



The  
University  
Of  
Sheffield.

# Disagreement and Deep Disagreement: Should you trust yourself?

**Simon Reginald Barker**

Thesis submitted in partial fulfilment of the requirements  
for the degree of Doctor of Philosophy

**The University of Sheffield**  
**Faculty of Arts and Humanities**  
**Department of Philosophy**

October 2019



# Abstract

What, on realising disagreement, is an agent justified in believing when it comes to the substance of that disagreement? What significance ought we to afford disagreement in the practices by which we came to hold the beliefs over which we disagree? What do we do when the disagreement reaches down to the norms that determine the kinds of things we see as reasons for belief?

In this thesis I employ the methodological lens of ‘peer disagreement’ to develop a response to these questions that I call the ‘Higher-Order Trust’ approach to disagreement.

There are two important aspects to this position:

First, I suggest that a satisfactory theory of disagreement must consider both ‘The Problem of Ordinary Disagreement’ – roughly, the question of how epistemic peers ought to respond when they disagree in so far as they hold conflicting beliefs – and ‘The Normative Problem of Deep Disagreement’ – roughly, the question of how peers ought to respond when a disagreement in beliefs is explained by the disputants following conflicting epistemic norms.

Secondly, I suggest that the realisation of disagreement affords one *higher-order evidence* about the epistemic practices by which one came to hold the relevant beliefs. Thus, the question of how one ought to respond to the realisation of disagreement depends upon whether one has available some form of *epistemic self-trust* in the practices drawn into question. To understand the normative significance of disagreement, then, we first have to understand what it means to have epistemic self-trust. In this light, I suggest that there is a form of *affective self-trust* that can be available in cases of ordinary disagreement and realised deep disagreement between peers. When such trust is available, it may be rational to stay steadfast in the face of disagreement. Where it is not, one should respond in more conciliatory fashion.

# Acknowledgements

Well, that was tougher than expected...

Fortunately, I had sterling support throughout.

I should start by thanking my supervisors. Miranda Fricker, who I started this thing with, and whose work originally inspired in me the idea that epistemology could do something meaningful. One day I will write that piece... T. Ryan Byerly for stepping up in the last minute and not complaining once at being made to read my convoluted prose. Paul Faulkner for taking the reins half-way through and sticking by me as my own self-trust issues threatened to derail the whole thing.

Thanks to numerous other staff for being around and being helpful and making Sheffield a great place to study in (even more so when you are actually there). Honourable mentions, though, go to Jennifer Saul and Komarine Romdenh-Romluc for making the graduate experience fun, welcoming, and an education in how to be a better person. Not to mention interesting conversations about what can and can't be shown in a philosophy of art class. Thanks to the departmental staff, especially Kate Atkinson, Joanne Renshaw, and Sally Weston, who helped me navigate through the Kafkaesque corridors of University bureaucracy (not to mention Victoria House).

The graduate community as a whole has my thanks too. It is one of the best philosophy communities I have encountered, both for quality of discussion and the willingness for people to stick around for a drink anytime and everytime. I would name names, but I'll either miss a bunch of people out, or the list will fill most of the page up. You're all great. I shall give special mention to Charlie Crerar and Trystan Goetze, partners through surprisingly successful adventures in funding, manic and inefficient attempts to organise the perfect conference, and learning just how fun it is to edit a collection of philosophy essays (great collection by the way, just tremendous).

I'd thank my family, but they won't read this, so I will tell them in person.

Finally thanks to my ever steadfast, ever trusting, ever supportive partner Jaanika. Somehow you wrote your own PhD whilst keeping me from crumbling into total ruin... Pretty amazing if you ask me.

# Contents

Abstract .....	i
Acknowledgements.....	ii
Introduction.....	1
Disagreement, why care? .....	1
Methodology and theoretical background .....	3
Summary of Chapters .....	14
Chapter 1. Peers and Problems .....	19
Introduction .....	19
1.1 Peerhood and Ordinary Disagreement .....	19
1.2 The Normative Problem of Deep Disagreement.....	33
Conclusion.....	46
Chapter 2. The Problem of Ordinary Disagreement: Conciliationism.....	47
Introduction .....	47
2.1 Framing the debate .....	47
2.2 The content of Conciliationism .....	48
2.3 Arguments for Conciliationism.....	51
2.4 Arguments against Conciliationism.....	60
Chapter 3. The Problem of Ordinary Disagreement: Steadfast and Non-uniform views.....	68
Introduction .....	68
3.1 Self-trust and Steadfastness .....	68
3.2 The Justificationist View of Disagreement.....	79
3.3 Total Evidence.....	84
Chapter 4. Higher-Order Evidence and Defeaters.....	96
Introduction .....	96
4.1 What is higher-order evidence.....	97

4.2	Forms of defeat .....	99
4.3	Higher-order evidence and defeaters .....	106
	Conclusion.....	116
Chapter 5. Higher Order Evidence, Defeater-Defeaters, and Epistemic Self-Trust.....		117
	Introduction .....	117
5.1	Forms of defeater-defeater .....	117
5.2	Epistemic self-trust .....	120
	Conclusion .....	146
Chapter 6. The Higher-Order Trust Approach and Deep Disagreement .....		147
	Introduction .....	147
6.1	Re-introducing The Normative Problem of Deep Disagreement .....	147
6.2	Theoretical positions on the problem of deep disagreement .....	151
6.3	Epistemic norms as social practices .....	157
6.4	Responding to deep disagreement: The Higher-Order Trust approach extended .....	167
	Conclusion.....	173
	Conclusion .....	174
	Bibliography:.....	180

# Introduction

## Disagreement, why care?

Disagreement is ubiquitous in our social lives. We disagree with people about matters mundane and world-shaking. We disagree with people about everyday problems and problems that have nothing to do with our everyday lives. We disagree about literature, films, philosophy, politics, science, and everything else under the sun. Disagreements can be illuminating, interesting, and stimulating. At times, disagreements can even be fun. At other times, disagreements can be frustrating, annoying, and lead to nothing but an unhelpful impasse. Disagreements can be conducted with consideration and sensitivity and lead to respect, admiration, even friendship. Equally so, we can be rude, abusive, and threatening when we disagree with others; breeding nothing but anger, fear, hurt and pain, destroying lives and relationships in the process. If communication binds people together in social relations, disagreement can be both the glue and the solvent for those relations.

Perhaps most fundamentally of all, though, disagreement is a source of information. It is through disagreement that we find out that others believe differently to us and it is through disagreement that we find out why others believe differently to us. Indeed, at times, it is only through disagreement that we learn that others *can* believe differently to us. Everything else that disagreement is and leads to, follows from its role as a source of information. And, it is in this informational capacity that disagreement is interesting to epistemologists. Disagreement can only play this role, however, when we realise (or at least are in a position to realise) that others disagree with us. Thus, since epistemologists are interested in disagreement in its informational capacity, what we are interested in is ‘realised disagreement’.

Just as the non-epistemological dimensions of disagreement can lead to boon or bust, so it is that the information gained from ‘realised disagreement’ can be beneficial or detrimental to our epistemic endeavours. When the people we disagree with are in the right, or have good reasons for their beliefs, we might improve the accuracy of our beliefs or learn something new and important that relates to those beliefs if only we listen to what they have to say. When the people we disagree with are in the wrong, or have based their beliefs on faulty reasoning, being too open to their viewpoints can put at risk beliefs that are already in good standing.

Unless we are totally dogmatic or spineless when it comes to our beliefs, of course, none of us will respond to disagreement in the same way in all instances. Sometimes we will respond by steadfastly sticking to our own opinions, at other times we will be more conciliatory, lowering our confidence that we are right and at times even acceding the matter entirely to our interlocutors. At still other times disagreement can lead us to conclude that none of us have a clue what we are talking about. These are ways that we can – and do – respond to disagreement and the information that we acquire through realising disagreement. But, for any instance in which someone sticks to their guns, accedes to their opponent, suspends belief, or even withholds judgement entirely, they can also ask whether they should have responded to that disagreement in that way, or whether they should have responded differently. More generally, then, we can ask: How ought one to respond to disagreement when one realises that one disagrees with others?

There are two ways in which we might understand this question. One way – as per the points above – is in terms of the content of our responses to disagreement. In this sense, we might rephrase the question to ask: What, on realising disagreement, is an agent justified in believing when it comes to the substance of that disagreement?

As far as any actual disagreement goes, and as far as those party to that disagreement might be interested in its epistemological dimensions, we might suppose that it is this question that motivates that interest. So, any satisfactory account of disagreement ought to say something informative – and offer some guidance – as to what we are justified in believing in the face of disagreement. Just as we can ask ‘what an agent is justified in believing’, however, any answer we give to that question raises a more fundamental one: *why* is that attitude justified?

Thus, the second question at the heart of the epistemology of disagreement is what significance we ought to afford disagreement in the practices by which we come to hold the disputed beliefs. It is only by answering this question that we will be able to explain why one response or other to disagreement is justified, and so have any hope of providing anything more than piecemeal guidance on how we ought to respond to disagreement. A satisfactory theory of disagreement, then, must also offer an explanation of the epistemic nature of disagreement, its normative significance, and its place in our epistemic practices. In other words, what we are after is a ‘normative theory of disagreement’.

In this thesis, I approach these questions through the methodological lens of *peer disagreement*. Roughly, peer disagreements occur when, justificatorily speaking, no particular response to disagreement is clearly favoured. In such cases there is a genuine puzzle as to how disputants



should respond to the realisation of disagreement. So, the thought goes, consideration of such cases can give us a unique vantage point when it comes to theorising the normative significance of realised disagreement. Discussion of peer disagreement has formed the backbone of the literature on the epistemology of disagreement, and I follow that trend in this thesis.

There are two important aspects to the account of disagreement I develop through doing so.

First, I suggest that a satisfactory account of disagreement needs to consider both ‘The Problem of Ordinary Disagreement’ – roughly, the question of how epistemic peers ought to respond when they disagree in so far as they hold conflicting beliefs – and, what we might call, ‘The Normative Problem of Deep Disagreement’ – roughly, the question of how peers ought to respond when a disagreement in beliefs is explained by the disputants following conflicting epistemic norms.

Secondly, I suggest that the realisation of disagreement affords one *higher-order evidence* about the epistemic practices by which one came to hold the relevant beliefs. Thus, the question of how one ought to respond to the realisation of disagreement depends upon whether one has available some form of *epistemic self-trust* in the practices drawn into question. To understand the normative significance of disagreement, we first have to understand what it means to have epistemic self-trust.

In this introductory chapter, I shall consider what is a suitable methodology for theorising the normative and epistemological dimensions of disagreement, roughly characterise the core theoretical issues at play, and discuss a few further theoretical assumptions pertinent to the thesis proper. Then, I provide a chapter-by-chapter breakdown of the contents of this thesis.

## Methodology and theoretical background

### Why peer disagreement?

To address the question of how we ought to respond to disagreement it will be necessary to analyse some cases of disagreement – actual or hypothetical. From a methodological perspective, then, it is important to consider what sort of cases might inform discussion of the epistemology of disagreement.

The difficulty facing any attempt to theorise disagreement informed by cases is that the justificatory structure of any given disagreement will be multi-dimensional. Relevant to what an agent is justified in believing in the face of disagreement, for instance, will be information that they have about their interlocutor’s competence, about their own competence, the evidence on which their interlocutor

based their opinion, the agent's access to that evidence, the evidence on which the agent based their original opinion, and any information they might have about the state of their and their interlocutor's intellectual and cognitive capacities in the particular context. These dimensions of justification, of course, are not specific to the context of disagreement but are the kinds of consideration that influence what we are justified in believing in just about any social-epistemic interaction. What this means, though, is that for any disagreement in which there appears to be a clear and relatively un-controversial answer to the question of what doxastic response is justified, these more general dimensions of the justificatory context will provide sufficient explanation of that. Consideration of such cases, then, will not get us far in ascertaining what role and significance the realisation of disagreement *itself* ought to have in our epistemic practices.

To illustrate, this point, consider the following hypothetical example. Suppose that I believe that there are no negative consequences to rising global temperatures, but then come to realise that the vast majority of climate-scientists disagree with me. In such a situation, I clearly *ought* to accede to the expert consensus. However, we hardly need an account of the significance of my realisation of disagreement itself to explain that. For one thing, as I am not an expert and not qualified to assess the relevant evidence, I am entirely dependent upon climate-change experts for having an informed opinion on climate change. For another, the collective of climate-change experts is sure to be far more reliable on the issue at hand than me and I should realise that. Either of these features of the disagreement sufficiently explain why I am justified in (and only justified in) acceding to the expert consensus. Thus, whilst it is not necessarily so that my realisation of the disagreement has no normative significance in its own right, consideration of how I ought to respond *in this case* provides no clear evidence that it does have any such significance, nor of what kind of significance it might have if it does.

