defend regarding evidential support. One might hold that a piece of evidence, e , supports a hypothesis, h , for an agent, S , when it "just seems" to S that e entails h , or makes h more probable, for example. It is far beyond the scope of this work to assess all proposals here⁴¹. As such, I will focus on Conee and Feldman's *best explanation* account, as I judge this to be the most promising internalist proposal.

Conee and Feldman defend an account of evidential support which relies on principles of best explanation. They write:

"Perceptual experiences can contribute toward the justification of propositions about the world when the propositions are part of the best explanation of those experiences that is available to the person. Similarly,

⁴¹ See Conee and Feldman (2008: 83-104) for an illuminating discussion of evidential support.

the truth of the contents of a memory experience may be part of the best explanation of the experience itself. Thus, the general idea is that a person has a set of experiences, including perceptual experiences, memorial experiences, and so on. What is justified for the person includes propositions that are part of the best explanation of those experiences available to the person. Likewise, one's inferences justify by identifying to one further propositions that either require inclusion in one's best explanation for it to retain its quality or enhance the explanation to some extent by their inclusion" (2008: 97-98).

On this *best explanation* view of evidential support, some piece of evidence, *e*, supports a hypothesis, *h*, when *h* is part of the best explanation of *e* available to a given agent. So, my evidence *that there is water running down the window* supports the hypothesis *that it is raining* because the hypothesis *that it is raining* is part of the best available explanation of the evidence *that there is water running down the window* for me.

Conee and Feldman state that *e* supports *h* when *h* is *part* of the best explanation of *e*. The reason for requiring that *h* merely be *part* of the best explanation is that, per Conee and Feldman, "[t]he best available explanation of one's evidence is a body of propositions about the world and one's place in it that make best sense of the existence of one's evidence" (ibid. 98). The important detail to appreciate here is that the best explanation of one's evidence is a *body* of propositions. Here, one might ask: is it not the case that the best explanation of the evidence *that there is water running down the window* is exhausted by the proposition *that it is raining*? In response to this, an advocate of the best explanation view might reply that the best explanation of that piece of evidence must also include propositions such as *when it rains, water can fall on windows; water is the sort of thing that can run down windows; there are rainclouds overhead*; and so on. The *best* explanation then, is a body of propositions which, taken together, "make

best sense of the existence of one's evidence" (ibid.). The idea here is that the task of making sense of the existence of one's evidence requires a *set*, or *body* of propositions. As such, the proposition *that it is raining* functions only as *part* of the best explanation of the evidence.

Note also that Conee and Feldman speak not of "the best explanation", but of "the best *available* explanation". This is to rule out explanations which include propositions the agent does not have as evidence or, perhaps, is not currently able to use. They write:

"It may be that the best scientific explanation of a person's current experiences includes detailed scientific theories or distant logical consequences that the person does not understand. Such an explanation is not available to the person" (ibid.).

It may be that the best explanation of some piece of evidence includes information of which the agent in question is entirely unaware.

Consider the case of Simon. Simon believes that he has shingles because he has a particular kind of rash. He believes that only shingles could cause this kind of rash. As a result, it seems that Simon has supporting evidence for the proposition that he has shingles. But suppose that Simon is wrong; he has measles, not shingles. Let us add further that Simon has never heard of measles. As such, the hypothesis *that I have shingles* is the best explanation *available* to Simon because he does not have other evidence which rivals the explanatory power of this hypothesis. Clearly, the hypothesis *that I have measles* is better insofar as it accurately reflects Simon's condition, but it is unavailable to him insofar as it is not something to which he can appeal. As such, the hypothesis *that I have measles* cannot be part of the best explanation for Simon *at that time*, per Conee and Feldman. The best *available* explanation should therefore be regarded as the best explanation *for an agent* given the body of evidence they have *at that time*. In this case, we see that Simon's body of evidence does not include *that measles is an infectious disease, that*

measles can cause rashes, and so on. As such, the proposition *that I have measles* cannot feature as part of the best explanation of his evidence that he has a rash.

If a piece of evidence supports a hypothesis when the hypothesis is part of the best *available* explanation of the evidence, this allows that what best explains my evidence, *e*, at some time, t_1 , might be different to what best explains *e* at a later point in time, t_2 . This is because I might have an explanation available to me at t_2 which is not available to me at t_1 . I might have further data at t_2 which allows me to appeal to an explanation which was previously unavailable. Further, it suggests, for similar reasons, that what best explains *my* evidence at t_1 might be different to what best explains *your* evidence at t_1 . Consider the case above once more. Whilst Simon is entirely ignorant of measles and its symptoms, you might not be. Faced with the same evidence, viz., the visual experience of Simon's rash, you might have an explanation available to you which is unavailable to Simon, viz., that Simon has measles (rather than shingles). And so, in the same situation, what best explains the evidence is different given your differing *background beliefs*. This tells us that whether a hypothesis can feature as part of the best available explanation depends on the background beliefs of the agent in question.

Now that we have a better understanding of how the internalist might conceive of the evidential support relation, let us return to the question of whether the internalist evidential hypothesis can explain the truth-directedness of our testimonial discriminations. To explain the truth-directedness of our testimonial discriminations, an explanatory hypothesis must account for our ability to discriminate between bits of testimony in such a way as to ensure systematic success in acquiring knowledge via testimony. Recall, in chapter one, I argued that we should understand truth-directed testimonial discriminations as a matter of doing what the reasonable person would do to get to the truth by way of a reliable cognitive ability. Reliability is built into truth-directed testimonial discriminations and so, any explanation of the truth-directedness of our testimonial discriminations

must accommodate that property. If the explanation to which we appeal allows for systematic failure, in that it allows that we could tend to get things wrong in making truth-directed testimonial discriminations, then it will fail to explain the truth-directedness of testimonial discriminations insofar as it fails to accommodate the property of reliability.

Furthermore, if the explanation to which we appeal allows for systematic failure, then, on such a view, any true belief we attain from testimony would be the result of veritic epistemic luck⁴² (Engel, 1992: 57) (Pritchard, 2005: 136). This is because, if there is nothing about the epistemic practice in question which explains our tendency to get things right reliably, then it would be a matter of luck that relying on such a practice would lead to success. As such, it cannot be that the explanatory hypothesis to which one appeals in accounting for the truth-directedness of our testimonial discriminations allows for systematic failure.

If one adopts the best explanation account from Conee and Feldman (2008), holding that a piece of evidence supports a hypothesis when the hypothesis features as part of the best available explanation of an agent's evidence, then we will need some detail as to how evidence helps direct one to the truth or probable truth. That is, we need to know how the best explanation account of evidential support manages to ensure systematic success in arriving at the truth through a sensitivity to evidence. Here, I suggest, the internalist hits a wall.

We have seen that, for Conee and Feldman, "[t]he best available explanation of one's evidence is a body of propositions about the world and one's place in it

⁴² I will concede that this claim must be indexed to our world. We can of course imagine possible worlds in which people always tell the truth. In such a world, gullibility is not problematic. In such a world, accepting whatever you are told will still see you gain true beliefs reliably. However, the explanation of *why* this is so has nothing to do with the epistemic practice you use. Rather, it comes down to the fact that speakers in this environment only report truths. In such an environment, discriminating between bits of testimony in a truth-directed manner is not necessary. By contrast, as we live in a world wherein speakers sometimes report the truth and sometimes report falsities, we must discriminate between bits of testimony in a truth-directed manner if we are to attain testimonial knowledge and justified testimonial belief.

that make best sense of the existence of one's evidence" (2008:98). They write further that:

"This notion of making sense of one's evidence can be equally well described as fitting the presence of the evidence into a coherent view of one's situation. So it may be helpful to think of our view as a non-traditional version of coherentism" (ibid.).

This is an important feature of their account because it indicates that what makes some explanation "the best" is its power to fit the agent's evidence into a coherent view of their situation. Admittedly, however, it is unclear what a "view of one's situation" is.

Return to the case of Simon, who believes he has shingles. For Simon, the hypothesis that he has shingles is part of the best available explanation of his evidence insofar as it helps to fit the presence of his evidence, that he has a rash, into a coherent body of propositions or "view of his situation". That is, the explanation which includes within it the proposition that Simon has shingles helps him to make sense of his evidence. Assuming that you are aware of measles and its symptoms, then the hypothesis *that Simon has shingles* does not fit into as coherent a body of propositions for you, relative to the hypothesis *that Simon has measles*. This is because you accept a range of propositions about what measles is and how it affects the body. As such, the hypothesis *that Simon has measles* helps to fit the evidence *that Simon has a rash* into a more coherent set of propositions *for you*. Given that the hypothesis *that Simon has shingles* does not help to fit the evidence *that Simon has a rash* into as coherent a body of propositions for you, that hypothesis is not as well supported. This idea then, of "fitting one's evidence into a coherent view of one's situation" (ibid.) can also be thought of as accommodating one's evidence in such a way as to allow it to sit coherently with a body of propositions one takes to be true. This position therefore prioritises coherence, for it is the most coherent explanation which is deemed "the best".

But why think that the most coherent explanation will do anything to get us to the truth or likely truth? If evidence supports a hypothesis when the hypothesis is part of the best available explanation of the evidence, and the best explanation of an agent's evidence is a body of propositions which best coheres with one's evidence, then it must be the case that coherence is truth-directed. I aim to show, however, that adopting this view of evidential support allows for a systematic failure in making truth-directed testimonial discriminations simply because coherence and truth (or likely truth) *need not* be associated.

Klein and Warfield argue that "by increasing the coherence of a set of beliefs, the new, more coherent set of beliefs is often less likely to be true than the original, less coherent set" (1994: 129). This is problematic because it suggests that an increase in the coherence of a set of beliefs need not bring about an increase in the likelihood that the set of beliefs is true. In fact, they argue the opposite: that an *increase* in coherence through the addition of a belief, or beliefs, which are not certain or entailed by other members in that set, will thereby bring about a *decrease* in the probability of the truth of that set. They aim to show this with the case of a detective investigating "Dunnit".

Imagine that at some time, t_1 , the detective has a set of beliefs: $\{B_1, B_2, B_3\}$. Later, at t_2 , the detective has added a new belief to that set, which yields the set: $\{B_1, B_2, B_3, B_4\}$. We can understand his set of beliefs at t_2 like so:

" B_1 : Dunnit had a motive for the murder.

B_2 : Several credible witnesses claim to have seen Dunnit do it.

B_3 : A credible witness claims that she saw Dunnit two hundred miles away from the crime scene at the time the murder was committed.

B_4 : Dunnit has an identical twin whom the witness providing the alibi mistook for Dunnit" (Shogenji, 1999: 341).