Consideration of cases of disagreement in which general features of the justificatory context favour one response or other, then, will not be informative when it comes to developing a proper understanding of the significance of disagreement – either in the cases at hand, or more generally. In so far as any inquiry into disagreement should be informed by the consideration of cases, then, we want to identify a class of disagreements in which, justificatorily speaking, there is nothing about the context of the disagreement that clearly favours one response over another. In this kind of case, there will be a genuine puzzle as to how disputants ought to respond to the realisation of disagreement. As such, considering the question of what disputants are justified in believing in the context can also inform our theorisation of the significance that the realisation of disagreement ought to have in our epistemic practices more generally. It is from this methodological perspective

that I suggest we should view the overwhelming focus in the literature on disagreement upon cases of disagreement between epistemic peers. Indeed, it is along similar lines that Jennifer Lackey suggests that:

In order to truly assess the significance of disagreement itself, there cannot be any relevant epistemic asymmetries between the parties to the dispute to shoulder the explanatory burden, in other words, such parties should be epistemic peers... (Lackey 2010: 208-09).

And Jon Matheson that:

We have seen that we cannot derive true interesting and universal principles about [...] everyday disagreements, but how much conciliation should be made in everyday disagreements will depend on what the correct story is about idealized disagreements coupled with how the asymmetrical reasons add up (Matheson 2014: 327).

I shall follow this approach in this thesis. It is, however, worth reiterating that the aim here is not to develop an account of peer disagreement only – but, rather, to develop an account of the normative significance of disagreement more generally through a methodological focus upon peer disagreement. Whilst I shall primarily be discussing cases of disagreement between peers, then, the hope is that by discussing these cases we will have a better grasp of the significance and role that disagreement ought to play in our practices in all cases.<sup>1</sup>

On that note, we can recognise a second consideration important to this methodological approach. Just as we might suspect that a consideration of everyday cases will not help us to isolate the significance of disagreement itself, so we might worry that an analysis of heavily idealised cases will fail to generalise to everyday cases in any interesting or informative way. An account of disagreement that is built solely on heavily idealised cases may, to borrow some jargon from the sciences, be ‘internally valid’ – yet, in so far as these cases are sufficiently dissimilar to any we encounter in our actual social engagements, fail to be ‘externally valid’. Where, as Clarke et. al. put it: “External validity”, or extrapolation, refers to the problem of exporting the results or methods of one study to a different population or setting (Clarke et al. 2014: 346).<sup>2</sup> Given the practical outlook of the interest in disagreement, an account of disagreement that fails the problem of external validity would hardly be a satisfying one. For that reason it will be important to identify a conception of epistemic peerhood that is suitable both for theorisation of the significance of disagreement *and* that we might hope to be somewhat representative of the disagreements we are

---

<sup>1</sup> That is not to say, of course, that a satisfactory theory of disagreement would have to provide an answer to the ‘response’ question in *all* cases of disagreement. That would seem to be an entirely unrealistic goal.

likely to encounter in experience. I address these methodological issues in more detail in Chapter 1.

## Justification

Given that the aim of this thesis is to answer questions about what agents are justified in believing in response to disagreement it is important to clarify the conception of justification that is relevant to these questions.

As I put it above, the underlying philosophical question that a normative account of disagreement needs to address is the question of what significance an agent *ought* to afford the realisation of disagreement in their epistemic practices. Presuming ‘ought implies can’ any normative significance attributed to the realisation of disagreement must be of a kind that the agent *can* afford that realisation.<sup>2</sup> Correspondingly, any doxastic response that an account of disagreement recommends in response to realisation of a particular disagreement must be one that the agent can come to – specifically – by affording that realisation the role and significance described.

From an epistemological perspective, this means that a viable account of disagreement cannot recommend responses or ways of responding to disagreement that depend entirely upon facts to which the agent does not have access. For, generally, if an agent does not have some kind of access to some fact X, then, that agent cannot afford X any kind of significance in their epistemic practices. In which case, it cannot be so that that agent *ought* to afford X any kind of significance when responding to disagreement.

---

<sup>2</sup> There is debate over whether the *epistemic ought* implies can, the *locus classicum* of which is Alston’s (1988) influential argument against deontic conceptions of justification. Alston’s argument presents us with a dilemma: either, he claims, we accept that we have voluntary control over our beliefs or we reject deontic conceptions of justification. Given that it seems we do not have voluntary control over our beliefs, so the argument goes, we ought to reject deontic conceptions of justification (see Feldman (1988a, 2001) and Wolterstorff (Wolterstorff 2010) for responses to this argument on its own terms; see Neta (2014) and Chuard and Southwood (2009) for useful summaries of the debate.) As Chuard and Southwood point out, Alston’s presentation of the argument appears to presuppose that ‘ought implies can’ should be read as saying that ‘an agent *cannot*  $\Phi$ , unless that agent has *voluntary control* over  $\Phi$ ’ing’. We might refer to this reading of ‘ought implies can’ as the principle that ‘ought implies voluntary control’. But as Chuard and Southwood, I think correctly, argue there are alternative formulations of the ought implies can principle that we might prefer, e.g. ‘ought implies logically possible’ or ‘ought implies can do otherwise’. And, if we accept one of these weaker formulations, the conflict between deontic conceptions of justification and doxastic involuntarism dissolves. As far as the current discussion goes, then, it is worth pointing out that the argument in the main text does *not* presuppose that ‘ought implies voluntary control’, but works just as well if we presume that ‘ought implies logical possibility’ or ‘ought implies can do otherwise’. Thus, and in so far as much of the motivation for denying that the epistemic ‘ought implies can’ stems from attempts to respond to Alston’s argument, I take it that this suffices for us to sidestep further debate over the nature of the epistemic ought.

In light of this, I would suggest that questions of how we *ought* to respond to disagreement are not questions about what we are justified in believing when justification is understood in a simple externalist or objective sense – i.e. as determined solely by facts about reliability, or as determined solely by facts about the objective relations between propositions. Instead, questions of how we ought to respond to disagreement pertain to justification conceived in a narrow *internalist* sense. By this, I mean the form of epistemic support we have for our beliefs that, as Richard Foley puts it, is ‘closely connected to internally defensible’ believing (Foley 2001: 13) and, in Tyler Burge’s words, corresponds to the idea of *having available* a justification for one’s belief. To quote Burge:

[J]ustifications in the narrow sense, involve reasons that people have and have access to. These may include self-sufficient premises or more discursive justifications. But they must be available in the cognitive repertoire of the subject. (Burge 2013b: 230).

Justification in this narrow sense is determined by the content of the beliefs and other attitudes that the agent holds and the relations between these beliefs and other attitudes. Since these are facts that an agent, in principle, can access – they are also facts in respect to which it can be so that an agent *ought* to believe, or act epistemically, one way or another.

Following Burge (2003, 2013b), then, let us use ‘warrant’ as a generic term to refer to any species of epistemic support and ‘justification’ to refer to the kind of warrant that an agent has for their beliefs when that agent has access, on reflection, to an argument with the content of the relevant belief as its conclusion.<sup>3</sup> If we use the term *reason* to refer to the kinds of consideration that provide the content of such arguments, we can also say that justification is the kind of warrant that corresponds to an agent’s having *epistemic reasons* for what they believe. In that light, let us describe the relationship between epistemic reasons and justification such that: if P is an *epistemic reason* for S to believe Q, then, in lieu of defeaters, were S to believe Q and that belief be properly based upon P, S would be justified in believing Q.<sup>4</sup>

Justification used in this way might also be described as *subjective justification*. I.e. It depends upon the relations between agent’s reasons for belief and the content of their beliefs as *the agent takes those relations to be*, not (at least directly) as they objectively are.<sup>5</sup> Thus, whatever role an agent ought to afford the realisation of disagreement will depend upon the normative significance of that

---

<sup>3</sup> See McDowell (1993: 196) for a description of justification along very similar lines.

<sup>4</sup> Unless otherwise specified, talk of ‘reasons for belief’ in this thesis should be understood as talk of epistemic reasons for belief

<sup>5</sup> For good discussions of the distinction between objective and subjective justification/theories of justification, see Kvanvig (1984) and Feldman (1988b)

realisation vis-à-vis what they are subjectively justified in believing. Likewise, what they ought to believe in response to disagreement will be a matter of what they are subjectively justified in believing in the face of disagreement.

Given this suggestion, it is important to note a further feature of justification so conceived. As I understand it, an agent's beliefs will not be justified if they doubt/deny or ought to doubt/deny that their beliefs are objectively justified. We can state this as the following condition on justification:

*An objective attitude condition on justification (OAC):*

If S believes Q on the basis of P and S doubts/denies, or ought to doubt/deny, that P is *objectively* good reason to believe Q, S is not *subjectively* justified in believing Q.<sup>6</sup>

Thus, if S ought to believe that P is objectively *not* reason to believe Q, given OAC, S could not bootstrap her way into justification for believing Q by supposing that 'P is nonetheless reason *for me* to believe Q'. What motivates OAC is the fact that subjective justification is still *epistemic* justification. And, as such, the kinds of argument that underwrite any justification that an agent has for their beliefs *must* be arguments to the truth of those beliefs. An argument that includes only the premises 'P', 'P is a reason for me to believe Q', and 'P is not objectively a reason to believe Q', on the other hand, will not be an argument to the truth of Q – for just the reason that to doubt or deny that 'P is objectively a reason to believe Q' is to doubt or deny that P and Q are connected in such a way that the truth of P indicates the likely truth of Q.<sup>7</sup>

Whilst subjective justification is not determined (at least directly) by facts about reliability or the objective relations between an agent's evidence and their beliefs, then, it is conditional on the attitudes an agent has or the attitudes that they ought to have about the relevant facts of reliability and objective relations between their evidence and their beliefs. As we shall see later in the thesis,

---

<sup>6</sup> Compare Feldman's description of what he calls 'moderate subjective justification (MSJ)': 'S is *subjective epistemically justified* in believing p iff S has good reasons to think that S is objectively epistemically justified in believing p' (Feldman 1988b: 414). As per OAC, 'subjective justification' so conceived is determined by the reasons one has available pertaining to the objective status of one's beliefs. Feldman's definition, however, is a positive condition – in contrast to the negative nature of OAC. With some additional caveats, this allows Feldman to ground the claim that MSJ implies objective justification. Without going into detail on Feldman's argument, we can simply note that, since OAC does not require having reasons to believe one's beliefs are objectively justified – only an absence of reasons to doubt/deny that – this entailment does not follow on OAC.

<sup>7</sup> Following this, we might note, it is natural to think of the kinds of consideration that count as epistemic reasons for belief as being roughly co-extensive with the kinds of consideration that count as evidence.

this caveat to how we understand justification – and so questions about how we ought to respond to disagreement is an important one.

## **Scepticism**

As discussed so far, discussion of the normative dimensions of disagreement concerns the question of what one would be justified in believing in the face of disagreement. We might note, then, that this discussion rests on the assumption that justified belief is not only possible, but, at least in our world, a common state of affairs. Or, to put it otherwise, the debate about the normative significance of disagreement presupposes that global or near-global scepticism about justification is wrong. At the same time, since certain of the responses that we might have to disagreement are of a sort with scepticism – i.e. suspending belief or withholding judgement altogether – the possibility for a normative account of disagreement that has significant sceptical consequences is on the table from the get go. Given how ubiquitous disagreement is, however, significant sceptical consequences within the context of disagreement may threaten global or near-global scepticism. Granted, since it would depend upon the contingent facts about the prevalence of disagreement within society, this would be a kind of ‘empirical’ rather than ‘philosophical’ scepticism. Crucially, though, it would undermine the working assumption that justified belief is both possible and a typical state of affairs nonetheless. For this reason, I assume throughout this thesis that any significant sceptical consequences that follow from a given account of disagreement, or other relevant theoretical concept, are at least *prima facie* evidence against that account. As we shall see at various points, this assumption is not without importance to the debate.

## **Belief, confidence, and doubt**

As noted, the epistemology of disagreement is to a large degree concerned with the question of what an agent is justified believing when it comes to the substance of that disagreement.

One way in which we might discuss this question is in terms of the tri-partite schema of ‘belief’, ‘disbelief’, and ‘suspension of judgement’. Another approach is to parse talk of belief as talk of the *rational confidence* or *credence* an agent has or ought to have in the truth of the target propositions.

I discuss these two ways of modelling the possible responses to disagreement in detail in Chapter 2. For the moment, though, it is enough to note that this thesis will for the most part follow the language of belief. Where it serves the argument best or better reflects the literature, however, I shall at times talk of cases in terms of credences or rational confidence.

Throughout the thesis much reference is also made to ‘doubt’, ‘doubts’ and ‘reasons to doubt’. The term ‘doubt’ can be and is used in a range of ways, in everyday language and in philosophical literature. One might without contradiction say something like ‘I am confident that the bill will have the votes to pass, but I must admit I have my doubts’. Here ‘doubt’ is used merely to express the fact that the speaker’s confidence in the truth of the relevant proposition is not at the level of certainty. One might also say something like ‘Do I think the bill has the votes to pass? I doubt it.’ Here ‘doubt’ has stronger connotations and is used, not just to qualify the speaker’s confidence, but to express the absence of confidence. Used in the first weaker sense, we might say, doubt does not preclude belief, or other attitudes, such as trust, faith, or optimism, that in some way entail a presumption of truth (or likely truth). ‘Doubt’ used in the latter stronger sense does. In normal usage, context is usually enough to make clear which of these senses of doubt is in use. Nonetheless, it is worth clarifying that in this thesis I use the term doubt strictly in the stronger sense.<sup>8</sup> Thus, let us say that an agent can be said to doubt Q when they have, in some sense, taken Q under consideration, yet have taken no attitude toward Q that would involve the presumption that Q is, or is likely to be, true. Where I offer principles such as OAC that include a ‘doubt’ condition or where I refer to ‘reasons to doubt’, the reader should have this technical sense of doubt in mind.

### **Uniqueness and reasonable disagreement**

A theoretical issue that runs somewhat orthogonal to the question of ‘how we ought to respond to disagreement’ is the question of whether there can be such a thing as ‘reasonable disagreement under full disclosure’ – whereby *reasonable disagreement* would occur when disputants realise they disagree, share all of their evidence, and are equally justified in maintaining their original positions. I will not be engaging with this debate in any detail in the main thesis, but it is worth taking some space to explain why.

Much of the discussion around the possibility of reasonable disagreement is framed around the plausibility of a thesis that Richard Feldman refers to as the Uniqueness thesis (Feldman 2007; Sosa 2010; White 2005). As Feldman describes it:

This is the idea that a body of evidence justifies at most one proposition out of a competing set of propositions (e.g., one theory out of a bunch of exclusive alternatives) and that it justifies at most one attitude toward any particular proposition. As I think of things, our options with respect to any proposition are believing, disbelieving, and suspending

---

<sup>8</sup> Though I hope the reader will bear with me if I do slip on occasion. I am confident that I have not, but in the tradition of the preface paradox, I must admit I have some doubts!



judgment. The Uniqueness Thesis says that, given a body of evidence, one of these attitudes is the rationally justified one. (Feldman 2007: 205).

As Feldman, I think rightly, points out, if ‘Uniqueness’ is true, then, reasonable disagreement will not be possible. Equally so, we might suppose, if reasonable disagreement is not possible, that will tell us something about the normative significance of disagreement. I.e. No matter what the disputants were justified in believing before they realised their disagreement, at least one of those party to a disagreement ought to revise their beliefs once they have realised that. By and large, this seems correct. Thus, it is also tempting to suppose that the opposite is equally true. I.e. If Uniqueness is false, then, reasonable disagreement is possible, and so it can be so that none of those party to disagreement ought to revise their beliefs, even under full disclosure. Hence, so the thought goes, the *falsity* of Uniqueness would also tell us something important about the normative significance of disagreement.

It is something like this second line of thinking, I think, that motivates a good deal of the attention on the Uniqueness thesis, not only in relation to reasonable disagreement but also in relation to the normative significance of disagreement (for instance Ballantyne and Coffman 2012; Christensen 2016a). Given that, it may seem something of an omission from this thesis that I do not discuss Uniqueness or reasonable disagreement (beyond this section). My reason for not doing so is that I do not think this line of argument is correct. Whilst the questions about reasonable disagreement and Uniqueness are interesting in their own right, and not entirely unrelated to the concerns of this thesis, they are not of central concern to the questions about disagreement’s normative significance and I shall not take a stand on either. Let me explain why.

Most plausibly, Uniqueness is true on an account of justification that leans strongly into the idea of objective justification, it is less compelling once we move further away from such a picture to place greater emphasis upon the subjective features of an agent’s epistemic standing.<sup>9</sup> Whilst I have not suggested that we necessarily need to choose between objective and subjective conceptions of justification,<sup>10</sup> I have suggested that the form of justification relevant to the current discussion is

---

<sup>9</sup> I include the ‘leans strongly into’ caveat since defences of Uniqueness have primarily been forwarded by those who endorse ‘Evidentialist’ theories of justification – which are internalist theories of justification. What seems to be common to those who offer these defences, however, is the inclusion of some caveat to suggest that the relationships of evidential fit on which justification depends hold independently of the agent’s perspective on those relations. So, for instance, Roger White suggests that ‘evidential support relations hold necessarily. If E supports P, then necessarily E supports P.’ (White 2013: 313) and Jon Matheson that he assumes that ‘justification is both objective and real. By ‘objective’ I mean that there are objective standards regarding what epistemic reasons there are and how good various epistemic reasons are.’ (Matheson 2015: 9).

<sup>10</sup> Following Foley, I take it that these are better conceived as concepts dealing with different questions in epistemology (Foley 2001).

subjective justification. At the least, then, I would suggest that any account of the significance of disagreement that rests solely upon Uniqueness will need to do some heavy lifting to defend that foundation. Let's suppose for now, then, that Uniqueness is not true in the way that would count towards the current discussion. Even on this supposition, as I argue below, the falsity of Uniqueness would not yet entail that reasonable disagreement under full disclosure is possible. And, as such, it does not settle any questions about the normative implications of realised disagreement.

Following White (2005), let's refer to any view that rejects uniqueness as a 'Permissivist' view. There are a number of forms of Permissivism we might entertain. We might, for instance, allow that two people could be justified in having different levels of *confidence* in a proposition Q in respect to a single body of evidence P, but not justified in outright differences in *belief*. Alternatively, we might allow that one person be justified in believing Q, and the other justifiably suspend judgement on Q – but not that one justifiably believes Q and the other  $\sim$ Q. We can refer to these views respectively as 'Weak' and 'Moderate Permissivism'. Alternatively, we might go further and allow that one person could be justified in believing Q on the basis of P and another justified in believing  $\sim$ Q. Call this last position 'Strong Permissivism'. If the possibility of reasonable disagreement is entailed by any of these views, it will be entailed by 'Strong Permissivism'. So, let us consider whether that is the case. Suppose the following can be true:

Strongly Permissive Case (SPC): S<sub>1</sub> justifiably believes Q on the basis of P; S<sub>2</sub> justifiably believes  $\sim$ Q on the basis of P.

At the least, the possibility of cases like SPC will entail that cases like the following are possible:

Reasonable Unrealised Disagreement (RUD): At t<sub>1</sub>, S<sub>1</sub> justifiably believes Q on the basis of P; S<sub>2</sub> justifiably believes  $\sim$ Q on the basis of P; S<sub>1</sub> and S<sub>2</sub> are not aware of the other's beliefs about Q.

RUD describes a situation in which individuals do not realise that they disagree. As we are using the term, however, it is a condition of *reasonable disagreement* that those involved realise that they disagree and have fully disclosed to each other the evidence upon which they based their conflicting beliefs. If the falsity of Uniqueness entails that reasonable disagreement is possible, then, the possibility of SPC will entail also that the following is possible:

Reasonable Realised Disagreement (RRD): At  $t_2$ ,  $S_1$  justifiably believes  $Q$  on the basis of  $P$  and realises that ' $S_2$  believes  $\sim Q$  on the basis of  $P$ ';  $S_2$  justifiably believes  $\sim Q$  on the basis of  $P$  and realises that ' $S_1$  believes  $Q$  on the basis of  $P$ '.

Presumably, RRD will only be true in cases where RUD could also be true<sup>11</sup> – but, that being so, SPC will only *entail* the possibility of RRD, if there can be no cases in which RUD is true but RRD is not possible. However, and irrespective of what the other is justified in believing, for each of  $S_1$  and  $S_2$ , RRD is informationally distinct from their position in RUD. I.e. in RUD the disputants do not have any information about what the other believes, in RRD they do. Thus, if SPC entails the possibility of RRD, it will only do so if it is not the case that the kind of information that is acquired in realised disagreement under full disclosure *always* has normative significance. To say that it does always have such significance, of course, would be to take a substantive position on the normative significance of disagreement. The important point in the current context, though, is that to deny that such information always has significance is *also* to take a substantive position on the normative significance of disagreement. Whilst 'Strong Permissivism' clearly does entail the possibility that individuals could be justified in holding conflicting beliefs on the same evidence when they *do not realise that they disagree*, whether or not it entails the possibility of reasonable disagreement as understood here will crucially depend upon what is the correct account of the normative significance of disagreement. Similarly, for any weaker permissivist views. Thus, far from it being so that establishing the falsity of Uniqueness would settle questions about the normative significance of disagreement – it is the other way around. It is only by answering questions about the normative significance of realised disagreement that we can settle the question of whether the falsity of Uniqueness entails the possibility of reasonable disagreement.

The quasi-modal form of this argument also helps us to recognise the parallel point. Just as the falsity of Uniqueness need not entail that reasonable disagreement is possible, so, establishing the possibility of reasonable disagreement would not entirely settle the questions about the normative significance of disagreement. For, even so, the key philosophical question would be *why* is reasonable disagreement possible? Would it be possible because, in *all* cases of disagreement under full disclosure, disputants are permitted to presume that their opponent's beliefs are not justified? Or would reasonable disagreement only be possible in *some* cases – where disputants had some

---

<sup>11</sup> To say otherwise would effectively be to say that the realisation of disagreement could justify  $S_1$  and  $S_2$  in holding conflicting beliefs on the basis of the same evidence when they were not justified in doing so prior to realising their disagreement. If disagreement is significant in anyway it is not in that way!

form of positive support for favouring their own views? Reasonable disagreement could be possible, yet these questions remain.

In summary, settling the questions about reasonable disagreement and Uniqueness will not get us far in developing a satisfactory normative theory of disagreement. Interesting as these debates are, I shall not engage with them further in this thesis.

That covers the main theoretical and methodological assumptions in this thesis. With that done, I provide a summary of the chapters.

## Summary of Chapters

### Chapter 1: Peers and Problems

In this chapter, I consider how we should understand epistemic peerhood to suit the methodological aims of the thesis and what are the key questions that an epistemological theory of disagreement should address.

In section 1.1, I introduce three conceptions of peerhood: *ideal peers*, *justificatory peers* and *effective peers* and consider the merits of employing each of these as a methodological starting point for theorising disagreement. Those party to a disagreement will be peers in the ideal sense if none have the relative epistemic advantage vis-a-vis the substance of their disagreement, justificatory peers if they have reason to believe that that is so. Disputants will be effective peers when none have reason to believe that any specific parties have the relative epistemic advantage, where those reasons are independent of the substance of the disagreement. I suggest that the concept of effective peerhood should be the methodological starting point from which to develop a normative theory of disagreement. Thus, the first issue that an epistemology of disagreement needs to address is the problem of determining how disputants should respond to the realisation of disagreement when they disagree with their effective peers. Call this ‘The Problem of Ordinary Disagreement’. I elaborate on these claims in section 1.1

In section 1.2, I introduce the notion of *deep disagreement*, discuss the different ways in which we might characterize such disagreements, and explain why we should see ‘The Normative Problem of Deep Disagreement’ – how to respond to deep disagreement with an effective peer when that disagreement is deep – as distinct from ‘The Problem of Ordinary Disagreement’.

## **Chapter 2: The Problem of Ordinary Disagreement – Conciliationism**

In the first section of this chapter, I offer a simple taxonomy of the possible accounts of disagreement one might offer in response to ‘The Problem of Ordinary Disagreement’. ‘Conciliationism’ says that one ought always to revise one’s beliefs in the face of disagreement with an effective peer. ‘Steadfast Views’ say that one is always permitted to maintain one’s beliefs. ‘Non-Uniform’ views say that the appropriate response will vary from case to case. Having introduced this taxonomy, the rest of the chapter considers the case for and against the Conciliationist view of ordinary disagreement.