At t_1 , the detective has a set of beliefs which does not contain B_4 . At t_2 , however, this changes, as he adds it. In adding B_4 to the set $\{B_1, B_2, B_3\}$, the detective's set of beliefs

hang together better. The uncomfortable tension between B_2 and B_3 in the set $\{B_1, B_2, B_3\}$ is heavily reduced when the detective adds B_4 . But, as Klein and Warfield state, "since the new set contains more beliefs than the old set and the added beliefs neither have an objective probability of 1 nor are they entailed by the old set, the *more* coherent set is less likely to be true than is the original, *less* coherent set" (1994: 131). In short, a larger set is more likely to have a false belief in it relative to a smaller set.

Consider the larger set of beliefs $\{B_1, B_2, B_3, B_4\}$. For this set to contain *only* true beliefs, more is required than for the set $\{B_1, B_2, B_3\}$ to contain *only* true beliefs. The truth of the set $\{B_1, B_2, B_3, B_4\}$ requires the truth of the set $\{B_1, B_2, B_3\}$ *plus* the truth of B_4 . When one recognizes that one can increase the coherence of a belief set by adding beliefs which help to make that set more coherent, the issue is clear. Insofar as the detective adds B_4 to $\{B_1, B_2, B_3\}$, the detective thereby increases the coherence of his belief set while simultaneously decreasing the likelihood that the set contains only true beliefs precisely because he is increasing the size of that belief set. From this, Klein and Warfield take it that "coherence, *per se*, is not truth conducive" (ibid. 129).

In response, Merricks has argued that Klein and Warfield go wrong insofar as they evaluate the truth conduciveness of coherence "on the level of systems or sets of beliefs" (1995: 307). Instead, Merricks suggests, we should be looking at the level of *particular beliefs*, asking whether increased coherence makes a particular belief more or less likely rather than whether it makes a particular *set* more or less likely. Underlying Merricks's approach here is his support of the thesis that "[j]ustification is truth conducive just so long as, *ceteris paribus*, the more justified a particular belief is, the more likely it is that that belief is true" (ibid.). And so, if greater coherence fails to entail that a particular belief is less likely true, Klein and Warfield cannot reach their conclusion, so argues Merricks.

Consider again the case of Dunit. Merricks allows both that the set $\{B_1, B_2, B_3, B_4\}$ is more coherent than the set $\{B_1, B_2, B_3\}$, *and* that it is more likely to contain a false belief. However, Merricks insists that this tells us nothing about any *particular belief*. What we need to do is to look at whether any particular belief is made *less likely* when it is part of the more coherent set $\{B_1, B_2, B_3, B_4\}$ (ibid. 308). In testing this, let us introduce:

B_5 : *Dunit committed the murder.*

Does the addition of B_4 to the set $\{B_1, B_2, B_3, B_5\}$ thereby make B_5 less likely? It is seemingly the case that the *more coherent* set makes B_5 more likely. The conclusion Merricks draws is that “Klein and Warfield show only that one can increase the coherence of a system of beliefs while, at the same time, increasing the probability that the system contains a false belief. But this does not entail that coherence is not truth conducive” (ibid. 309).

The debate concerning the truth-conduciveness of coherence is lively and complex. While some theorists strongly disagree with Klein and Warfield (see Merricks (1995), Cross (1999), Shogenji (1999) (2001b)), others passionately defend their conclusion (see Klein and Warfield (1996), Akiba (2000), Olsson (2001) (2002)). I will not take a particular stance on this issue. For my argument against the best explanation account offered by Conee and Feldman (2008) to succeed, I need only reach the claim that *coherence can come apart from the truth, thereby allowing for systematic error*. Notice that this claim is weaker than that of Klein and Warfield (1994), in that it allows one to hold that coherence *can* be an indicator of truth, albeit a weak indicator.

If all that is necessary for evidence to support a hypothesis is that the supported hypothesis is part of the best available explanation of the agent’s evidence, which can be further understood as belonging to a set of propositions, or an *explanation*, which coheres better with the agent’s evidence relative to an available rival explanation, then the individual who has many false, but mutually

coherent, beliefs is doing very well, evidentially. Consider the case of Martha, who has many false, but mutually coherent beliefs. She believes that the government is obsessed with controlling its citizens. As such, she has very low trust in them, accepting little of what they say. Martha has come to have beliefs like this through interacting with like-minded individuals whom she judges to be beacons of knowledge. These individuals always seem to have an explanation for big political events, and it all fits together so well. When the government announces that they will be initiating a mass vaccination programme to protect its citizens against COVID-19, Martha has her suspicions. She doesn't doubt that many people are being injected with something, as she has seen footage of this on the television and has been told as much by several friends. Rather, she doubts that the government are vaccinating people to protect against COVID-19, since, according to Martha, there is no such thing as COVID-19. Martha instead believes that all of this is a big ruse to control people. She believes *that the government is injecting people with 5G chips*.

Does Martha have evidence for her belief *that the government is injecting people with 5G chips*? I argue that, if we adopt the best explanation account of evidential support, we are committed to saying so. On the best explanation view, evidence supports a hypothesis when the hypothesis features as part of the best available explanation of it. As we have seen, what it is for an explanation to be "the best" for a given agent is for it to fit the agent's evidence into a more coherent body of propositions relative to any available rival explanation. As such, to see whether Martha has evidence for her belief *that the government is injecting people with 5G chips*, we must establish whether the proposition *that the government is injecting people with 5G chips* features as part of the best available explanation for Martha's evidence.

Since Martha forms her belief *that the government is injecting people with 5G chips* in response to her observation of the government's vaccination programme,

her evidence here is the observational data of the vaccination programme. The next step is to evaluate the extent to which the hypothesis *that the government is injecting people with 5G chips* fits her observational evidence into a coherent body of propositions. To do this, we will have to look at her background beliefs, as we will need something against which we can measure the hypothesis for coherence. Given Martha's low trust in the government, her beliefs about the government controlling people, and her belief that COVID-19 does not exist, the hypothesis *that the government is injecting people with 5G chips* helps to fit the presence of her evidence into a more coherent body of propositions relative to available rival explanations. She need not amend *any* of her beliefs. The hypothesis *that the government is injecting people with 5G chips* sits in a beautiful harmony with the rest of Martha's beliefs. The hypothesis *that the government is injecting people with 5G chips* certainly affords Martha a better explanation than the rival hypothesis *that the government is vaccinating people against COVID-19*. The latter hypothesis, by contrast, does not help to fit Martha's observational evidence into a coherent set of propositions insofar as it fails to cohere with many of her background beliefs.

Clearly, there is something wrong here. Martha's supporting evidence is supportive of her hypothesis *only* insofar as it helps to achieve a coherent body of propositions. But this does nothing to push her towards the truth or likely truth. The deeper issue is that, if evidential support is understood by appealing to explanatory coherence, as Conee and Feldman suggest, then there is no reason to believe that, through demonstrating a sensitivity to evidence, one thereby demonstrates a sensitivity to considerations in favour of the truth, or falsity, of some proposition.

Here, one might respond that I am entertaining what is clearly an *atypical* scenario. Perhaps I should start with the assumption that, most of us have more true than false beliefs. Such an assumption surely reflects reality, so this should not be controversial. Now, according to the best explanation view, when a piece of evidence supports some hypothesis, that hypothesis will belong to a set of

propositions which strongly coheres with our evidence. At least in the case where evidence consists in an agent's beliefs, an agent's evidence is probably true because we are operating on the assumption that most of our beliefs are true. The thought is that, because of this, there is a higher probability that the supported hypothesis will also be true; that the truth of the evidence will increase the likelihood of a true hypothesis. But this doesn't work either.

Consider once more the case of Simon. Let's say that Simon has mostly true beliefs. The evidence he receives, in the form of the proposition that he has a rash, supports the hypothesis that he has shingles insofar as it, taken together with the true background beliefs he has about shingles, helps to fit the presence of his evidence, that he has a rash, into a coherent body of propositions which he holds to be true. Despite forming a coherent body of propositions and despite Simon mostly having true beliefs, his evidence supports a false proposition. Even allowing that most of our beliefs are true does nothing to assist us here. And if that is the case, then what reason do we have to believe that any given case of our evidence supporting some proposition, for example, that the speaker is offering veridical testimony, does not play out exactly like this? The question arises then: how does this account of evidential support rule out the possibility of systematic error?

To be clear, my concern is not that sometimes our evidence can lead us astray. I do not dispute that this is a possibility and I do not suggest that a necessary condition for something, e , to be evidence of some hypothesis, h , is that e guarantees h with certainty. Instead, the issue I outline above is that, if evidence supports a hypothesis when the hypothesis is part of the best available explanation of the evidence, and the best explanation of an agent's evidence is a body of propositions which best coheres with one's evidence relative to available rival hypotheses, then the support which a hypothesis receives is not concerned with the probability of the hypothesis obtaining but the degree to which a hypothesis, when taken together with a set of related propositions, coheres with the presence of an

agent's evidence. The problem is that this notion of evidential support fails to allow evidence to act as a bridge to the truth, or likely truth. Rather, it has evidence act as a bridge to *internal coherence*. And given that a set of propositions can be mutually coherent despite being false, we must concede that coherence and truth, or likely truth, come apart, at least in principle. In short, there is no reason to think that a sensitivity to evidence, when realised in this internalist manner, can rule out systematic error simply because this notion of evidential support need not be concerned with the truth. This is a serious problem. If one cannot rule out systematic error, one cannot rule out the possibility that demonstrating a sensitivity to evidence, conceived in this manner, has us tend towards false belief.

As stated already, the evidential hypothesis attempts to account for the truth-directedness of our testimonial discriminations by appealing to a sensitivity to evidence. But if the thing to which we are being sensitive is primarily concerned with coherence and coherence comes apart from the truth, then why think that demonstrating a sensitivity to evidence would explain truth-directedness? There must be a connection between evidence and truth if we are to explain the truth-directedness of our testimonial discriminations by way of a sensitivity to evidence. Given the failure of the internalist evidential hypothesis to accommodate that vital connection, I argue that we must eliminate it.

IV

Having eliminated the internalist evidential hypothesis, we are now left to examine the *externalist evidential hypothesis*. As previously stated, internalists are committed to the thesis that the factors which determine the justificatory status of an agent's belief must be *internal* to an agent. Externalists reject this thesis, instead holding that at least some external factors play a role in determining the justificatory status of an agent's belief that *p*. Such factors may involve *causal connections* (Goldman, 1967), *a counterfactual sensitivity to the truth* (Nozick, 1981), *basis-relative safety*

(Sosa, 2007), or that one's beliefs be outputted by a *reliable belief-producing process* (Goldman, 1979). In chapter three, I introduced process-reliabilism. According to this position, an agent's belief that p is justified when that belief is produced by a belief-producing process which tends to output true beliefs. Insofar as process-reliabilism requires that reliability be understood in terms of the overall tendency of a process to produce more true beliefs than false beliefs, it is an *external* factor. This is because, whether a given process is reliable or not is independent of how an agent takes things to be. The externalist evidential hypothesis, therefore, can appeal to factors *outside* of an agent's mind in explaining the truth-directedness of our testimonial discriminations.