In section 2.2, I consider two ways in which we might understand the requirement on Conciliationism to revise one’s beliefs – in terms of categorical belief states, or in terms of subjective credences. In section 2.3, I discuss and develop what I take to be the two core arguments for Conciliationism: the ‘Truthometer Analogy’ and the ‘Independence’ argument. The Truthometer Analogy suggests that disagreement with an effective peer is analogous to finding oneself faced by conflicting readings from equally reliable thermometers. If the analogy holds, Conciliationism is the correct response to ‘The Problem of Ordinary Disagreement’. The Independence argument attempts to ground the intuitions behind the Truthometer Analogy in a more substantive principle of rationality. The principle of ‘Independence’ states, roughly, that it is not permissible to rely on considerations that are not independent of the substance of disagreement to justify favouring one’s own position in that disagreement. The main motivation for this principle is that it rules out dogmatic responses to disagreement. Since, effective peers have no independent reason to favour one view or other, Conciliationism follows as a matter of course.

In section 2.4 I consider the suggestion that Conciliationism leads to unpalatable sceptical consequences. Having developed this line of argument I suggest that we have motivation to consider the feasibility of a steadfast or non-uniform view of disagreement.

## **Chapter 3: The Problem of Ordinary Disagreement – steadfast and non-uniform views**

In this chapter, I consider three alternatives to Conciliationism: the ‘Default Self-Trust View’ variations of which have been forwarded in defence of steadfastness by David Enoch and Karl Schafer; Jennifer Lackey’s non-uniform ‘Justificationist View’; and Thomas Kelly’s non-uniform ‘Total Evidence View’.

The Default Self-Trust View claims that, given a general *entitlement* to trust my epistemic practices, I can count the fact that another person disagrees with me as evidence against their competence. I agree that self-trust is of central importance to the epistemology of disagreement and return to the topic in Chapter 5. I suggest, though, that since entitlements are defeasible, the question remains whether that entitlement is defeated by the realisation of peer disagreement. If it is, then, the ‘Default Self-Trust View fails.

On Lackey’s ‘Justificationist View’, the steadfast response to disagreement may be justified in cases where an agent has *personal information* that can ground an argument to the conclusion that the other party to the disagreement was subject to some kind of performance error. In other cases, conciliation is required. I point out that the conditions under which this argument will be available will be met in only a small range of possible disagreements. As such, the main shortcoming of this account is that it appears to be vulnerable to the unwelcome sceptical consequences of Conciliationism.

Kelly’s ‘Total Evidence View’ rest on two ideas: that we ought to reject the principle of Independence if there is a way of accommodating its anti-dogmatic motivations within a more general epistemic framework; and that we can do so by differentiating between first-order and higher-order evidence. I agree with Kelly on both points but suggest that Kelly’s conception of higher-order evidence fails to rule out dogmatic responses to disagreement. To develop a satisfactory account of disagreement around the concept of higher-order evidence, we need to consider more generally the question of how such evidence functions and how the higher-order evidence acquired in case of disagreement with an effective peer fits within that more general picture. I do this in Chapter 4

## **Chapter 4: Higher-order evidence and defeaters**

In this chapter, I consider how best to understand the normative significance of higher-order evidence.

In section 4.1, I offer examples of negative higher-order evidence outside of the context of disagreement and suggest that the key to understanding the normative significance of higher-order evidence lies in understanding how, and if, it can be accommodated within a more general schema of defeaters and justification. In section 4.2, I lean upon John Pollock’s influential account of defeasible reasoning to introduce the relevant kinds of defeater: *rebutting* and *undercutting*. In addition, I suggest that Pollock’s definition of undercutting is insensitive to an important

distinction and offer a further distinction between *contextual undercutting defeaters* and *general undercutting defeaters*. As a result, three kinds of defeater need to be recognized.

In section 4.3, I consider whether any of the higher-order evidence cases can be explained in terms of rebutting, contextual undercutting, or general undercutting. I suggest they can and that higher-order defeaters of all three kinds are possible, thus, whether a conciliatory or steadfast response to disagreement is required will depend simply on whether the agent in question has a defeater-defeater for the higher-order evidence.

## **Chapter 5: Higher-order evidence, defeater-defeaters, and epistemic self-trust**

In this chapter, I turn to consider the kinds of consideration that might count as defeater-defeaters in cases of defeat from higher-order evidence, including cases of ordinary disagreement between effective peers. Echoing the distinction between first-order and higher-order forms of defeater forwarded in the previous chapter, I suggest in section 5.1 that, that where one has independent evidence that rebuts or undercuts the defeating higher-order evidence, one will have a *first-order defeater-defeater* for that higher-order evidence. In cases of disagreement between effective peers no such evidence is available. Thus, I suggest that, unless there can also be *higher-order defeater-defeaters*, conciliation will always be the appropriate response to the realisation of such disagreement. In the rest of the chapter I address the question of whether there can be higher-order defeater-defeaters. Following the suggestion that defeating higher-order works by preventing one from trusting one's epistemic practices, I do so by exploring what we might mean by epistemic self-trust, what role it has to play in the structure of justification and reasons, and what forms self-trust might take. I conclude by introducing what I call the 'Higher-Order Trust' approach to disagreement the core claim of which are that there is a form of *affective self-trust* that can be available in case of disagreement with an effective peer. When it is steadfastness is a possibility, when it is not, the conciliatory response to disagreement will be appropriate.

## **Chapter 6: The Higher-Order Trust approach and deep disagreement**

In the final chapter of the thesis, I extend the Higher-Order Trust approach to disagreement developed in Chapter 4 and 5 to 'The Normative Problem of Deep Disagreement'. In sections 6.1 and 6.2, I re-introduce 'The Normative Problem of Deep Disagreement' and discuss the possible theoretical responses to that problem. In 6.3, I situate the notion of *epistemic norms* within the broader understanding of epistemic practices and self-trust forwarded in the previous chapter. In the final section of the chapter, I consider in more detail what the Higher-Order Trust account

has to say about how those party to deep disagreement between effective peers ought to respond in two types of case: cases in which disputants are not in a position to recognise the nature of their disagreement, and cases in which disputants are in a position to recognise the nature of their disagreement.



# Chapter 1. Peers and Problems

## Introduction

What role and significance ought we to afford disagreement in the way that we come to form and hold beliefs, and why?

In the introduction to this thesis, I suggested that we can address this question through consideration of cases in which, justificatorily speaking, no one side or the other has a clear epistemic advantage in respect to the substance of the disagreement. In the vernacular of the debate, these are disagreements between *epistemic peers*.

In section 1.1, I introduce three conceptions of peerhood: *ideal peers*, *justificatory peers* and *effective peers* and consider the merits of employing each of these as a methodological starting point for theorising disagreement. I argue that we should use the notion of *effective peers* as our methodological starting point. Thus, the first issue that an epistemology of disagreement needs to address is the problem of determining how disputants should respond to the realisation of disagreement when they disagree with their effective peers. Call this ‘The Problem of Ordinary Peer Disagreement’.

In section 1.2, I introduce the notion of *deep disagreement*, discuss the different ways in which we might characterize such disagreements, and explain why we should see ‘The Normative Problem of Deep Disagreement’ – how to respond to deep disagreement with an effective peer when that disagreement is a deep disagreement – as distinct from ‘The Problem of Ordinary Disagreement’.

## 1.1 Peerhood and Ordinary Disagreement

### 1.1.1 What is ordinary disagreement?

Consider the following two cases, the first a slightly modified version of a much discussed example from David Christensen (2007), the second a somewhat higher-stakes case:

DINERS-1

Jack and Jill go to dinner with a group of friends. At the end of the dinner, the group agree to split the bill evenly, including a 20% tip. They leave it to Jill to work out what each owes. Having calculated the share of the bill in her head, Jill states that ‘Each of us owes \$45’. To her surprise, Jack responds to this, claiming ‘That isn’t right. Each of us owes \$43’. Jack and Jill have dined together many times in the past and always split the bill. In the majority

of past cases they have agreed on the total, and in the cases where they have not, neither has a better track-record than the other.

#### DOCTORS-1

Doctor Quinn is treating a patient, X, with Condition-C. Quinn is trained and experienced in the relevant field and has assessed the available case evidence – including patient history and test results. Quinn has also checked and doublechecked the relevant medical guidelines. Following her assessment of the case-evidence and guidelines, Quinn believes that the patient is most likely to recover under treatment option-A. Quinn then comes to realise that a second doctor on the case, Doctor Zaius, having considered the same case-evidence and guidelines, believes that the patient is most likely to recover under treatment option-B. Quinn has every reason to believe that Zaius is as well trained and experienced as she.

Despite the clear and significant differences between these two cases, both practically and epistemically, there are a number of similarities relevant to the current discussion.

Firstly, Diners-1 and Doctors-1 are alike in so far as, epistemically speaking, they are both quite *ordinary disagreements*. In each case, what is at issue is a non-normative matter of fact, and the substance of the disagreement over that matter can be well characterised in terms of the disputants holding conflicting beliefs on that issue. Secondly, they are alike in so far as neither member of the two pairs appears to have any *epistemic advantage* over the other of the kind that would allow them to settle the disagreement without controversy. In neither case does one of other of the pair appear to have better evidence than the other (indeed, in both cases they share at least a significant part of their evidence) or be more competent when it comes to assessing that evidence, and in neither case is there any evidence that one party or other was subject to any specific kind of performance impairment or enhancement.

In respect to these points, we might say that these are cases of disagreement between *epistemic peers*. And so, more specifically that these are cases of *ordinary peer disagreement*. In this kind of case, there is a genuine puzzle about how disputants ought to respond to realisation of the disagreement. Ought they to retain their original beliefs – even whilst it seems that the only evidence any of them might have to justify doing so is itself a matter of contention? Or, ought they to revise the beliefs – even whilst it seems that if they could be confident that they competently assessed the evidence they share, they could also be confident that their opponent had not?

This gives us a general idea of the puzzle presented by peer disagreement. As we shall see in the following discussion, however, there are a number of different relationships between those party to disagreement that we might describe in terms of epistemic peerhood – and determining which of these generate this puzzle and which are apt for the methodological focus on peer disagreement is crucial to the general discussion of this thesis. I turn to discuss these different relationships in the next section.

### 1.1.2 Kinds and contours of epistemic peerhood

As I have characterized them so far, peer disagreements arise when none of those party to disagreement have the kind of epistemic advantage that would allow them to determine without controversy which of the respective positions is more likely to be correct than the others. Given this characterization, perhaps the most straightforward way to conceive of the peerhood relation is by reference to the likely truth of the beliefs held by the various parties to the disagreement, conditional on the substance of that disagreement. Amongst those who suggest that we think of the peerhood relation in this way are Adam Elga and David Enoch. So Elga states that:

[Y]ou count your friend as an epistemic peer with respect to an about-to-be judged claim if and only if you think that, conditional on the two of you disagreeing about the claim, the two of you are equally likely to be mistaken. (Elga 2007: 500).

And Enoch:

[B]y your “peer” I will understand someone who is, somewhat roughly, antecedently as likely as you are to get things right (on matters of the relevant kind) (Enoch 2010: 956).

Let’s say that those party to a disagreement who are peers in this sense are *Ideal Peers*, and define what it is for disputants to be such as follows:

*Ideal Peers*

IF  $S_1$  and  $S_2$  disagree and, *independent of the substance of the disagreement*, neither have the relative epistemic advantage when it comes to determining the likely truth of the disputed propositions,  $S_1$  and  $S_2$  are *Ideal Peers*.<sup>12</sup>

Epistemic peerhood, understood this way, is a relationship determined by the facts about the context of the disagreement as they are and irrespective of any information that the disputants

---

<sup>12</sup> To be clear, neither Elga nor Enoch suggest that this relationship is sufficient to generate the characteristic problems of peer disagreement. As the above quote makes clear, Elga is interested in situations in which one *believes* that one’s interlocutor to be one’s ideal peer. Enoch, elsewhere, is explicit that he takes what I call later in this discussion *Justificatory Peerhood* to be the relevant notion (see Enoch 2010: 476).

might have about those facts. Given this, we might note, we do not need to specify *what kind* of epistemic advantages disputants might have. As Elga makes the point:

In defense of my use, suppose that you think that conditional on the two of you disagreeing about a claim, your friend is more likely than you to be mistaken. Then however intelligent, perspicacious, honest, thorough, well-informed, and unbiased you may think your friend is, it would seem odd to count her as an epistemic peer with respect to that claim, at least on that occasion. You think that on the supposition that there is disagreement, she is more likely to get things wrong (Elga 2007: 500).

Whilst, then, there are a variety of factors that determine the relative likelihood that beliefs held by those party to a disagreement will be true, all that matters in respect to their status as Ideal Epistemic Peers is that those probabilities come out as equal. Thus, for example, given all the facts about how they came to hold their respective beliefs and conditional on the disagreement, if Quinn was in any way more likely to correctly judge the bearing of hers and Zaius's shared evidence, Quinn and Zaius would not be ideal peers. Crucially, this would be so, even if neither Quinn nor Zaius had reason to believe that Quinn had any such advantage. Likewise, if Quinn has evidence that justifies her in believing that she does have the relative epistemic advantage in respect to the disputed propositions, they would still be peers in this idealised sense if that evidence is misleading.

Echoing Elga's point above, I think this way of conceiving the peerhood relation is a natural starting point for discussion. After all, what those party to the disagreement are interested in is determining which, if any, of the disputed propositions are true, or likely to be true. And knowing the antecedent probabilities that any beliefs held by the disputants will be true seems like a fair way of doing that. If those party to a disagreement somehow knew that they were ideal peers, on the other hand, that route would not be available – and so they would clearly face the puzzle described in the previous section (for example, the puzzle of what to think about the restaurant bill).

As discussed in the previous chapter, however, questions about how we ought to respond to disagreement concern what we are justified in believing in the face of disagreement. And, as such, the normative significance of disagreement cannot be determined *solely* by facts to which those party to disagreement do not have direct access. The facts that determine whether disputants are ideal peers, though, are facts to which the disputants do not have direct access. Thus, even if it is so that disputants are ideal peers, that alone cannot determine how they ought to respond to disagreement.

The possibility that disputants be ideal peers whilst being justified in believing that they are not, well illustrates this point. Equally significant is the parallel observation – that disputants might, in principle, be justified in believing that they are ideal peers even whilst they are not in fact ideal

peers. Thus, not only will it not be sufficient for it to be so that disputants ought to respond to their disagreement as if they ideal peers that they are ideal peers, neither is it necessary that they are ideal peers. This suggests a second conception of the peerhood relationship – framed not in terms of the facts about disputants relative standing vis-à-vis the disputed propositions as they are, but in terms of what disputants are justified in believing those facts are. In that light, let's say that disputants are *Justificatory Peers* when the following condition is met:

*Justificatory Peers*

If,  $S_1$  and  $S_2$  disagree and *independent of the substance of their disagreement*  $S_1$  is justified in believing that  $S_2$  is her Ideal Peer, then,  $S_2$  is  $S_1$ 's *Justificatory Peer*

So defined, and reflecting the earlier points, it is neither necessary nor sufficient that disputants be ideal peers for them to be justificatory peers, nor for such cases to create a dilemma for those involved. Since the appropriate response to disagreement is a matter of what disputants are justified in believing and that is determined by an agent's mental states and intellectual/cognitive make-up, if disputants *are* justified in believing that they are ideal peers they ought to respond to that disagreement as if they *are* ideal peers, whatever their relative epistemic standing actually is. Philosophically, then, disagreements between justificatory peers appear to be a methodologically suitable starting point from which to develop an account of disagreement. The problem with a focus on justificatory peers, however, is not so much philosophical, as practical. In short, the condition that disputants be justified in believing they are ideal peers is one that is likely to be met in the real-world rarely at best. Thus, there is some concern that an account of disagreement grounded in analysis of cases of disagreement between justificatory peers will not well generalize to the kinds of disagreement we encounter in our everyday lives. Or, to put it in the terms used in the Introduction, any theory of disagreement based only on consideration of cases of *Justificatory Peer Disagreement* will run into problems with *External Validity*.

The important point here is that the definition of justificatory peers is derivative of that of ideal peers. Thus, since those party to disagreement will only be ideal peers if they are epistemic *equals* in respect to that disagreement, they will only be justificatory peers if they are justified in believing that they are epistemic equals in respect to that disagreement. However, as noted before, we do not have direct access to facts about our relative epistemic standing in respect to propositions over which we disagree. As such, and in contrast to ideal peerhood, the status of those party to a disagreement as justificatory peers depends upon the information that they have about the context of the disagreement. Given this, it becomes important to consider the different dimensions of

evaluation on which we might judge that those we disagree with are our epistemic equals. As we shall see, there are problems on all fronts.

In general, there are two kinds of information by which disputants might judge each other's relative epistemic standing vis-à-vis the disputed propositions:

1. Information about the evidence on which each party has based their beliefs (Evidence)
2. Information about how competently each party has assessed their evidence, or otherwise come to hold the relevant beliefs (Evidence of Competence)

Of course, all sorts of information can provide us with evidence of how competently another person has formed their beliefs. Broadly speaking, though, these can be split into two further categories:

- 2.a. Information about each party's general competence in the relevant domain of inquiry (General Evidence of Competence)
- 2.b. Information about features of the specific context that improve or worsen the various party's performance in that context (Contextual Evidence of Competence)

Crucially, for those party to a disagreement to qualify as justificatory peers, the information that they have about the context of disagreement will only afford them reason to believe that they are ideal peers if (i) it is sufficient evidence to justify believing that they are equals on *each* dimension or (ii) it is sufficient evidence for them to determine that any specific advantages/disadvantages disputants might have along those dimensions balance out overall. I take it that establishing that one's interlocutors are one's epistemic equals when there is evidence of significant disparity along particular dimensions of evaluation will be yet more difficult than when it does not. For that reason, I shall focus only upon the difficulties with establishing equality on each dimension. If establishing equality in this way is prohibitively difficult, then, doing so in the latter way will only be yet more so.

With that said, let us consider the four dimensions of evaluation in turn. First, the demand that disputants have reason to believe that their evidence is equally good.

One way that disputants might have reason to believe that they have *equally good* evidence is if they have reason to believe that they have the *same* evidence. On the face of it, Diners-1 and Doctors-1 would appear to be good examples of disagreement over the same-evidence. One important condition on same-evidence disagreements between justificatory peers, however, is that disputants will typically only be justified in believing that they have the same-evidence after, what Richard

Feldman calls, *full disclosure*. Full disclosure, as Feldman explains, is the stage of in which disputants: ‘have thoroughly discussed the issues. They know each other’s reasons and arguments, and that the other person has come to a competing conclusion after examining the same information’ (Feldman 2006: 220). Disputants who are otherwise justified in believing that they are epistemic equals but who are not at this stage, however, ought not to believe that they have the same evidence. Why? Because, typically, when two equally competent people disagree, we would expect them to have based their opinions on different bodies of evidence. And so, unless that can be ruled out, I would suggest, disputants will not be justified in believing that they have the same evidence. Same-evidence disagreements between justificatory peers, thus, should be understood as disagreements that persist after full-disclosure.

Given that same-evidence disagreement between justificatory peers require full disclosure, however, it follows that such disagreement will only be possible if the disputants’ evidence consists solely of, as Nathan King puts it: ‘the sort of thing that is discursive and shareable through articulation, and is such that it can in principle provide one with a dialectical advantage over one’s dissenters’ (King 2012: 254) King refers to this as a *dialectical conception of evidence*. We can describe considerations that count as evidence in the context of disagreement on the dialectical conception as being considerations of a sort such that, where P is the potential evidence to be shared:

On realising that S<sub>2</sub> based their belief Q on P, S<sub>1</sub> could, in principle, judge the bearing of P on Q without additional information about S<sub>2</sub>’s reliability, or somehow relying upon S<sub>2</sub>’s being reliable.

On this conception, considerations such as intuitions, seemings, perceptual experiences, and so on, would not count as evidence.<sup>13</sup> For instance, I could communicate to you that, when I was in the living room ‘I had the perceptual experience of a white table’. But you could not treat that information as evidence that there is a white table without having some reason to believe that my perceptual experiences are reliable, or else simply ‘trusting’ that my perceptual experiences are reliable. Put aside the question of whether this is a good conception of evidence. Since all evidence on a dialectical conception can in principle be shared, this conception allows that disputants could

---

<sup>13</sup> As it happens, the way that justification is conceived in this thesis suggests a more ecumenical conception of evidence than this – i.e. my perceptual experience that Q can count as evidence for you that Q, just if you have access to an argument from ‘S had perceptual experience Q’ to the conclusion that Q. For sure, I could not base my beliefs directly upon your perceptual experiences and in that sense your perceptual experiences could not literally be my evidence (King 2012: 257), but I could base my beliefs upon information I have about your perceptual experiences.

fully disclose to each other what evidence they considered. Thus, they could come to be justified in believing that they have the same evidence. If, in such a situation, the disagreement persists, then, the disputants will be justified in believing that they disagree despite having equally good evidence.

Being justified in believing that a person you disagree with has the same-evidence as you, however, is not sufficient justification for believing that they are your ideal peers. Thus, being in this position would not be sufficient for disputants to be justificatory peers. Whether we classify them as *evidence*, however, intuitions, seemings, perceptual experiences, or any other ‘evidence-like’ considerations still play a role in determining how reliably we hold true beliefs. And, indeed, if we do not classify these as being a part of an agent’s evidence, it seems natural to group them under the range of considerations that determine how competently an agent assesses the evidence that they do have. E.g. If my rational insight into what my evidence supports is reliably off-base, then, I am less likely to competently assess the bearing of any evidence I have. Restricting the conception of evidence to exclude these types of consideration may allow us to maintain that disputants can have literally the same-evidence, then, but only by making the other dimensions of evaluation more significant when it comes to the disputants’ status as justificatory peers.

So much for ‘same-evidence’ disagreements. What about cases where disputants do not share their evidence?

As noted, it will typically be possible for disputants to be justified in believing that they have the same-evidence only if full-disclosure is possible. In contrast, where, for theoretical or practical reasons, full-disclosure is not possible, I have suggested that it will typically be reasonable to believe that someone that you do not enjoy some other clear advantage over, yet disagree with, has different evidence to you. The second way in which disputants might be justified in believing they are evidential equals, then, is if they have reason to believe that they have different but equally good evidence. To illustrate the possibility of ‘different-evidence’ peer disagreement, consider the following example:

#### SPIES

Agent-M is an analyst working for a US intelligence agency. Agent-M has collected a large body of data pertaining to the interference of a foreign agent G in the recent US election, on which basis she has concluded that ‘G interfered’. Agent-M liaises with Agent-S, a British analyst investigating the same issue, who informs Agent-M that he has concluded that ‘G did not interfere’. Agent-M and Agent-S know each other both to be exceptionally



competent analysts and know that they can trust each other. However, both Agent-M's and Agent-S's evidence is constituted by information that is classified by their respective governments. Though they might like to, neither are in a position to share all of their evidence with the other. As it happens, Agent-M's belief that 'G interfered' is justified by her evidence.

In this case, it seems reasonable for Agent-M to suppose that Agent-S has different evidence. What is more, since Agent-M is an exceptionally competent and trustworthy analyst, it seems reasonable for Agent-M to believe that B has evidence that justifies believing 'G did not interfere'. However, given the practical features of the case, it is not possible for Agent-M and B Agent-S to fully disclose to each other the evidence that they have. Thus, whilst Agent-M can draw conclusions about the bearing of Agent-S's evidence, she does not have access to the specific content of that evidence. As a result of this, a parallel puzzle arises as in the same-evidence cases. Should Agent-M continue to believe that 'G interfered' – even though she is aware of evidence that *may* defeat the support she has for that? Or should Agent-M somehow revise her beliefs – even though the only evidence to which she has direct access supports her belief?<sup>14</sup>

The key point here, and one which is general of different-evidence cases, is that what Agent-M is justified in believing about Agent-S's evidence depends entirely upon what she is justified in believing about Agent-S's competence. Just as restricting the conception of evidence to allow 'same-evidence' cases places greater significance upon the competence dimension of evaluation when it comes to assessing whether disputants are justificatory peers; so, anyway, it is the competence dimension that has primary significance in different-evidence cases.

When it comes to determining whether someone is your ideal peer – and so whether you and they are justificatory peers – then, much of the action and perhaps more than we might have first thought is on the competence dimension. How likely, then, is it that those party to a disagreement could be justified in believing that they are equals on this dimension? I would suggest not at all likely. Or, at least not if the requirement is that disputants be justified in believing that no party to disagreement has *any advantage at all*.

The reason for thinking this is simply that the range of factors that can affect how competently a person assesses their evidence, or otherwise comes to believe something or other, is so wide as for exact parity in a particular case to be extremely unlikely. Quinn and Zaius might well be *similarly*

---

<sup>14</sup> For the most part this is treated as less of a puzzle in the literature than the puzzle in the same-evidence cases. However, I shall suggest in Chapter 4 that these two kinds of case are more alike than they first may seem.

competent in general ways. Indeed, given the details of their educational backgrounds, career positions, and track-record in practice, it seems quite reasonable to suppose that they are similarly competent in general ways. But, any number of factors are likely to prevent them from being *equally* competent. Perhaps Zaius is slightly more careful when it comes to checking all of the details of their patient histories, perhaps Quinn is slightly better at the purely probabilistic aspects of medical reasoning, or maybe she is better at the abductive side of medical reasoning – just a little more sensitive to the particular and peculiar features of individual cases than Zaius. And so on. The problem here is not that Quinn and Zaius could not have exactly the same degree of competence when it comes to the relevant domain of inquiry (thought that seems unlikely) – but rather that, given the range of ways in which they *could* differ in competence, it is hard to see how they could have evidence that justifies believing that neither of them have even the slightest advantage on this dimension.