We saw in the previous section that those sympathetic to the internalist position must conceive of evidence, and evidential support, in such a way as to agree with the internalist thesis. On this picture, evidence is not something which involves *any* external factors nor is evidential support explained by anything external to an agent. The externalist can opt for a different route, taking evidence to be something that involves external factors, or that evidential support is explained, at least partially, by external factors. In the previous section, I found the internalist evidential hypothesis to fail insofar as it appeals to a notion of evidence which is not truth-directed. In this section, I introduce a notion of evidence which I regard as appropriately truth directed.

According to Williamson, two conditions must be satisfied for a candidate-item to count as a piece of evidence, e , for a hypothesis, h , for some subject, S :

"First, e should speak in favour of h . Second, e should have some kind of creditable standing" (2000: 186).

The first condition can be regarded as a condition on *evidential support*. If e does not support h , it does not count as evidence. This requires little defence. The proposition *that bats are mammals* does not support the proposition *that Sheffield is a rainy city* in any way whatsoever, so we cannot regard the former as evidence in

favour of the latter. The second condition concerns the *standing* of *e* isolated from *h*. That is, the second condition does not concern the extent to which *e* supports *h*. Instead, the second condition focuses on whether *e* is the right sort of thing. Call this a condition on *evidential standing*.

To the question of what it means for a piece of evidence to “speak in favour” of a hypothesis, Williamson offers, as an approximation, the following response:

“*e* should raise the probability of *h*. That is, the probability of *h* conditional on *e* should be higher than the unconditional probability of *h*; in symbols, $P(h | e) > P(h)$ ” (ibid.).

To satisfy the evidential support condition, it must be the case that any candidate piece of evidence raises the objective probability of the hypothesis it purports to support. This means that the probability of the hypothesis given the candidate-evidence must be higher than the probability of the hypothesis alone.

Let us use an example from Williamson to better understand this (ibid.). Take the hypothesis *that John is single*. And let us take as our evidence *that John is a member of Club A*. To ascertain whether the proposition *that John is a member of Club A* counts as evidence for the hypothesis *that John is single*, it must be the case that the evidence makes the hypothesis more probable. This necessitates it being the case that the probability *that John is single* given *that John is a member of Club A* is higher than the probability *that John is single* taken by itself. But is it not the case that, depending on further information, the probability of a hypothesis given some evidence may be *lower* rather than *higher* than the probability of the hypothesis? Williamson appreciates this, stating that “whether *e* raises the probability of *h* may depend on background information” (ibid.). If one knows that Club A is a club for singles then this *increases* the probability of the hypothesis given the evidence, but if one knows that Club A is a club for married couples, then this *decreases* the probability of the hypothesis given the evidence. So, the probability

of the hypothesis obtaining given the evidence will also depend on the *background information or background beliefs* we have.

Williamson's evidential support condition presents a stark contrast to the best explanation conception of evidential support discussed in the previous section. Firstly, we see that Williamson appeals to *external factors* in requiring that, for e to count as evidence for h for S , $P(h | e) > P(h)$. This is external in the sense that whether e increases the probability of h or not is something that does not depend on S 's mental life whatsoever. Further, in requiring that e increase the probability of h , Williamson's concept of evidence is truth-directed given that what it is for something to be evidence for h for S is for it to increase the probability of h . This means that evidence is directly concerned with the *objective probability* that h is true. In this sense, by believing in accordance with your evidence, your beliefs are based on considerations which direct you towards probable truth.

Let us examine how this notion of evidential support functions when plugged into the evidential hypothesis. The evidential hypothesis states that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence. On Williamson's account, evidence is the sort of thing that increases the objective probability of the hypothesis for which it is evidence. As such, evidence can act as a guide to the truth insofar as, on the current account, what it is for something to be evidence is for it to be something that increases the objective probability of the hypothesis it purports to support.

Given this conception of evidential support, we can explain the truth-directedness of our testimonial discriminations by forwarding that audiences are sensitive to evidence which, importantly, has a bearing on the truth or falsity of the testimony offered by speakers. This sort of evidence is evidence for the truth of a limited number of hypotheses⁴³: that the report just offered is veridical; that the

⁴³ I am willing to accept that there may be other hypotheses one might include here. As this has no impact on the position I forward, I will simply state that the list I offer is not exhaustive.

report just offered is false; that the speaker is trustworthy; that the speaker is untrustworthy. Audiences are sensitive to this evidence and discriminate between bits of testimony in accordance with it. Given that evidence, on the current picture, supports a hypothesis when the probability of the hypothesis given the evidence is greater than the probability of the hypothesis alone, an audience's evidence will direct them towards the truth insofar as the audience's evidence will increase the probability of one of the previously mentioned hypotheses. So, by demonstrating a sensitivity to evidence which is understood as supporting a hypothesis only when the evidence increases the probability of the hypothesis, an audience's testimonial discriminations get to be truth-directed.

However, as we have seen, the evidential support condition is not sufficient by itself to give us an account of when e is evidence for h for S . Williamson also requires that "e should have some kind of creditable standing" (ibid. 186). Let us now examine this requirement further and see how it affects the externalist evidential hypothesis.

V

The evidential support condition is not enough by itself to tell us when something counts as evidence for S . This is because, as Williamson writes, "e may raise the probability of h in the sense that $P(h | e) > P(h)$ even if S knows that e is false or has no idea whether e is true; but then, for S , e would not be evidence for h " (ibid.). For this reason, Williamson proposes:

"EV: e is evidence for h for S if and only if S 's evidence includes e and $P(h | e) > P(h)$ " (ibid.).

As we have already discussed evidential support, the goal in this section is to understand the evidential standing condition *that S 's evidence includes e* .

Prior to dealing explicitly with Williamson's evidential standing condition, consider an example. Imagine two passengers on a train. Let's say that, at some time, t_1 , passenger A is asleep while passenger B is awake and that, at a later time, t_2 , both passengers are awake. At t_1 , the train passes by a field of cows. Gazing out of the window, B sees the field of cows and forms the belief "that the train passed a field of cows". At t_2 , A and B differ with respect to their evidence in support of the claim *that the train passed a field of cows*. What is the difference? It is that B has evidence A does not. Whilst A might concede that seeing a field full of cows while on the train is excellent evidence for believing *that the train passed a field of cows*, the evidence that A has restricts the doxastic moves A can make. Were A to nevertheless believe *that the train passed a field of cows*, A would have a true belief, but it would not be supported by the evidence A has.

The example above helps us to think of evidence as *that which is possessed by an agent*. Here, one might ask: is the fact that the train passed the field cows not evidence for the proposition *that the train passed a field of cows*? Why restrict our talk of evidence to only that which is evidence *for an agent*? To resolve this concern, we must recognize that we can speak of evidence in multiple ways. We can talk about something, e , *being* evidence. We can also talk about e being evidence *for an agent*, S . In the former sense, we do not talk about evidence *possession* while we must in the latter sense. Williamson has in mind this second sense of evidence, as something *had* by S . I will provide a brief outline of the former conception of evidence to serve as a contrast.

Achinstein argues that there are several concepts of evidence (1978) (2001). I will discuss his concept of *veridical evidence*, according to which e is evidence of h when it provides "a good reason to believe that h " in an "objective" manner (ibid. 25). On this conception of evidence, Achinstein maintains, "whether e is a good reason to believe h does not depend on whether in fact anyone believes e or h or

that *e* is a good reason for believing *h*" (ibid.). In this sense, *e* functions as a "sign" or "symptom" of *h*, similar to how we think of symptoms of diseases:

"A rash may be a sign or symptom of a certain disease, or a good reason to believe the disease is present, even if the medical experts are completely unaware of the connection and so are not justified in believing that this disease is present, given what they know" (ibid. 25).

When *e* is veridical evidence for *h* it provides a good reason to believe *h* irrespective of whether anyone has any awareness of it. Insofar as veridical evidence provides a good reason for belief irrespective of whether any agent is aware of the support *e* gives to *h*, its status as evidence does not require its being possessed. The thought here is that, if you *did* believe *h* on the basis of *e*, you *would* be doing so for an objectively good reason. However, whether anyone believes that *e* supports *h* has no influence, according to Achinstein, on the status of a piece of veridical evidence as evidence (ibid.).

By contrast, Williamson's discussion of evidence *requires* that it must be had by an agent (2000: 187). This is due to his focus on a question different to that considered by Achinstein. Achinstein is concerned with offering analyses of differing concepts of evidence, some of which necessitate possession by an agent and some of which do not⁴⁴ (1978) (2001). Williamson, by contrast, is concerned with the question of *what counts as evidence*. He writes:

"Why does it matter what counts as evidence? Consider the idea that one should proportion one's belief in a proposition to one's evidence for it. How much evidence one has for the proposition depends on what one's evidence is" (2000: 189).

Here, Williamson has in mind the concerns raised by Carnap when he introduces *the total evidence requirement*. According to Carnap:

⁴⁴ For example, Achinstein's concept of *subjective evidence* which requires, amongst other things, *S* to believe that *e* is evidence that *h* (2001: 23).

"[I]n the application of inductive logic to a given knowledge situation, the total evidence available must be taken as a basis for determining the degree of confirmation". (1950: 211)

The point here is that what counts as evidence for us is important. The set of my total evidence impacts how I follow rules such as "proportion your belief to your evidence". It also affects the credence I give to a hypothesis, and the hypotheses that I rule out as inconsistent with my evidence. As Williamson puts it, "disputes between different theories of evidence are not merely verbal" (2000:189).

In hoping to make sense of rules such as "proportion your belief to your evidence", it is important to understand what counts as *an agent's* total evidence. Consider, theory A might take an agent's total evidence to count as $\{e_1 \dots e_n\}$ while theory B takes an agent's total evidence to count as $\{e_{10} \dots e_n\}$. The consequence of this is that, insofar as the theory to which one subscribes impacts the membership of the set of evidence you have, what your evidence directs you to believe, or what it justifies, can fluctuate significantly. Williamson is engaged in the project of producing a theory which answers the question of what constitutes an agent's *body of evidence*. Notice that this is different to considering the question of whether some item, *e*, counts as a piece of evidence. Insofar as something is evidence *for* an agent, it should be understood as something belonging to a body of evidence which is *had by an agent* (ibid.).