Even if disputants *did* have evidence that none was even slightly more competent than each other generally, however, they would still not be justificatory peers on the competence dimension unless they were also justified in believing that none performed any better or worse to their abilities in the case at hand. Jack and Jill might be generally as reliable as one another when it comes to simple mathematics, for instance, but that doesn't count for much if Jill has reason to believe that Jack has gotten completely blathered on tequila over dinner. It is important to keep in mind, then, that what I have called 'general evidence of competence' is *only* significant in so far as it is evidence of the disputant's competence *in the case at hand*. After all, those party to disagreement are interested in who is *right* – not who is more reliable. If it is difficult to see what evidence disputants could have to justify belief that none are in the slightest bit more competent generally, however, it is even more difficult to see how they could be justified in believing that they all performed just as well to their abilities in the particular context. Simply consider the sheer number of factors that can impair or enhance one's performance in a given moment, including: tiredness, what food you last ate, when you last ate, coffee, alcohol, headaches, circadian rhythms, sunlight, and any number of other factors. Or consider the number of factors that might affect the *care* with which one considers a question and the evidence bearing on that question – lack of interest, enthusiasm, unrelated joys and traumas, related joys and traumas, the need to pick the kids up from school, or to put out the rubbish... and so on. To be justified in believing that no party has even the slightest advantage on this dimension, one would presumably need to be justified in believing that no relevant factors of this sort were at play. But, given how many such factors are relevant, I'm not sure that we could ever be justified in believing that.

Common to all of the problems I am pointing to, is that they follow on the assumption that disputants' status as justificatory peers depends upon their being able to determine precise probabilities as to how likely each is to be right. Precise probabilistic expressions of the reliability of epistemic agents may be a useful theoretical tool in epistemology, however, but they are a convenient fiction. In the real world, we do not and typically cannot determine each other's reliability (especially in a particular context) with any degree of precision. Perhaps, then, we should simply loosen the conditions on peerhood, such that justificatory peers need only be justified in believing that they are 'close to being ideal peers.' So, for instance, Lackey suggests disputants might be still face the familiar puzzles of peer disagreement if:

(1) [they] are aware that they hold differing doxastic attitudes, and (2) prior to recognizing that this is so, [they] take themselves to be roughly epistemic peers with respect to this question (Lackey 2010: 305).

We might refer to this looser notion of justificatory peerhood as *approximate peerhood*. Approximate peers need not be justified in believing that they are epistemic equals – only that they are approximately so.<sup>15</sup>

Intuitively, the epistemic conditions on approximate peerhood will be significantly less demanding, and so more significantly more likely to be realised, than those on justificatory peerhood. For this to be so, however, we need to be careful in how we understand what it means to say that disputants are *approximately* equals. One way we might parse this is to retain the focus upon relative probabilities from the *ideal* and *justificatory* peer relationships. Thus, we would say that approximate peerhood requires that disputants are justified in believing that the antecedent probabilities that each party is right are *approximately* the same. If we think of approximate peerhood in this way, however, it is hard to see how we could avoid drawing upon the same convenient fiction of how we reason as does the notion of justificatory peerhood. For, however we might specify what difference in supposed antecedent probabilities might still count as 'approximately the same' – any difference that falls outside of that range will prevent disputants from being approximate peers (just as any difference at all prevents them from being justificatory peers in the more ideal sense). But if disputants are to establish that the relative likelihoods that their respective beliefs are true fall within the threshold of different – and not, say, just slightly outside of that – they will need to be just as precise with those probabilities as if they were to establish that none of them is just slightly more likely to be correct than the other. Just as we might say that justificatory peerhood as

---

<sup>15</sup> Lackey refers to this kind of situation as 'Ordinary Disagreement' – I should be clear that this is not how I use that term. As I use it, whether a disagreement is 'ordinary' is a matter of the content of the disagreement; for Lackey it is a matter of the disputants' relative epistemic standing.

a methodological starting point has problems with external validity – so approximate peerhood understood this way would too.

More plausibly, we might think of the relationship of approximate peerhood as a matter of having evidence that there are only slight *material differences* in the disputants' respective epistemic positions. So, for instance, if Jill had evidence that Jack had drunk more wine than her at the dinner table, she would not need to calculate the difference that his consumption of wine made to the comparative probabilities that each was right to determine whether or not he was her approximate peer. Instead, it would be enough for Jill to consider how much more wine he had consumed than her. If Jack had consumed *a lot* more, then she should not think that they are approximate peers. If Jack had consumed only *a little* more than Jill and was not showing any signs of being significantly more inebriated, depending on her other evidence of the context of the disagreement, Jill would be justified in thinking that she and Jack are approximate peers. Approximate peerhood understood this way would seem more easily realised than understood probabilistically. The problem with this from a methodological perspective, however, is that even slight material differences can have significant epistemic consequences. Suppose that Quinn and Zaius are both careful in checking all of the details of patient histories, but Zaius was just slightly less careful this time around. Quite conceivably, Zaius' comparative lack of care may have led to him missing some subtle aspect of the patient's history that the more careful Quinn had not missed. Even a subtle point about the patient's history, however, could have significant consequences when it comes to determining which treatment option is best. Evidence of a slight difference in care taken when assessing the evidence, thus, might well afford both Quinn and Zaius a compelling reason to favour Quinn's opinion (at least until they can recheck the patient history) – even whilst their evidence would be such that they could justifiably believe they were somewhat 'close to being ideal peers'.

Thus, approximate peerhood understood this way is not subject to the worries about external validity I have suggested we might have about justificatory peerhood. Yet, it will have problems with what we could call *internal validity*. For, if there can be slight material differences in the epistemic standing of approximate peers vis-à-vis the substance of their disagreement – then it is quite possible that disputants could be approximate peers whilst still having sufficient reason to favour one side or other in the disagreement. If that is the case, we could not be sure that any conclusions we draw about how disputants ought to respond to those cases are best explained by the significance of the disagreement itself and not those other material differences.

The problems with the concepts of justificatory and approximate peerhood, we might observe, come around because each requires satisfaction of a positive epistemic condition on peerhood.

Whilst, however, any methodologically interesting conception of peerhood has to include *some* epistemic condition, it need not be a positive condition. Indeed, in at least two of the cases we have considered so far, the problems with determining how to respond to the disagreement arise despite the disputants not being in a position to determine that they are peers in the ideal sense. In SPIES, for instance, the fact that Agent-M has no access to Agent-S's evidence makes it difficult for Agent-M to ascertain whether or not Agent-S's evidence is literally as good as her evidence. Indeed, it is probably safe to assume that one of them will have better evidence than the other, even if separately each body of evidence is sufficient to justify their respective beliefs. Nonetheless, even knowing that one body of evidence is likely to be better than the other will not help Agent-M determine how to respond to the disagreement unless she also has an idea *which* of those bodies of evidence is better. Similarly, it is probably safe for Jill to assume that if she wasn't somehow impaired when calculating the restaurant bill, Jack probably was. After all, it is a relatively simple calculation of the kind that both have been reliable on in the past and so they are not particularly likely to disagree unless something is up with one of them. Nonetheless, given that she has no evidence of impairment other than the fact that they disagree, Jill cannot conclude on that basis alone that it was specifically Jack who was impaired, or indeed that it was she. And that being so she still has no way of resolving the disagreement that would not require her to rely on her original – and contested – reasoning to do so. The point here, then, is that the 'puzzle of disagreement' is not generated so much by the evidence that the disputants have – but rather by evidence they lack – i.e. evidence that justifies believing a *specific* party has the epistemic advantage. So long as disputants have no evidence of this kind, or at least none that is independent of the disagreement, then, they will face a puzzle – even if they are otherwise justified in believing that they are not peers in the ideal sense.

As we might put it, then, what is significant is not that disputants are justified in believing they actually are peers, indeed they may rarely be, but that the context of the disagreement is such that – justificatorily speaking – they are *effectively* peers. Along these lines, then, let us introduce the concept of an *Effective Peer*, defined as follows:

#### *Effective Peers*

If,  $S_1$  and  $S_2$  disagree and *independent of the substance of their disagreement*  $S_1$  is not justified in believing specifically that  $S_1$  or  $S_2$  holds the relative epistemic advantage vis-à-vis the disputed propositions, then,  $S_2$  is  $S_1$ 's *Effective Peer*.

Given the connection I made in the Introduction between reasons and justification, this definition is equivalent to the following:

If,  $S_1$  and  $S_2$  disagree and  $S_1$  has no undefeated reason *independent of the substance of the disagreement* to believe specifically that  $S_1$  or  $S_2$  holds the relative epistemic advantage vis-à-vis the disputed propositions, then,  $S_2$  is  $S_1$ 's *Effective Peer*.<sup>16</sup>

There is an important caveat to this definition. Though the conditions of effective peerhood so defined are negative, for the sake of discussion we can take these to entail that  $S_1$  is justified in believing that  $S_2$  is generally competent in the relevant domains.<sup>17</sup> This is because if  $S_1$  can deny  $S_2$ 's competence, either  $S_1$  will have independent reason to believe that she has the relative advantage vis-a-vis the disagreement *or*  $S_1$  should deny that she is competent. In the latter case the negative conditions of effective peerhood will be satisfied, but to coin a phrase, disagreements between incompetent peers are not the disagreements we are looking for. In light of this caveat, we can stipulate ahead of time that all cases of ordinary disagreement discussed in this thesis are cases in which, independently of the substance of the disagreement, disputants would be justified in believing the other parties to that disagreement are generally competent in the relevant domains.

As per the comments above, disagreements between effective peers generate the familiar puzzles in respect to how disputants ought to respond to the realisations of disagreement. For that reason, a focus on disagreement between effective peers is internally valid in a way that we saw a focus on those between the material differences interpretation of approximate peers is not. At the same time, the conditions of effective peerhood are relatively undemanding. For certain, most disagreements that we encounter will not satisfy these conditions, but I see no reason to think that the kinds of disagreement in which those conditions are met would be particularly uncommon. Hence, I would suggest we need not have the worries about generalising from cases of disagreement between justificatory peers when that is understood in the strict sense. Indeed, and on this same point, I would suggest that very many of the disagreements that motivate the epistemology of disagreement – i.e. disagreements about science, medicine, philosophy, and so on – are good candidates for *effective peerhood*. A hypothetical case like Diners-1, for instance, whilst as likely as any to be a case of ideal peer disagreement, might strike us as slightly artificial, just because

---

<sup>16</sup> Of course, if one were justified in believing no party to a disagreement had the advantage, one would also lack reason to believe that any specific party had the advantage – so the class of disagreements between effective peers includes any cases of disagreement between justificatory peers however rare they might be.

<sup>17</sup> Or, at least that  $S_1$  could not justifiably deny that  $S_2$  is competent independently of the substance of the disagreement. I won't dwell on the point, but it might help to keep this caveat in mind when we come to discuss deep disagreement.

we would usually expect there to be clear evidence of error one way or the other when people disagree on such a simple issue. On the other hand, as a problem increases in complexity and difficulty, requiring the assessment of large and complex bodies of evidence and high levels of expertise to do so, the chances that differences in access to that evidence, general competence, or situational factors will lead to error on the part of one party or other would seem to increase drastically. Yet, at the same time, the sheer range of possible sources of error inevitably makes it more difficult to identify any particular factor as making the difference. Just as we might expect those party to the kinds of disagreement that motivate discussion not to be peers in the ideal sense, then, I would argue that there is a good chance that they will often be effective peers. Not only, does the concept of effective peerhood better suit our methodological purposes, so too, I think, it better captures the intuitive appeal of the idea of epistemic peerhood to start with.

With this conception of peerhood in mind, then, we can better specify the question that lies at the heart of the epistemology of disagreement. Simply put:

How ought one to respond to disagreement with an effective peer or peers?

In so far as we are currently concerned with cases of ordinary disagreement, call this ‘The Problem of Ordinary Disagreement’. In the second half of this chapter, I shall suggest that there is a second and parallel problem that an epistemological theory of disagreement must address: ‘The Normative Problem of Deep Disagreement’.