It is because of Williamson's appeal to a conception of evidence which sees it as something which must be possessed by an agent that he cannot allow a candidate piece of evidence, say, *that the train passed a field of cows*, to count as evidence for an agent, *S*, when *S* has no idea whether it is true. The thought is that when we consider what beliefs are justified given an agent's evidence, we must restrict ourselves to looking only at the agent's available body of evidence. And so, that which is not possessed by *S* cannot be evidence for *S*. This is why, at t_2 ,

passenger A cannot be justified in believing *that the train passed a field of cows*. It is because his total evidence does not support that hypothesis.

Whilst we now have some understanding of the sort of thing Williamson is talking about when he talks about evidence, we are still without the deeper understanding necessary to appreciate his evidential standing condition. We now know that evidence must be *had* by an agent, but we can pry into this further. Consider, Conee and Feldman would happily agree with Williamson that evidence is *had* by agents. However, recall that Conee and Feldman take evidence to consist in the mental life of an agent, as beliefs and experiences (2004: 55-56). So, let us examine what form Williamson takes our body of evidence to take. This is an important matter to resolve because, if we are to explain the truth-directedness of testimonial discriminations by appealing to a sensitivity to evidence, we had better understand that thing to which we are demonstrating a sensitivity. As Williamson states “[h]ow much evidence one has for the proposition depends on what one's evidence is” (2000: 189).

Williamson's answer to the question of what counts as evidence for *S* is informed by his broader epistemological position. Whilst it is beyond the scope of this work to provide a particularly detailed picture of his *knowledge-first* account, I will provide a brief exposition of it here.

Williamson notes that the overwhelmingly large number of unsuccessful attempts to provide an analysis of knowledge are “obvious signs of a degenerating research program” (2014: 2). In response, Williamson “reverses the order of explanation” (2000: 185). Instead of analysing knowledge in terms of belief, truth, and some other non-circular conditions, he treats *knowing* as *fundamental* (ibid.) (2014: 2). Moreover, instead of making sense of knowledge by turning to concepts like *evidence* and *justification*, he claims that we should make sense of evidence and justification by turning to the concept of *knowledge*.

For Williamson, “knowledge, and only knowledge, constitutes evidence” (2000: 185). So, to the question of what counts as one’s evidence, Williamson answers: *knowledge*. The set of one’s evidence is equated with the set of one’s knowledge (ibid.). If this view is correct, then it would entail that the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence* is equivalent to the hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to knowledge*. But why think that knowledge *and only knowledge* constitutes evidence? Does the agent who mistakenly, but confidently, believes that she left her keys on the table not have evidence despite lacking knowledge? Why not instead think that this agent’s *belief* is part of her evidence?

Consider the following case: You look outside and see that the ground is dry. Whilst you might concede that the proposition *that the ground is wet* increases the probability of the hypothesis *that it rained last night*, you don’t have that proposition as evidence given that you take it to be false. Now, what if, due to some strange quirk of the light hitting your window, it appears to you as though the ground is wet? Do you have as evidence *that the ground is wet* now? In such a case you *believe* that the ground is wet, but do not *know* it. One might argue that, yes, in this case you do have evidence in the sense that you have the belief *that the ground is wet* which is afforded to you by your perceptual experience. This is not the route Williamson takes, however.

Williamson notes that “our evidence sometimes rules out some hypotheses by being inconsistent with them” (ibid. 196). This is clearly an important role evidence plays. *That the ground was coated in hydrophobic chemicals last night* is clearly incompatible with my evidence *that the ground is wet*. As such, my evidence rules out that hypothesis. But let us suppose that evidence consists not in knowledge, but in mere *belief*. If I merely *believe* that the ground is wet, and therefore have evidence that the ground is wet, then the hypothesis *that the ground*

was coated in hydrophobic chemicals last night is not ruled out by my evidence. This is because, if mere belief is what constitutes evidence, then my belief, and therefore my evidence, can be false and this will *not* rule out the truth of the hypothesis *that the ground was coated in hydrophobic chemicals last night* because beliefs need not be true to be held. Otherwise put, if mere belief constitutes one's body of evidence, then in assessing a range of hypotheses, I do not need to restrict myself to those hypotheses which are compatible with the truth of the proposition *that the ground is wet*, because my total evidence consists in the propositions I merely believe; not *truly believe*.

We can now return to the question of whether you have evidence *that the ground is wet* when it merely *appears* to you as though the ground is wet. In this case, you have the false belief *that the ground is wet*. Insofar as your belief is false, you cannot have *that the ground is wet* as your evidence. At best, your evidence can be *that the ground appears to be wet*, but this is different to the proposition *that the ground is wet*. To have *that the ground is wet* as evidence seems to require, at the least, *truly believing* the proposition *that the ground is wet*, and at the most, *knowing* the proposition *that the ground is wet* (ibid 201-202).

Here, one might suggest that the argument above merely tells us that evidence must be *true*, not that evidence must be *known*. If evidence consists in *true belief*, then my having *that the ground is wet* as evidence will rule out the hypothesis *that the ground was coated in hydrophobic chemicals last night* because, in assessing a range of hypotheses, I am forced to consider only those which are compatible with the truth of the proposition *that the ground is wet*.

Williamson's contention, however, is not that *true belief* constitutes evidence, but that *knowledge* does. So, consider the following case from Williamson:

"Suppose that balls are drawn from a bag, with replacement. In order to avoid issues about the present truth-values of statements about the future, assume that someone else has already made the draws; I watch

them on film. For a suitable number n , the following situation can arise. I have seen draws 1 to n ; each was red (produced a red ball). I have not yet seen draw $n+1$. I reason probabilistically, and form a justified belief that draw $n+1$ was red too. My belief is in fact true. But I do not know that draw $n+1$ was red" (ibid. 200).

The question Williamson then poses is whether he has *that draw $n+1$ was red* as evidence. His belief is true and seemingly justified. So, the suggestions on the table now include both that justified true belief constitutes evidence and that true belief constitutes evidence. However, neither suggestion seems to work here. This becomes clear when one appreciates that the proposition *that draw $n+1$ was black* is consistent with his evidence *even though it is false*. But if *that draw $n+1$ was black* is consistent with his evidence, it must mean the proposition *that draw $n+1$ was red* is not part of his evidence, because if it was, then *that draw $n+1$ was black* would directly contradict his evidence. Williamson takes it that examples like this demonstrate that *knowledge and only knowledge* constitutes evidence. In this example, he knows neither *that draw $n+1$ was black* nor *that draw $n+1$ was red*. His failure to know either proposition thereby entails his failure to *have* either proposition as evidence.

We can now return to Williamson's evidential standing condition. Recall that, according to the evidential standing condition, S 's evidence must include e (ibid. 187). We can make sense of this as requiring that, for some candidate piece of evidence, e , to count as evidence for h for S , it must be the case that S *knows* that e . Conjoined with the evidential support requirement, for e to be evidence for h for S , e must increase the probability of h such that $P(h | e) > P(h)$, and e must also be contained within S 's body of evidence, meaning that e must be known by S .

For Williamson, it is *knowledge* that counts as evidence for S . This impacts the evidential hypothesis insofar as it impacts the set of items to which an agent

demonstrates a sensitivity. It means that, in making truth-directed testimonial discriminations, an agent is demonstrating a sensitivity to what she *knows*.

An advantage of this position is that it easily handles the question of *how* we demonstrate a sensitivity to evidence. Consider an alternative way of understanding what counts as evidence for *h* for *S*. Consider Achinstein's concept of *potential evidence* (1978: 35) (2001: 27). On this concept, *e* is evidence that *h* iff:

"(a) *e* is true, (b) *e* does not entail *h*, (c) $P(h,e) > k$, (d) $P(\text{there is an explanatory connection between } h \text{ and } e, h.e) > k$ " (1978: 35).

That is, some candidate piece of evidence counts as potential evidence that *h* iff: a) the evidence is true; b) the evidence does not entail the hypothesis; c) the probability of the hypothesis obtaining given the evidence is greater than some number, *k*; d) the probability of there being an explanatory connection between the hypothesis and the evidence, given that both *h* and *e* are true, is greater than some number, *k*.

If we wish to explain the truth-directedness of our testimonial discriminations by way of a sensitivity to evidence, then, if we take evidence to be something like Achinstein's *potential evidence*, we must ask: *how* are we able to demonstrate a sensitivity to it? The benefit of adopting E=K is that this hard work is already done for us. We are sensitive to our evidence because our evidence is what we know. The theorist who wishes to adopt an account relying on something like *potential evidence* will need to give us a story as to how we are sensitive to it. I do not wish to suggest that such a challenge is insurmountable. Rather, I wish to claim only that the theorist who defends the externalist evidential hypothesis whilst understanding evidence as something like Achinstein's *potential evidence* owes us a story of how we can demonstrate a sensitivity to evidence.

Given the arguments provided here, I take it that the externalist evidential hypothesis is well-equipped to explain the truth-directedness of our testimonial discriminations. To explain the truth-directedness of our testimonial discriminations

by way of a sensitivity to evidence, it must be the case that evidence acts a bridge to the truth in some sense. If we adopt Williamson's notion of evidence, we can achieve this. For Williamson, a necessary condition of e counting as evidence for h for S is that e increases the objective probability of h , such that $P(h | e) > P(h)$. If e increases the objective probability of h then demonstrating a sensitivity to e will see one demonstrating a sensitivity to something directly related to the objective probability of h . And if that is the case, then it is clear how evidence can act as a bridge to the truth.

Furthermore, if we adopt Williamson's *knowledge-first* account (2000), we can quickly respond to any concern over how we are sensitive to evidence. Given $E=K$, we can simply state that, when agents demonstrate a sensitivity to their evidence, they demonstrate a sensitivity to their own knowledge. The set of our evidence is equivalent to what we know (ibid. 185).

Given all of this, I conclude that *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. According to the view I have advanced, evidence should be realised in an externalist manner to allow it to act as a bridge to the truth. I have proposed we make sense of evidence by appealing to Williamson's $E=K$ thesis, which equates the set of one's evidence with the set of one's knowledge (ibid.). In realising evidence in such a way, we can explain how audiences are sensitive to evidence; they are sensitive to that which they know. And through including a support condition on evidence such that, for e to be evidence for some hypothesis, h for some agent, S , $P(h | e) > P(h)$, we can see how demonstrating a sensitivity to evidence can account for the truth-directedness of our testimonial discriminations. The guiding thought is that, by demonstrating a sensitivity to something which is connected to truth, viz., evidence, our testimonial discriminations get to be *truth-directed*. In the next section, I will respond to a number of objections that one may raise against my view.

VI

1. The empirical data shows that we fail to demonstrate a sensitivity to evidence

Consider again the deception detection studies discussed in chapter three. These experiments tend to work in a very similar fashion. Participants, also called *receivers* or *hearers*, watch clips of speakers, also called *senders*, offering some testimony. Roughly half of the speakers in the total set will be offering false testimony knowingly and half of the speakers will be offering veridical testimony knowingly. The experiments are conducted to assess the capability of humans to detect deceptions (Vrij, 2008: 147). However, the results from decades of research into deception detection indicate that we lack any reliable capacity to detect deceptions (Bond and Depaulo, 2006: 231). The average detection rate stands at little more than 50%, which is only fractionally higher than what we expect through chance alone (Vrij, 2014: 321). In the third chapter, I appealed to these studies in arguing against the claim that there is a reliable process for making truth-directed testimonial discriminations⁴⁵. One might use that same data to argue against the evidential hypothesis.

I will first note, as I did in chapter three, that there is a difference between *experimental* and *everyday scenarios*. The experimental scenarios present a stark contrast to those everyday scenarios with which we are familiar. The reliabilist, I suggested, might propose that the reliable truth-directed testimonial discriminations process functions well in *everyday scenarios* but not in *experimental scenarios*. I think that this distinction is epistemically significant and helps to explain the poor results in the studies. It does not follow from this, however, that we lack a capacity to make truth-directed testimonial discriminations. I take it that, in response

⁴⁵ Of importance to note, however, is that the conclusion at which I arrived after introducing the empirical data was that we lack a reliable process for discriminating between bits of testimony *on the basis of analysing speaker behaviour*. However, this claim should be specified only to those speakers with whom we are unfamiliar.

to the empirical data, any theorist must accommodate the empirical data into their account of truth-directed testimonial discriminations.

As noted previously, if we are to make a distinction between experimental and everyday scenarios, we need to identify conditions that differ between the two scenarios. One suggestion I advanced concerns *the availability of evidence*. In the experimental setting, there is a paucity of evidence. Now, if the evidential hypothesis is correct, this explains why we perform so badly in the experiments: the way we make truth-directed testimonial discriminations is by being sensitive to evidence. But the experimental scenario is one in which we lack evidence. As such, we are deprived of that which we need to discriminate between bits of testimony in a truth-directed manner. Much as light is required for us to discriminate between objects visually, evidence is required for us to discriminate between bits of testimony in a truth-directed manner.

This raises a potential issue, however. If $E=K$ is right, then that which constitutes our evidence is our knowledge. If that is so, then why should the experimental scenario be a scenario in which evidence is lacking? We all have a great deal of evidence if $E=K$. When the participants enter the laboratory, they are not deprived of the knowledge they have of speakers and testimonial exchanges. So, in what sense are they *evidentially deprived*?

To this, we must be clear about the *kind* of evidence to which we are sensitive in testimonial exchanges. Firstly, we need to understand the sort of hypotheses which might receive support from any candidate piece of evidence. Earlier, I tentatively suggested four hypotheses: that the report just offered is veridical; that the report just offered is false; that the speaker is trustworthy; that the speaker is untrustworthy. So, any evidence to which we are sensitive must be evidence which could be advanced in support of these four hypotheses⁴⁶. Such evidence may take

⁴⁶ There may very well be further hypotheses, or fewer hypotheses. I do not see this as a key issue and will not spend time discussing it.

the form of a speaker's interests, a speaker's position towards lying, information about which the speaker is testifying, what sort of social norms are in play and how they will exert pressure on the speaker, the tendency of speakers to report truthfully in such situations, our knowledge of what is said and what is contrary to what is said, and so on. The issue is that much of this evidence, whilst useful in situations where we know the speaker, or at least are familiar with the social context in which we receive the testimony, is irrelevant in the experimental setting. Knowledge of the truth or falsity of what is said is of course relevant, but the experimental scenario often restricts what is reported by the senders to that which is unknown by the receivers. If this is the case, then the receivers are without the type of evidence they need to make truth-directed testimonial discriminations.

Here, it is worth considering the results of a study conducted by Park, Levine, McCornack, Morrison & Ferrara (2002). They gave 202 undergraduate students a questionnaire, asking them to recall, and reflect on, a clear example in which they successfully detected a deception. The participants were asked how they discovered the lie and how long it took to detect it. 32% of the participants responded that they discovered the lie via third party information (ibid. 151). A mere 02.1% responded that they discovered the lie in question via verbal / nonverbal behaviour (ibid.). Moreover, the researchers found that "only a small proportion of the recalled lies were detected at the time they were told, and a much larger proportion was discovered an hour or more after the fact. Some lies were detected months or even years later" (ibid. 152-153).

If the results presented above afford us an accurate picture of how we tend to behave, then we should not be surprised by our poor performance in the deception detection experiments. Deception detection, one might conclude given this research, is not accomplished by staring into the eyes of your interlocuter, judging the truth of their report as soon as they finish speaking. Rather, deception detection necessitates a sensitivity to a bunch of different sorts of evidence and can

take a long time to accomplish. If we explain the truth-directedness of our testimonial discriminations by appealing to a sensitivity to evidence, we are well-equipped to explain our poor performance in the deception detection studies. We can maintain that the methods on which we tend to rely for making truth-directed testimonial discriminations require the sort of evidence which the experimental scenario prohibits. When we have that evidence available to us, however, we can, and do, make truth-directed testimonial discriminations.

2. We must be aware of our evidence

To make truth-directed testimonial discriminations is, according to the view defended here, to be sensitive to *evidence*. If you receive evidence which suggests that p , you should respond to it in the right way by believing that p . Another way to put this is that you should *respect* your evidence (Williamson, 2000: 164). But how can one respect their evidence unless one knows what it is? I might falsely believe the proposition *that the ground is wet*, in which case I will not *know* that the ground is wet and thus, not have evidence. Nevertheless, I might treat that false belief as evidence. But if making truth-directed testimonial discriminations requires a sensitivity to evidence, does this not demand that we be aware of what our evidence is? How do I know that a given case of me being sensitive to “evidence” is really a case of me being sensitive to evidence and not something which merely looks like evidence?

The issue here might be seen as a sceptical worry. There are good cases and bad cases. In a good case, things will appear to you thus-and-so and the world will be thus-and-so. It might appear to you that the ground is wet, and the ground is indeed wet. Alternatively, you might be in the bad case, in which things appear to you thus-and-so but are not. In the bad case, it appears to you that the ground is wet, but the ground is not wet. The problem is that appearances in the bad case are

subjectively indistinguishable from the good case. So, appearances alone will fail to aid your understanding as to whether you are in the good or the bad case.

If we adopt $E=K$, we must say that we have evidence in the good case and not in the bad case. The problem is that, in the bad case, I will falsely believe that I have evidence and will respond to it as *if it really is my evidence*. But surely, we need some way of discriminating between evidence and that which merely looks like evidence. After all, if I don't know what my evidence is, how can I respond to it appropriately?

The idea of *respecting one's evidence* can also be framed as a matter of "proportioning one's belief to one's evidence" (ibid. 192). Such a phrase has been discussed already but might also be familiar from Hume's famous line that "[a] wise man [...] proportions his belief to the evidence" (1748/1999: 77). One might ask: how it is possible to follow such a rule if one cannot know what one's evidence is? Williamson, however, suggests that this is not necessary. Think of the rule: "proportion your voice to the size of the room" (2000: 192). Our following of this rule does not necessitate knowledge of the size of the room we are in, and it would be insincere to suggest that what purports to be a common following of this rule is mistaken. Rather, Williamson argues, "[t]o have applied the rule 'proportion your voice to the size of the room', one needs beliefs about the size of the room, but they need not have been true – although if they were false, one's application was faulty" (ibid.). And the same is true of our following of the rule "proportion your belief to your evidence". What is required are beliefs about your evidence, and consequently, your knowledge (given $E=K$). Sometimes those beliefs may be false, in which case, one's application is faulty. Following the rule does not require knowing what your evidence is, however. Just as one can proportion one's voice to the size of the room without knowing the size of the room, one can proportion one's beliefs to one's evidence without knowing their evidence. Admittedly, this is quite the implication: *we are not always in a position to know what our evidence is*. Given

$E=K$, we are not always in a position to know what we know. Importantly, however, it does not follow from this that we cannot respond to evidence.

A further, related concern is that if we don't know what our evidence is, then isn't it just a matter of luck that we are sensitive to evidence whenever we are? This I reject. Consider the internalist position, that evidence need not increase the objective probability of the hypothesis obtaining. Adopting such a view allows for the conclusion that, in the bad case, one's evidence is equal to that in the good case. So, when you falsely believe *that the ground is wet* in the bad case, you have just as much evidence as when you truly believe *that the ground is wet* in the good case. Perhaps one finds this implication attractive, but if this is right, then it is a matter of *luck* that that appearances are as they seem.

Contrast the internalist position above with the externalist position defended in this chapter. On the externalist view, one's evidence is different in the two cases because what it is for something to be evidence for a hypothesis is for the evidence to increase the objective probability of the hypothesis. In the bad case, one's "evidence" does not do this, so it is not evidence. Notice then, that it is not a matter of luck whether appearances are as they seem. The grounds of one's belief, on the position I defend, are different in the two cases. In the good case, one's grounds for belief are evidential and in the other case they merely *seem* evidential. Admittedly, this means it is a matter of luck whether we have evidence or not, but on the internalist position it is a matter of luck whether evidence guides us to the truth or not. This element of luck is, unfortunately, inescapable. As McDowell puts it "[w]hether we like it or not, we have to rely on favors from the world" (1995: 886). He continues, "the hope is that this admission of luck is tolerable, because it comes only after we have credited reason with full control over whether one's standings in the space of reasons are satisfactory" (ibid. 885). That is, the luck that features here has nothing to do with getting to the truth because getting to the truth by way of

evidence is not “lucky”. Rather, the good fortune you experience when you are in the good case and not the bad case is in *attaining evidence*.

Importantly, this does not entail that the presence of luck is epistemically troublesome. Recall the distinction noted in the first chapter, between *Evidential Epistemic luck* (EE luck) and *Veritic Epistemic luck* (VE luck) (Engel, 1992: 67). One is EE lucky when it is a matter of luck that one has evidence, but VE lucky when it is a matter of luck that one’s belief is true. As discussed already, EE luck is compatible with knowledge whilst VE luck is not. So, that one has evidence may be a matter of luck, but as noted, EE luck is not knowledge-impeding.