## 1.2 The Normative Problem of Deep Disagreement

In the previous section we saw that ‘The Problem of Ordinary Disagreement’ arises in cases of disagreement between effective peers. Effective peerhood does not require that disputants be justified in believing that no party to the disagreement has any dispute-independent advantage when it comes to the disputed propositions. Indeed, disputants can be effective peers even whilst they would be justified in believing that one party has a quite significant advantage, just so long as they are unable to independently determine which specific party that is. In this section, I shall argue that recognising that the puzzles of disagreement arise in case of disagreement with an effective peer also allows us to recognise a second problem that a normative theory of disagreement ought to address. That is the problem of how to respond to the realisation of disagreement when the other parties to that disagreement are one's effective peers and that disagreement is a *deep disagreement*. Deep disagreements we can characterize as disagreements that are explained by the

disputants following conflicting epistemic norms. As we shall see, there is some pull to the idea that those party to a deep disagreement cannot be peers in the ideal sense – and so also cannot be justificatory peers. They can, however, be effective peers. Thus, if the starting point for a satisfactory normative theory of disagreement is to consider cases of disagreement between effective peers – it seems pursuant on us to also consider the normative implications of deep disagreement.

### 1.2.1 What is Deep Disagreement?

Consider the following case:

#### POLITICIANS

Fax and Prudence are politicians belonging to rival political parties. Recently, a glut of polls from reputable pollsters have indicated that the public have an extremely low level of trust in the political class. Both Fax and Prudence have sincerely professed that they think a crucial aspect of politics is to act so as to engender just the kind of public trust the polls refer to. Additionally, both Fax and Prudence recognise that, were they not personally trusted by the public, engendering such trust would require considerable effort and expenditure on their behalf. Shortly after the release of yet another poll indicating the dearth of public trust, Fax asks Prudence what she thinks about the public-trust problem. Prudence replies ‘What problem?’ Surprised, Fax asks Prudence if Prudence has seen the recent polls. Surely, Fax says, given the poll results ‘It is quite clear that the public do not trust either of us.’ Prudence responds: ‘Of course I’ve seen the polls and they definitely spell trouble for you. Fortunately for me, the public trust me implicitly,’ before adding, with no apparent irony or insincerity, ‘I mean, if they didn’t, I would have to change the whole way I go about doing politics!’

As it happens, Fax based her beliefs about who the public trust on the polls alone; Prudence treats both polls and practical considerations as relevant to determining who the public do and don’t trust. Since Prudence would incur a greater cost were she to act as if the public do not trust her, Prudence determined that the poll results must not be evidence of how the public feel about Prudence. What is more, the difference in how the pair came to their beliefs in this case represents a quite general feature of their epistemic practices – Prudence sees practical considerations as generally relevant to determining what is true, Fax does not.



In this case there is a disagreement at the surface level about the propositions ‘the public do not trust Prudence’ and so also ‘the public do not trust either of us’. At this level the disagreement resembles a case of ordinary disagreement’ I.e. the disputants hold conflicting beliefs toward some non-normative factual proposition or set thereof. Unlike ordinary disagreement, however, the substance of the disagreement is not confined to the level of belief. There is also an underlying difference in terms of the kinds of consideration that each recognises as reasons to believe as they do. Thus, we might say, Fax distinguishes between the ‘epistemic’ and ‘practical reasons’. Prudence, on the other hand, does not make such a distinction. For all intents and purposes, Prudence treats the ‘practical reasons’ for belief as ‘epistemic reasons’ for belief. It is this underlying difference that explains how the pair have come to disagree.

Disagreements like this have come to be known in the literature as *deep disagreements*. Typical examples of deep disagreement include: disagreements between practitioners of alternative and conventional medicine (Kappel 2018); those who follow scientific principles of inquiry and those who follow the authority of religious texts (Lynch, 2010; also discussed in Kappel 2012; Matheson 2018; Pritchard 2018); and disagreements between people on opposite sides of the political spectrum (Pritchard 2018).

Deep disagreements have been characterized as arising when disputants disagree over ‘fundamental epistemic principles’ (Kappel 2012; Lynch 2010; Matheson 2018), and—influenced by remarks in Wittgenstein’s *On Certainty* – when disputants are committed to different *hinge* or *framework propositions* (Fogelin 1985; Pritchard 2018). Let’s call the first approach the ‘Principle’ model and the second the ‘Wittgensteinian’ model.<sup>18</sup> Before moving on to discuss the problems that the possibility of deep disagreement poses it will be worth considering the differences and points of contact between these two models. By doing so, we might best identify the features of deep disagreement most relevant to its distinctiveness in respect to questions of epistemic normativity.

Let’s start with the Wittgensteinian model. As noted, discussion of the place of hinge commitments in epistemology is inspired by comments from Wittgenstein’s *On Certainty* such as:

---

<sup>18</sup> The terms hinge commitment and framework proposition are both used in the literature to refer to the same strand of thought in Wittgenstein’s *On Certainty* and can be treated as synonymous. Since hinge commitment is the more prevalent term in the contemporary literature, I shall follow that trend in my own usage. (Wittgenstein himself does not explicitly refer to either ‘hinge commitments’ or ‘framework propositions’ – but, as is his wont, employs the terms ‘hinge’ and ‘framework’ at various point as metaphors to elucidate his wider discussion.)

That is to say, the questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like *hinges* on which those turn. (Wittgenstein 2013: §343. Italics added).

The mathematical proposition has, as it were officially been given the stamp of incontestability. I.e.: “Dispute about other things; this is immovable—it is a *hinge* on which your dispute can turn (Wittgenstein 2013: §655. Italics added).

There is ongoing debate about the meaning of passages such as these, as well as what conception of a hinge commitment is relevant to epistemology irrespective of the exegetical questions. Still, Ranalli succinctly explains the general idea here:

As a rough and ready characterization [...] hinge commitments are the background presuppositions of our world views and general areas of inquiry, such as physics, history, or geology. In particular, hinge commitments are given a certain epistemic role within our world views and inquiries (Ranalli 2018b: 4).

Standard examples of hinge commitments include propositions such as ‘there is an external world’; ‘I exist’; ‘my sense faculties are reliable’. Amongst many other examples, Wittgenstein himself includes ‘my hands don’t disappear when I am not paying attention to them’ (Wittgenstein 2013: §153), ‘no one ever was on the moon or *could* come there’ (Wittgenstein 2013: §171), and ‘the earth [is] a ball floating free in space and not altering essentially in a hundred years’ (Wittgenstein 2013: §146). As these examples make clear, hinge commitments need not have any epistemic or normative content to class as such. Instead, the status and significance of any hinge is determined by the nature of the agent’s commitment to the proposition in question and the role that that commitment plays in their epistemic practices. What that entails precisely will depend upon one’s preferred account of hinge commitments (for discussion, see Ranalli 2018a; Coliva 2016).<sup>19</sup> Nonetheless, in line with the Ranalli quote above, we can roughly characterize this role as the agent’s presumption of the certainty of the relevant hinge propositions across their epistemic practices. Moreover, it is common to the various accounts that hinge commitments are not arrived at by way of rational inquiry as we might expect of an agent’s beliefs, but rather form its underlying framework. Thus, to say that S has ‘God exists’ amongst her hinge commitments, is to say that, though she did not arrive at that position by way of rational inquiry, S holds it as certain that ‘God exists’ across her epistemic affairs and rejects whatever she feels to be inconsistent with that certainty (at least in ordinary circumstances). Importantly, hinge commitments have this role

---

<sup>19</sup> As Ranalli (2018b: 16) notes it is also a matter of debate as to whether there are such things as hinge commitments. For the purposes of this discussion I shall presume that we do have hinge commitments – the question is then, given that we have hinge commitments, how these can help us better understand the normative significance, or not, of deep disagreement.

without necessarily corresponding to any explicit attitudes held by the agent toward the relevant hinge propositions. As Pritchard explains:

[O]ur hinge commitments are ordinarily tacit, in that they are rarely made explicit in our normal rational activities, much less do we become aware of them qua hinge commitments. As Wittgenstein puts it, “they lie apart from the route travelled by inquiry” (OC, § 88). Relatedly, our hinge commitments manifest themselves most fundamentally not in what we say but rather in what we do. That is, it is our manner of acting that manifests the certainty that constitutes our hinge commitments (Pritchard 2018: 2).

What I have called the Wittgensteinian model, then, delineates the class of deep disagreements as those involving – whether explicitly or at the level of tacit presupposition – a clash of these hinge commitments. As Pritchard and Fogelin point out, given the nature of hinge commitments it is reasonable to suppose that such disagreements will be inherently intransigent if not, indeed, rationally irresolvable. So Fogelin (who uses the alternative term framework proposition) writes:

What I have called deep disagreements are generated by conflicts between framework propositions. They remain recalcitrant to adjudication because the sources of the disagreement—the framework propositions—are allowed to lie in the background, working at a distance. The way to put the debate on a rational basis is [to] surface these background propositions and then discuss them directly (Fogelin 1985: 8).

And, more pessimistically, Pritchard suggests that:

[S]ince one’s hinge commitments are essentially arationally held, it appears to follow that deep disagreements of this kind cannot be rationally resolved. Indeed, since it is one’s hinge commitments that determine the nature of one’s rational evaluations, we also seem to get the epistemic incommensurability claim noted above by default (Pritchard 2018: 3).<sup>20</sup>

Given this inherent intransigence and the fundamental role that hinge commitments are purported to have in our epistemic practices, deep disagreements so conceived are undoubtedly a theoretically interesting class of disagreement. At the very least they raise hard questions about how, practically speaking, we might best respond to such disagreements so as to avoid unconstructive *aporia* in contexts that require epistemic coordination between different people and groups. What is left somewhat mysterious, however, is how and whether this model picks out a class of disagreement that is normatively distinct from ordinary disagreement. After all, that a disagreement is intransigent is not in itself a matter of normative significance. If an expert finds themselves in disagreement with an abidingly stubborn dilettante, that disagreement will be intransigent, moreover in such a way that any argument that the expert might forward for the view in question

---

<sup>20</sup> Pritchard, to be clear, questions and ultimately rejects the suggestion that deep disagreements are rationally irresolvable. So long as there is enough sufficient ground between disputants, he suggests, disputants can work a change in their commitments through rational discourse, albeit only indirectly (Pritchard 2018: 7). Nonetheless, I take it that Pritchard would endorse the weaker claim that they are inherently intransigent.

the dilettante will reject, but there is nothing especially normatively puzzling about this case. Despite the intransigence, the expert clearly need not give any weight to the dilettante's views. So, if deep disagreements are to be a more interesting class of disagreements than disagreements with the abidingly stubborn, there must be more to them than their inherent intransigency. In that light, I would suggest, deep disagreements must manifest not just the commitment to upholding the certainty of the relevant propositions, they must stem from the disputant's commitment to a certain *epistemic-normative* way of doing things in light of those propositions. Which, neatly enough, brings us to the epistemic principle model.

The *locus classicum* for describing deep disagreements in terms of epistemic principles is offered by Lynch in his 2010 paper 'Epistemic Circularity and Epistemic Incommensurability'. It is worth quoting this at length:

[S]ometimes we move still further up what we might call the epistemic ladder: we begin to disagree over how we *ought* to support our views of the facts, about the sort of evidence that should be admitted, and whose methods more accurately track the truth. When we do that, we are engaged in a truly *epistemic* disagreement: a disagreement over epistemic principles. [...] I'll say [an *epistemic*] disagreement is *deep* when it meets the following conditions:

*Commonality*: The parties to the disagreement share common epistemic goal(s).

*Competition*: If the parties affirm distinct principles with regard to a given domain, those principles (a) pronounce different methods to be the most reliable in a given domain; and (b) these methods are capable of producing incompatible beliefs about that domain.

*Non-arbitration*: There is no further epistemic principle, accepted by both parties, which would settle the disagreement.

*Mutual Circularity*: The epistemic principle(s) in question can be justified only by means of an epistemically circular argument. (Lynch 2010: 264-65).<sup>21</sup>

---

<sup>21</sup> Reflecting points later, I take it that the first of these conditions does not stipulate that disputants must share the same epistemic goals, but that whatever goals that they do have must be 'epistemic goals'. Lynch's illustration of the principle fits with this reading, as he suggests: 'Finally, in some cases, the disputants are simply talking past one another—one is aiming at truth, for example, the other at practical success or some such.' (Lynch 2010: 265) See also the definition of deep disagreement Lynch and Silva offer in a more recent paper, which drops the need for the conditions in the previous definition by focusing on the individual agent's position in respect to the disagreement. Thus, they state:

'(DED) S1 is in a deep epistemic disagreement with S2 just in case: (i) S1 employs an epistemic principle, EP, to arrive at true beliefs in some domain of inquiry, (ii) S2 rejects that epistemic principle as a reliable guide to forming true beliefs in that domain, and (iii) S1 has no further epistemic principle, EP\*, that does

I shall return to each of Lynch's four conditions in due course during the discussion in this chapter and Chapter 6. For the moment, however, it is worth noting that Lynch's concern is with a primarily meta-epistemological question: Should the possibility of disagreements that are irresolvable on account of their metaphysical nature lead us to accept some form of relativism about epistemic justification? These four conditions, and specifically the fourth, thus, can be seen as intended *just so as to* delineate the class of disagreements relevant to that question. Whilst the meta-epistemological implications of deep disagreement are undoubtedly worthy of discussion, however, these issues run somewhat orthogonal to discussion of the normative implications of deep disagreement. For that reason, whilst Lynch's conditions capture much of what is important about deep disagreement – and I shall flag up their importance where relevant – I shall not adhere to them entirely through this thesis.