3. The evidential support condition is too weak

The evidential support condition discussed in §IV holds that evidence, *e*, supports a hypothesis, *h*, when $P(h | e) > P(h)$ (ibid. 186). This support condition, one might argue, is too weak insofar as it allows a lot of propositions to count as evidence for some hypothesis when they should not be counted as such. Achinstein takes issue with this notion of evidential support. In doing so, he writes the following:

“Events often occur which increase the probability or risk of certain consequences. But the fact that such events occur is not necessarily evidence that these consequences will ensue; it may be no good reason at all for expecting such consequences. When I walk across the street I increase the probability that I will be hit by a 1970 Cadillac; but the fact that I am walking across the street is not evidence that I will be hit by a 1970 Cadillac” (1978: 29).

The idea here is that, when *S* walks across the street, they thereby increase the probability of being hit by a 1970 Cadillac insofar as the probability of the hypothesis *that S will be hit by a 1970 Cadillac given that S is walking across the street* is higher than the probability of the hypothesis alone (ibid.). If we adopt Williamson’s evidential support condition, we are committed to concluding that the

proposition *that S is walking across the street* is evidence for the hypothesis *that S will be hit by a 1970 Cadillac* when we know *that S is walking across the street*. But this is absurd. This notion of evidential support clearly allows too much.

In response to this, Achinstein proposes a different probabilistic support condition:

$$P(h | e) > k \text{ (ibid. 30).}$$

This formula reads that the probability of some hypothesis h , given the evidence, e , is greater than some number, k . This can get around the issue raised by the counterexample above because, unless k is very low, or the probability of being hit by a 1970 Cadillac is very high, *that S is walking across the street* will fail to be evidence that *that S will be hit by a 1970 Cadillac* insofar as it will fail to increase the probability of that hypothesis to a value greater than k . But this definition suffers from criticisms too. Consider the following evidence and hypothesis:

e: that this badger lives in Sheffield.

h: that this badger cannot speak English.

It is quite clear that e is not evidence for h here. And when we utilise the “greater than h ” support condition ($P(h | e) > P(h)$), we see that it is *not* considered evidence because the probability of the hypothesis given the evidence is not greater than the probability of the hypothesis alone. Problematically, when we use the “greater than k ” support condition ($P(h | e) > k$), we see that it must be considered evidence. This is because the probability of h is already so high, so when taken together with e , ($h | e$) is greater than k . So, this definition is no good either.

In response to this, Achinstein has suggested that, in such cases, the issue is that there is no explanatory connection between e and h . What a good definition of evidential support needs then, is a requirement on the existence of an explanatory connection between e and h in addition to the requirement that the probability of h given e be greater than k .

As should be familiar from an earlier discussion, consider Achinstein's definition for *potential evidence*:

"*e* is potential evidence that *h* iff (a) *e* is true, (b) *e* does not entail *h*, (c) $P(h,e) > k$, (d) $P(\text{there is an explanatory connection between } h \text{ and } e, h.e) > k$." (ibid. 38).

Potential evidence must be true. It must also not entail *h*. This means that *e* can support *h* without always bringing it about. This contrasts with Achinstein's concept of *veridical evidence* which requires *h* to be true (ibid.). It is also the case that, for *e* to be potential evidence for *h*, the probability of *h* given *e* must be greater than *k*. Achinstein suggests for *k* to stand at 0.5 (ibid. 30). Further, for *e* to be potential evidence that *h*, the probability that there is an explanatory connection between *h* and *e*, given the truth of both *h* and *e*, must be greater than *k*. It is this condition which rules out our badger example as being a case of evidence for the hypothesis. The badger example clearly fails to satisfy the explanatory connection requirement.

This new definition can be plugged into Williamson's earlier proposal (EV), to yield:

EV*: *e* is evidence for *h* for *S* if and only if *S*'s evidence includes *e* and (a) *e* is true⁴⁷, (b) *e* does not entail *h*, (c) $P(h,e) > 0.5$, (d) $P(\text{there is an explanatory connection between } h \text{ and } e, h.e) > 0.5$. (ibid. 38) (Williamson, 2000: 186).

I will make the point here explicitly, however, that it is not crucial to my account that this *specific* notion of evidential support is true. I have included this amendment to show that the weakness of the previously discussed evidential support condition should not present a barrier to one's defending the externalist evidential hypothesis.

⁴⁷ This condition is not needed if one accepts E=K. This is because E=K requires that *e* be known and therefore, true.

VII

The externalist evidential hypothesis is compatible with a range of positions regarding justified testimonial acceptance. This is a further advantage of the theory. For example, one might defend Burge's non-reductive view, that testimonial acceptance is justified so long as one does not have stronger reasons not to accept what the speaker says (1993). One might defend Goldberg's process-reliabilist view, that testimonial acceptance is justified when it is the product of a reliable testimonial belief producing process (2010). Or one might defend McMyler's assurance view, that testimonial acceptance is justified for interpersonal reasons, based on trust (2011). All these views are compatible with the externalist evidential hypothesis, so long as the manner in which we make truth-directed testimonial discriminations is explained by our sensitivity to evidence. The non-reductivist has an easy time agreeing with this by virtue of requiring a sensitivity to defeaters. The reliabilist can accommodate the evidential hypothesis by arguing for a reliable evidence-monitoring process. And the assurance theorist can accommodate the evidential hypothesis by maintaining that we demonstrate a sensitivity to evidence *of trustworthiness*, which then issues a state of trust towards the speaker.

It is an open question *how* we come to have this capacity to make truth-directed testimonial discriminations. I do not want to commit to an answer on this matter, but I suggest that something like the following is plausible. I suggest that our ability to make truth-directed testimonial discriminations can be understood as a developed cognitive ability. As we develop, we slowly acquire knowledge about different *testimonial situations, social pressures, speaker behaviour* and the like. At a young age, we have little knowledge about this, but, by adulthood, we have amassed a great deal of knowledge. We learn that there are different kinds of "testimonial situations". There are situations in the classroom, where a teacher offers testimony. There are situations at home, where a caregiver offers testimony. There

are situations at work, where a peer gives testimony. There are low-stakes situations. There are high-stakes situations, and many more. These situations differ insofar as there are different expectations of the speaker. Whilst these aren't always clear to a young child, they become clear when the child develops and learns that there are *social pressures* exerting force on speakers to behave certain ways. An understanding of the sort of testimonial situation one is in can help one to recognise the sort of social pressures exerting force on speakers. Moreover, we come to learn that speakers in certain testimonial situations tend to behave in a certain way.

The benefit of a model like this is that it explains the tendency of very young children to accept what they are told, *by virtue of being told it*. So, the truth-directed testimonial discriminations capacity, or system, comes "online" so to speak, when it receives enough input. Without learning about testimonial situations, social pressures, speaker behaviours and so on, we can't make truth-directed testimonial discriminations.

I will reiterate that I am not committed to the above view. I include this to provide some further detail to the position I defend and the sort of implications it might have for the way we should think about truth-directed testimonial discriminations.

VIII

In this chapter, I first introduced the evidentialist position concerning epistemic justification to show how the evidentialist hypothesis receives support. After doing so, I focused on the question of when a piece of evidence supports a hypothesis. In doing so, I made a distinction between the *internalist evidential hypothesis* and the *externalist evidential hypothesis*. The former position subscribes to the internalist thesis that the factors responsible for conferring justification onto an agent's belief that *p* must be *internal* to that agent. To better understand this position, I looked at the work of Conee and Feldman (2008), who defend a *best explanation* view about

evidential support. They maintain that “[w]hat is justified for the person includes propositions that are part of the best explanation of those experiences available to the person” (ibid. 98). As such, they require no appeal to any factors external to an agent in expressing their view about the evidential support relation. The problem with this view, I argued, is that it relies entirely on the notion of *coherence*. And coherence, I argued, can come apart from the truth, thereby allowing for systematic error. We need a conception of evidence that will act as a bridge to the truth if the evidential hypothesis is to work. I argued that the internalist evidential hypothesis is unable to overcome this problem and must therefore be eliminated.

After this, I introduced the work of Williamson (2000) in my discussion of the *externalist evidential hypothesis*. Williamson’s evidential support condition requires that, for e to be evidence for h for S , the probability of the hypothesis *given the evidence* must be higher than the unconditional probability of the hypothesis ($P(h | e) > P(h)$) (ibid. 186). In requiring this, we can understand how evidence is appropriately connected to the truth in such a way that it can act as a bridge to the truth. After this, I focused on the question of *what* evidence is. This is an important question because the evidential hypothesis explains the truth-directedness of our testimonial discriminations by maintaining that we are *sensitive* to it. Without an understanding of this thing to which we are sensitive, we fail to have a detailed understanding of how the externalist evidential hypothesis works. In detailing this, I noted an important distinction in the way we can discuss evidence: we can talk about something being evidence, but we can also talk about something being evidence *for an agent*. Williamson, I noted, focuses on the question of *what counts as evidence for an agent*. As such, his concern is with *bodies of evidence* which requires that evidence be that which is possessed. Thereafter, I introduced his E=K thesis, which forwards that that “knowledge, and only knowledge, constitutes evidence” (2000: 185). This position, I argued, presents us with a way to defend the

evidential hypothesis. It also has the implication that, in demonstrating a sensitivity to evidence, we are demonstrating a sensitivity to what we know.

In §VI, I responded to several objections and made some clarifications. First, I argued that the deception detection data, far from being a problem for the theory, is exactly what we should expect. If we make truth-directed testimonial discriminations by demonstrating a sensitivity to evidence, then situations which strip us of that all-important evidence should be situations in which our ability to make truth-directed testimonial discriminations is adversely affected. This is exactly what we see.

I also argued that the externalist evidential hypothesis does not require us to be *aware* of what our evidence is. We may very well go wrong sometimes, judging a belief to be knowledge, thereby judging *e* to be evidence when it is not. These cases are unfortunate, but they do not cause irreparable damage to the position. The fact that we sometimes go wrong does not stop us demonstrating a sensitivity to evidence when we do.

I then suggested that we can bolster the evidential support condition by turning to Achinstein's notion of *potential evidence* (1978: 38). I should say that this is not crucial to my account, however. Of course, in stripping away detail to avoid objections, one's account becomes less explanatory. It is necessary to take a stance with respect to a variety of issues. The stance I have taken in this thesis is that our testimonial discriminations are explained by our sensitivity to evidence. Evidence should be understood in an externalist manner; as something which increases the objective probability of the hypothesis it purports to support. Through demonstrating a sensitivity to such evidence, our testimonial discriminations get to be truth-directed. This position is compatible with a range of views regarding justified testimonial acceptance and can be accepted by assurance theorists, reliabilists, reductionists and non-reductionists.

We are now at the end. This thesis has been written with the goal of introducing a new condition on the attainment of testimonial knowledge and justified testimonial belief: *that we make truth-directed testimonial discriminations*. But my major focus in this thesis has been with the question of what *explains* the truth-directedness of our testimonial discriminations.

In the second chapter, I rejected the non-reductive hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to counterevidence*. The issue with this hypothesis is that it can only provide a rival explanation to the evidential hypothesis if there is a clear distinction between *evidence* and *counterevidence*. I argued that no such distinction exists; that which is evidence that p can be counterevidence that $\sim p$, depending on the context. As such, the non-reductive hypothesis fails to provide an explanation of the truth-directedness of our testimonial discriminations which is distinct from the evidential hypothesis.

In the third chapter, I rejected the reliabilist hypothesis *that our testimonial discriminations are truth-directed because they demonstrate the operation of a reliable faculty / process*. I argued that the reliabilist must agree with the evidential hypothesis insofar as the manner in which the truth-directed testimonial discriminations process operates, according to the reliabilist, necessitates a sensitivity to evidence. Insofar as the truth-directed testimonial discriminations process is a process of monitoring evidence, the reliabilist hypothesis fails to provide an explanation of the truth-directedness of our testimonial discriminations distinct from the evidential hypothesis.

In the fourth chapter, I rejected the assurance hypothesis *that our testimonial discriminations are truth-directed because, in trusting a speaker, an audience is sensitive to the truth of what is said*. I argued that, for this hypothesis to account for the truth-directedness of our testimonial discriminations, there must be a connection between trust and truth such that trust can function as a guide to truth,

or probable truth. One can work to support this claim, I suggested, by maintaining that when an audience, *A*, trusts a speaker, *S*, for the truth of her report, *A* represents *S* as trustworthy. Given that what it is to be trustworthy is to be a sincere and competent speaker who tends to get things right, we see that trust is connected to truth by way of *trustworthiness*. However, this alone fails to support the claim that trust can function as a guide to truth, or probable truth, since trust might be extended, by and large, to those who are entirely *untrustworthy*. For trust to function as a guide to truth, or probable truth, our patterns of trust must map onto occasions of speaker trustworthiness. Without this, we allow that trust can tend towards failure. I argued that trust cannot do this without an appeal to evidence in support of *trustworthiness* or *trust-responsiveness*. I then concluded that this reliance on evidence sees the assurance hypothesis collapse into the evidential hypothesis.

I have argued in defence of the evidential hypothesis *that our testimonial discriminations are truth-directed because they demonstrate a sensitivity to evidence*. In doing so, I take it that evidence must be understood in such a way as to be connected to the truth. This, as we have seen, rules out internalist conceptions of evidence, such as the view defended by Conee and Feldman (2004) (2008). Instead, I have defended an externalist evidential hypothesis, appealing to Williamson's E=K thesis (2000). I will note, however, that there may be other ways to realise evidence in the way necessary for truth-directed testimonial discriminations (see McDowell (1995), Achinstein (1978) (2001)). On this point, I wish to make clear that, if a sensitivity to evidence is to explain the truth-directedness of our testimonial discriminations, evidence must be able to act as a reliable guide to the truth. It can only do that, I have argued, if it is connected to the truth. It is for this reason, that I have argued in support of the *externalist evidential hypothesis*.

Finally, I will note that the discrimination condition I have introduced in this thesis has merely been *applied* to the testimonial domain. The discrimination condition is a condition on justified belief and knowledge more generally. Whilst I

have been focused on testimony here, the arguments presented in the first chapter have a wider reach than the testimonial domain. As such, there is far more work to be done regarding the discrimination condition on knowledge. To gain a deeper understanding of the route to knowledge we must understand how we make discriminations in a way such that we can attain knowledge. I hope that this presents us with an exciting opportunity for new research in epistemology.

Bibliography

- Aamodt, M.G. & Cluster, H. (2006) "Who Can Best Catch a Liar? A Meta-Analysis of Individual Differences in Detecting Deception", *The Forensic Examiner*, Spring: 6-11.
- Achinstein, P. (1978) "Concepts of Evidence", *Mind*, 87(345): 22-45.
- Achinstein, P. (2001) *The Book of Evidence*, Oxford: Oxford University Press.
- Akiba, K. (2000) "Shogenji's Probabilistic Measure of Coherence is Incoherent," *Analysis*, 60: 356-359.
- Aristotle. (1999) *Nicomachean Ethics*, T. Irwin (Trans.). Indianapolis, IN: Hackett.
- Armstrong, D.M. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.
- Audi, R. (2003) *Epistemology: A Contemporary Introduction to the Theory of Knowledge*, 2nd edn. London: Routledge.
- Baier, A. (1986) "Trust and Antitrust", *Ethics*, 96(2): 231-260.
- Battaly, H. (2014) "Varieties of Epistemic Vice", in J. Matheson & R. Vitz (eds.), *The Ethics of Belief*, Oxford: Oxford University Press: 51-76.
- Baehr, J. (2011) *The Inquiring Mind: On Intellectual Virtues & Virtue Epistemology*, Oxford: Oxford University Press.
- Bond, C.F. & DePaulo, D.M. (2006) "Accuracy of Deception Judgements", *Personality and Social Psychology Review*, 10(3): 214-234.
- Burge, T. (1993) "Content Preservation", *Philosophical Issues*, 6: 271-300.
- Burge, T. (1997) "Interlocution, Perception, and Memory", in: T. Burge (ed.), *Cognition Through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection: Philosophical Essays, Volume 3* (2013), Oxford: Oxford University Press, 285-305.
- Carnap, R. (1950) *Logical Foundations of Probability*, Chicago: University of Chicago Press.

- Carter, J.A., Jarvis, B.W. & Rubin, K. (2015). "Varieties of Cognitive Achievement", *Philosophical Studies*, 172(6): 1603-1623.
- Chisolm, R. (1989) *Theory of Knowledge*, 3rd edn., London: Prentice-Hall.
- Coady, A.J. (1992) *Testimony: A Philosophical Study*, Oxford: Oxford University Press.
- Coffman, E.J. (2015) *Luck: Its Nature and Significance for Human Knowledge and Agency*, Hampshire: Palgrave Macmillan.
- Conee, E. & Feldman, R. (2008) "Evidence", in Q. Smith (ed.), *Epistemology: New Essays*, Oxford: Oxford University Press, 83-104.
- Conee, E. & Feldman, R. (2004) "Internalism Defended", in E. Conee & R. Feldman, *Evidentialism: New Essays*, Oxford: Oxford University Press, 53-82.
- Cross, C. B. (1999) "Coherence and Truth Conducive Justification," *Analysis*, 59: 186-93.
- Engel Jr., M. (1992). "Is Epistemic Luck Compatible with Knowledge?", *Southern Journal of Philosophy*, 30: 59-75.
- Faulkner, P. (1997) "David Hume's Reductionist Epistemology of Testimony", *Pacific Philosophical Quarterly*, 79: 302-313.
- Faulkner, P. (2000) "The Social Character of Testimonial Knowledge", *The Journal of Philosophy*, 97(11): 581-601.
- Faulkner, P. (2007) "On Telling and Trusting", *Mind*, 116(464): 875-902.
- Faulkner, P. (2011) *Knowledge on Trust*, Oxford: Oxford University Press.
- Faulkner, P. (2018) "Giving the Benefit of the Doubt", *International Journal of Philosophical Studies*, 26(2): 139-155, DOI: 10.1080/09672559.2018.1440952.
- Faulkner, P. (2020) "Testimony and Trust", in Simon, J., (ed.), *The Routledge Handbook of Trust and Philosophy*, Routledge, 329-340.
- Feldman, R. & Conee, E. (1985) "Evidentialism", *Philosophical Studies*, 48(1): 15-34.
- Feldman, R. & Conee, E. (2004) "Evidentialism", in E. Conee & R. Feldman, *Evidentialism: New Essays*, Oxford: Oxford University Press, 83-108.

- Firth, R. (1978) "Are Epistemic Concepts Reducible to Ethical Concepts?", in A. Goldman & J. Kim (eds.), *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, Dordrecht: Kluwer Academic Publishers, 215-229.
- Fricke, E. 1994: "Against Gullibility", in B. K. Matilal and A. Chakrabarti, (eds.), *Knowing from Words*, Amsterdam: Kluwer Academic Publishers, 125-61.
- Fricke, E. (2017) "Inference to the Best Explanation and the Receipt of Testimony", in K. McCain & T. Boston (eds.) *Best Explanations: New Essays on Inference to the Best Explanation*, 262-294.
- Gettier, E. (1963) "Is justified true belief knowledge?", *Analysis*, 23:121-123.
- Goldberg, S. (2007) *Anti-Individualism: Mind and Language, Knowledge and Justification*, Cambridge: Cambridge University Press.
- Goldberg, S. (2010) *Relying on Others: An Essay in Epistemology*, Oxford: Oxford University Press.
- Goldberg, S. (2012) "A Reliabilist Foundationalist Coherentism", *Erkenntnis*, 77: 187-196.
- Goldberg, S. (2020) *Conversational Pressure: Normativity in Speech Exchanges*, Oxford: Oxford University Press.
- Goldberg, S. & Henderson, D. (2006) "Monitoring and Anti-Reductionism in the Epistemology of Testimony", *Philosophy and Phenomenological Research*, 72(3): 600-617.
- Goldman, A. (1967) "A Causal Theory of Knowing", *The Journal of Philosophy*, 64(12): 357-372.
- Goldman, A. (1976) "Discrimination and Perceptual Knowledge", *The Journal of Philosophy*, 73(20): 771-791.
- Goldman, A. (1979) "What Is Justified Belief?", in A. Goldman (ed.) (2012) *Reliabilism and Contemporary Epistemology: Essays*, Oxford: Oxford University Press, 29-48.

- Goldman, A. (1999) "Internalism Exposed", *The Journal of Philosophy*, 96(6): 271-93.
- Graham, P.J. (2000) "The Reliability of Testimony", *Philosophy and Phenomenological Research*, 61(3): 695-709.
- Graham, P.J. (2010) "Testimonial Entitlement and the Function of Comprehension", in A. Haddock, A. Millar, D. Pritchard (eds.) *Social Epistemology*, Oxford: Oxford University Press, 148-174.
- Greco, J. (1993) "Virtues and Vices of Virtue Epistemology", *Canadian Journal of Philosophy*, 23(3): 413-432.
- Greco, J. (2003) "Knowledge as Credit for True Belief", in M. DePaul & L. Zagzebski (eds.) *Intellectual Virtue: Perspectives from Ethics and Epistemology*, Oxford: Oxford University Press, 111-134.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998) "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test", *Journal of Personality and Social Psychology*, 74(6), 1464-1480.
- Grice, H.P. (1957) "Meaning", *The Philosophical Review*, 66(3): 377-388.
- Hawley, K. (2014) "Trust, Distrust and Commitment," *Noûs*, 48(1):1-20.
- Hawley, K. (2019) *How to be Trustworthy*, Oxford: Oxford University Press.
- Hieronymi, P. (2008) "The Reasons of Trust," *Australasian Journal of Philosophy*, 86(2): 213-236.
- Hookway, C. (2003) "How to be a Virtue Epistemologist", in M. DePaul & L. Zagzebski (eds.) *Intellectual Virtue: Perspectives from Ethics and Epistemology*, Oxford: Oxford University Press, 183-202.
- Harper, W. (1996) "Knowledge and Luck", *The Southern Journal of Philosophy*, 34: 273-283.
- Hetherington, S. (2019) "The Luck / Knowledge Incompatibility Thesis", in I.M. Church & R.J. Hartman (eds.) *The Routledge Handbook of the Philosophy and Psychology of Luck*, Oxon: Routledge, 295-304.

- Holton, R. (1994) "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy* 72(1): 63-76.
- Hume, D. (1748/1999) *An Enquiry Concerning Human Understanding*, Kitchener: Batoche Books.
- Jaswal, V.K. (2004) "Don't Believe Everything You Hear: Preschoolers' Sensitivity to Speaker Intent in Category Induction", *Child Development*, 75(6): 1871-1885.
- Jaswal, V.K., Croft, A.C., Setia, A.R. & Cole, C.A. (2010) "Young Children Have a Specific, Highly Robust Bias to Trust Testimony", *Psychological Science*, 21(10): 1541-1547.
- Jones, K. (1996) "Trust as an Affective Attitude", *Ethics*, 107(1): 4-25.
- Jones, K. (1999) "Second-Hand Moral Knowledge", *The Journal of Philosophy*, 96(2): 55-78.
- Jones, K. (2012) "Trustworthiness", *Ethics*, 123: 61-85
- Kant, I. (1781) *Critique of Pure Reason*, N.K. Smith (Trans.), Hampshire: Palgrave Macmillan, 2007.
- Kappel, K. (2014) "Believing on Trust", *Synthese*, 191(9): 2009-2028.
- Kassin, S.M. & Fong, C.T. (1999) "'I'm Innocent!': Effects of Training on Judgments of Truth and Deception in the Interrogation Room", *Law and Human Behavior*, 23(5): 499-516.
- Kelp, C. (2013) "How to Motivate Anti-Luck Virtue Epistemology", *Grazer Philosophische Studien*, 88: 211-225.
- Kelly, T. (2002) "The Rationality of Belief and Some Other Propositional Attitudes", *Philosophical Studies*, 110: 163-196.
- Keren, A. (2014) "Trust and Belief: A Preemptive Reasons Account," *Synthese*, 191(12): 2593-2615.
- Keren, A. (2020) "Trust and Belief", in Simon, J., (ed.), *The Routledge Handbook of Trust and Philosophy*, Routledge, 109-120.
- Klein, P. & Warfield, T.A. (1994) "What Price Coherence?", *Analysis*, 54(3): 129-132.

- Klein, P. & Warfield, T.A. (1996) "No Help for the Coherentist", *Analysis*, 56(2): 118-121.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004) "Trust in Testimony: Children's Use of True and False Statements", *Psychological Science*, 15(10): 694-698.
- Lackey, J. (2007) "Why we don't deserve credit for everything we know", *Synthese*, 158: 345-361.
- Lackey, J. (2008) *Learning from Words*, Oxford: Oxford University Press.
- Lackey, J. (2009) "Knowledge and Credit", *Philosophical Studies*, 142(1): 27-42.
- Lackey, J. (2011) "Testimony: Acquiring Knowledge from Others", in A. Goldman & D. Whitcomb (eds.) *Social Epistemology: Essential Readings*, 71-91.
- Lehrer, K. & Paxson Jr., T. (1969) "Knowledge: Undefeated Justified True Belief", *The Journal of Philosophy*, 66(8): 225-237
- Lipton, P. (1998) "The Epistemology of Testimony", *Studies in History and Philosophy of Science*, 29(1):1-31.
- Lipton, P. (2007) "Alien Abduction: Inference to the Best Explanation and the Management of Testimony," *Episteme*, 4(3): 238-251.
- Lyons, J. (2022) "Cognitive diversity and the contingency of evidence", *Synthese*, 200(202). <https://doi.org/10.1007/s11229-022-03660-8>
- Marušić (2015) *Evidence and Agency: Norms of Belief for Promising and Resolving*, Oxford: Oxford University Press.
- Mascaro, O. & Sperber, D. (2009) "The Moral, Epistemic, and Mindreading Components of Children's Vigilance Towards Deception", *Cognition*, 112: 367-380.
- McDowell, J. (1995) "Knowledge and the Internal", *Philosophy and Philosophical Research*, 55(4): 877-893.
- McDowell, J. (1998) "Knowledge by Hearsay", in J. McDowell, (ed.) *Meaning, Knowledge, & Reality*, Cambridge, MA: Harvard University Press, 414-443.
- McMyler, B. (2011) *Testimony, Trust, and Authority*, Oxford: Oxford University Press.

- Merricks, T. (1995) "On Behalf of the Coherentist", *Analysis*, 55(4), 306-309.
- Montmarquet, J. (1992) "Epistemic Virtue and Doxastic Responsibility", *American Philosophical Quarterly*, 29(4): 331-341.
- Moran, R. (2006) "Getting Told and Being Believed", in J. Lackey & E. Sosa (eds.) *The Epistemology of Testimony*, Oxford: Oxford University Press, 272-306.
- Nagel, T. (1979) *Mortal Questions*, Cambridge: Cambridge University Press.
- Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.
- Olsson, E.J., (2001) "Why Coherence is not Truth-Conducive," *Analysis*, 61: 236-241.
- Olsson, E.J., (2002) "What is the Problem of Coherence and Truth?", *The Journal of Philosophy*, 99: 246-272.
- Park, H.S., Levine, T., McCornack, S., Morrison, K. & Ferrara, M. (2002) "How people really detect lies", *Communication Monographs*, 69(2): 144-157.
- Pettit, P. (1995) "The Cunning of Trust", *Philosophy & Public Affairs*, 24(3): 202-225.
- Pollock, J.L. (1971) "Perceptual Knowledge", *The Philosophical Review*, 80(3): 287-319.
- Pollock, J.L. (1974) *Knowledge and Justification*, Princeton: Princeton University Press.
- Pollock, J.L. & Cruz, J. (1999) *Contemporary Theories of Knowledge*, 2nd edn. Plymouth: Roman & Littlefield Publishers.
- Pritchard, D. (2005) *Epistemic Luck*, Oxford: Oxford University Press.
- Pritchard, D. (2007) "Anti-Luck Epistemology", *Synthese*, 158: 277-297.
- Pritchard, D. (2010) "Anti-Luck Virtue Epistemology", in D. Pritchard, A. Millar, A. Haddock (eds.) *The Nature and Value of Knowledge: Three Investigations*, Oxford: Oxford University Press, 48-65.

- Pritchard, D. (2014) "Knowledge Cannot Be Lucky", in M. Steup, J. Turri & E. Sosa (eds.) *Contemporary Debates in Epistemology*, 2nd edn. Chichester: Wiley Blackwell, 152-164.
- Reid, T. (1764) *An Inquiry into the Human Mind: On the Principles of Common Sense*, D.R. Brookes (ed.), Edinburgh: Edinburgh University Press, 1997.
- Rinard, S. (2019) "Believing for Practical Reasons", *Noûs*, 53(4): 763-784.
- Shogenji, T. (1999) "Is Coherence Truth-conducive?", *Analysis*, 59: 338-345.
- Shogenji, T. (2001b) "Reply to Akiba on the probabilistic measure of coherence", *Analysis*, 61, 147-150.
- Sosa, E. (1985) "The Coherence of Virtue and the Virtue of Coherence: Justification in Epistemology", *Synthese*, 64(1): 3-28.
- Sosa, E. (1988) "Beyond Scepticism, to the Best of our Knowledge", *Mind*, 97(386): 153-188.
- Sosa, E. (1991) *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press.
- Sosa, E. (1999) "How to Defeat Opposition to Moore", *Philosophical Perspectives*, 13: 141-153.
- Sosa, E. (2003) "The Place of Truth in Epistemology", in M. DePaul & L. Zagzebski (eds.) *Intellectual Virtue: Perspectives from Ethics and Epistemology*, Oxford: Oxford University Press, 155-179.
- Sosa, E. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Oxford: Oxford University Press.
- Sosa, E. (2015) *Judgment and Agency*, Oxford: Oxford University Press.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010) "Epistemic vigilance", *Mind & Language*, 25(4): 359-393.
- Steinpreis, R.E., Anders, K.A., & Ritze, D. (1999) "The Impact of Gender on the Review of the CVs of Job Applicants and Tenure Candidates: A National Empirical Study", *Sex Roles*, 41(718): 509-528.

- Titus, L.M. & Carter, J.A. (2023) "What the tortoise should do: A knowledge-first virtue approach to the basing relation", *Noûs*, 1-26. doi: 10.1111/nous.12460.
- Turri, J. (2011) "Believing for a Reason", *Erkenntnis*, 74(3): pp. 383-397.
- Turri, J. (2012) "Is Knowledge Justified True Belief?", *Synthese*, 184 (3): 247-259.
- Unger, P. (1968) "An Analysis of Factual Knowledge", *Journal of Philosophy*, 65, 157-70.
- Vrij, A (2008) *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd edn., Chichester: Wiley.
- Vrij, A. (2014) "Detecting Lies and Deceit: Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection", in C.A. Berger (ed.) (2014) *Interpersonal Communication*, Berlin: De Gruyter, 321-346.
- Vosniadou, S. & Brewer, W.F. (1992) "Mental models of the earth: a study of conceptual change in childhood", *Cognitive Psychology*, 24: 535-585.
- Weiner, M. (2003) "Accepting Testimony", *The Philosophical Quarterly*, 53(211): 256-264.
- Williamson, T. (2000) *Knowledge and Its Limits*, Oxford: Oxford University Press.
- Williamson, T. (2014) "Knowledge First", in M. Steup, J. Turri & E. Sosa (eds.) *Contemporary Debates in Epistemology*, 2nd edn. Chichester: Wiley Blackwell, 1-9.
- Wolterstorff, N. (2001) *Thomas Reid and the Story of Epistemology*, Cambridge: Cambridge University Press.
- Wright, S. (2015) "In Defence of Transmission", *Episteme*, 12(1): 13-28.
- Wright, S. (2016) "Sincerity and Transmission", *Ratio*, 29(1): 42-56.
- Zagzebski, L. (1994) "The Inescapability of Gettier Problems", *The Philosophical Quarterly*, 44(174): 65-73.
- Zagzebski, L. (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press.