For the moment, I want to focus at the more basic level on the question of what the Principle model of deep disagreement might add to the Wittgensteinian model.

Previously we saw that hinge commitments can correspond to propositions without normative-epistemic content. In contrast, we might understand epistemic principles, at least as treated in the disagreement literature, as theoretical statements that define or describe the application conditions of some normative epistemic concept – e.g. justification, knowledge, evidence, and so on. So, Lynch and Matheson, respectively, describe the notion: 'By an epistemic principle, I mean a normative principle to the effect that some source or way of forming beliefs has some valuable epistemic status.' (Lynch 2010: 262) and 'An epistemic principle claims that some epistemic property (justification, knowledge, warrant, etc.) obtains whenever some descriptive property obtains.' (Matheson 2018: 3) Since, then, deep disagreements on this model are limited to those concerning, or involving, epistemic principles – and epistemic principles encode the application conditions for normative epistemic properties – it is clear why deep disagreements as the class of disagreements over epistemic principles might present normative issues distinct from ordinary disagreement.

Where the principle model is less clear, is upon the question of what standing those party to deep disagreement have to the respective principles. Or to put it otherwise, we might ask what it means to disagree over epistemic principles. One thought that we might have is that those party to deep

---

not rely on or presuppose the reliability of EP with which he can show that EP is in fact a reliable means of arriving at true beliefs.' (Lynch and Silva 2016: 43)

disagreement simply *believe* the relevant principles to be true. And, indeed, at times those discussing the principle model do lean on this way of describing things. Thus, Matheson suggests that:

Fundamental epistemic principles are (epistemically) justified when they meet their own standards for justification. [...] when one's belief that one's epistemic principle is correct meets the relevant descriptive properties, it is epistemically justified. On this account, I can be justified in believing [the epistemic principle] Seeming when it seems to me that Seeming is true, and that *prima facie* justification is not defeated (Matheson 2018: 4).

Ranalli writes:

[W]hat are the attitudes of the disagreeing parties to the relevant fundamental epistemic principles? In the case of direct deep disagreement, it seems to be belief. So too, we might claim, it is belief in the indirect case, albeit dispositional belief (Ranalli 2018b: 11).

And Lynch:

Let us say that A disagrees with B over some epistemic principle just when A does not believe EP and B does. Thus one disagrees with someone over an EP in this sense when one either disbelieves the EP or withholds belief in it (e.g. when one doubts that it is true). A overtly disagrees with B over some EP just when A explicitly withholds assent from an EP B asserts (Lynch 2010: 265).

Despite quotes such as these, however, I think that it would be uncharitable to view the principles model strictly in terms of belief. Rather we might see such language as a matter of convenience. Why? Because most people do not explicitly believe the kinds of statement that would class as epistemic principles. Nor I would suggest, aside from epistemologists and some theoreticians of other stripes, do most people have the theoretical resources such that we might expect them to be disposed to believe such statements. Thus, were the principle model to require that those party to deep disagreement *believe* conflicting principles, we might expect vanishingly few disagreements to count as deep disagreements – including many of those that are taken to be paradigmatic of the kind. A more charitable reading of the principles model, then, is that those party to deep disagreement do not have to believe the principles in question. Instead, I would suggest, we can lean on the notion of hinge commitments to make good on this explanatory lacuna. I.e. We might describe deep disagreement as disagreements that arise when disputants have amongst their hinge commitments commitments to conflicting epistemic principles *and* it is that underlying conflict that explains the disagreement. On this understanding, those party to a disagreement could be committed to conflicting epistemic principles as hinges, whilst having no explicit attitudes to those principles. What matters is that the agent *act as if* those principles are true, not that they believe that those principles are true. Interestingly, Lynch echoes this thought when he writes:

The underlying issue isn't a matter of what we know or don't know, but of what we should or shouldn't do. It is not even best conceived as a problem about which principles we

‘ought’ to believe (because one might say we ought to believe what we know to be true). After all, the root issue at the heart of an epistemic disagreement—that which makes the dispute an ‘epistemic’ one—is the question of which methods we ought to employ. What we want is a reason for employing one method over another. That’s a practical matter. (Lynch 2010: 274).

By treating the Wittgensteinian and Principle models as complimentary, not conflicting, then, we can accommodate both the idea that deep disagreements are normatively distinct from ordinary disagreement and the idea that such disagreements need not require explicit beliefs. The Principles model captures the first, the Wittgensteinian model the second. What I want to suggest now, is that a natural way of expressing this synthesis of the two models is if we describe deep disagreements in terms of the *epistemic norms* that the disputants follow.

In general, following Littlejohn and Turri (2014) and Boghossian (2008a), we might gloss epistemic norms as rule-like statements of the following form:

*Where  $\Phi$  designates some normative epistemic act such as believing, inferring, trusting, and so on,*

Under condition C, one ought to/ought not to/is permitted to  $\Phi$

Or, imperatives of the form:

If C,  $\Phi!$  / don’t  $\Phi!$

In the most basic sense, to say that someone ‘follows an epistemic norm’ is to say that they somehow regulate the relevant epistemic actions in accordance with what they *perceive* to be the prescriptions of that norm in different circumstances. I shall say more on what this amounts to in Chapter 6. For the moment, however, the important point is that ‘following an epistemic norm’ is intuitively (i) something an agent does, not necessarily a matter of what they believe, and (ii) constitutively connected to what an agent treats as epistemic reasons. The former I take to be self-explanatory (at least for the moment, I return to this in Chapter 6). The latter follows from the suggestion in the Introduction that questions about what we ought to believe are questions about what we have reason to believe. Thus, to say that someone *perceives* the situation they are in as one in which an epistemic norm that they follow prescribes revising their beliefs is just to say that that person *perceives* that situation as one in which they have reason to revise their beliefs.

With this much said, let us say that two or more people are involved in a deep disagreement about Q when they hold conflicting beliefs about Q and that disagreement is explained by their having followed different epistemic norms in coming to hold those beliefs.

This characterization of deep disagreement captures the key aspects of each of the principle and Wittgensteinian models. (Incidentally, it also entails that deep disagreements satisfy Lynch's *competition* condition). Since epistemic norms are closely related to epistemic reasons and justification, similarly to the principle model, it is readily apparent that such disagreements might be normatively significant in a way distinct from ordinary disagreement. Since, following an epistemic norm is a matter of what we do, and not necessarily what we believe, as per the Wittgensteinian model, a characterization in terms of norms captures the thought that the source of deep disagreements is something below the level of belief – and, thus, does not require that disputants hold explicit attitudes about theoretical epistemological issues.

Now that we have some idea of what deep disagreements are, we are well positioned to consider how they represent a problem that is distinct from the problem of ordinary peer disagreement. The key question is whether those party to such disagreements can in some sense be epistemic peers.

If those party to deep disagreement cannot, in principle, be peers in any sense – then, it will always be clear from the justificatory context of deep disagreement which side or other to favour. In which case we need not consider whether deep disagreement has any special normative significance. If those party to deep disagreement can be epistemic peers in some sense, cases of deep disagreement in which it is not clear which side or other to favour may be possible. If so, it will be important to consider whether the fact that a disagreement is deep has any special normative significance over and above the significance of the fact that there is a disagreement in beliefs.

In the final section of this chapter, then, I consider whether or not those party to 'deep disagreements' can in some sense of the term be epistemic peers.

### **1.2.2 Deep Disagreement and Epistemic Peerhood**

Let's return to POLITICIANS. Following the discussion in the previous section we might describe the disagreement as one in which Fax and Prudence disagree about which politicians the public do and do not trust *because* they have come to hold the relevant beliefs by way of following the conflicting norms:

Fax's-norm: If the only considerations that support one's believing Q are practical considerations, then, for the sake of acquiring true beliefs, one ought not to believe Q.



Prudence's-norm: If the only considerations that support one's believing Q are practical considerations, then, for the sake of acquiring true beliefs, it is permissible for one to believe Q.<sup>22</sup>

I take it that, given this difference in norms, Fax and Prudence could not be ideal peers, nor could they be justificatory peers – at least as far as being justificatory peers in respect to the deepness of the disagreement. For the pair to be ideal peers it would have to be so that their norms are equally good. Since the propositional statements of the two norms are inconsistent, however, it seems that this will not be the case. Either one of those norms is true and the other false, or both norms are false. In which case, the pair are either *not* peers in the ideal sense, or they are peers – but only in so far as neither of them are competent agents. Disagreements between incompetent agents are not normatively interesting. Given this reason for thinking the pair cannot be ideal peers, it follows immediately that they cannot be justificatory peers either. Or, at least in the relevant sense.

The relevant sense of justificatory peerhood would be such that disputants are justified simultaneously in believing that they are ideal epistemic peers and that they are party to a deep disagreement. (If disputants are justified in believing they are ideal peers *qua* ordinary disagreement – I take it, they *ought* to respond to the disagreement as if it is an ordinary disagreement). Reflecting the point about ideal peers, however, those party to deep disagreement could only justifiably believe that they are peers in the ideal sense *and* involved in a deep disagreement, if they believe that each of their norms is false. For, given that the norms are, by definition, inconsistent – they could not justifiably believe that their norms are both true. And, if they were justified in believing that one was true and the other not, they would not be justified in believing that they are ideal peers. If the disputants are justified in believing that all of their norms are false, however, they are also justified in believing that none of them are competent to determine the truth of the disputed propositions. In which case, they *clearly* ought to withhold judgement on the truth of those propositions. (Again, disagreements between incompetent peers are not the disagreements we are looking for.)

If those party to a deep disagreement are to be peers in any sense at all, then, it will be if they are effective peers. As it happens, and as I shall explain below, Fax and Prudence *are* effective peers. Thus, it is possible in general that those party to deep disagreement can be effective peers. There

---

<sup>22</sup> As per the comments about Lynch's *commonality* condition at fn. 18 it is important to include the 'for the sake of' caveat in these norms. The point of the example is not that Prudence forms beliefs *in spite* of what Prudence perceives to be 'epistemic reasons for belief' – but rather that Prudence does not distinguish between epistemic and practical reasons. For all intents and purposes, Prudence *perceives* practical reasons as epistemic reasons. (As it happens, I don't think that this is an entirely uncommon epistemic position for people to take, but that is a different discussion.)

are two important considerations that make this possible. First, as defined earlier, one's interlocutors are one's effective peers when one does not have reason to believe that any specific party to the disagreement has the relative epistemic advantage, where that reason is *independent of the substance of the disagreement*. Second, since the norms that one follows are constitutive of what one treats as reasons for belief, the norms that those party to a deep disagreement follow also define what counts as a reason for belief that is independent of the substance of the disagreement. That being so, one will have a dispute-independent reason to believe that some party or other has the relative epistemic advantage in the case of deep disagreement if and only if one has a reason to believe as such that would, either: (i) count as an epistemic reason on *any* of the norms in play in the disagreement, or (ii) count as an epistemic reason on some *further* norm that both parties share and accept has authority in the situation. If neither of those things is the case, however, then those party to the disagreement will be *effective peers*.

Again, can illustrate this by way of POLITICIANS.

Fax is Prudence's effective peer if Fax has no reason to believe that she has the epistemic advantage over Prudence that is independent of the substance of the disagreement. As the case goes, the disagreement between Fax and Prudence is explicitly over the question of who the public do and do not trust. Moreover, whilst that disagreement is explained by their following different norms, it is not necessary that they hold any beliefs in respect to those norms. (Indeed, we might think it implausible that they would. They are politicians, not philosophers, after all.) Nonetheless, Fax's and Prudence's respective norms determine the kind of consideration each sees as providing reasons for belief. Thus, though their norms are, so to speak, hidden from view, the difference in norms cannot be separated from the substance of the disagreement. That being so, any reason that Fax might have to believe that she has the epistemic advantage over Prudence will fail to be independent of the substance of the disagreement unless it also counts as a reason to believe the same on the norms that Prudence follows. Now, as the case goes, Prudence as close to admits to Fax that she factored the practical considerations into her assessment of who the public do or do not trust. And, on Fax's norms, that admission would count as reason for Fax to believe that Prudence has made a significant error in her reasoning. However, the appropriateness of taking into account practical considerations in one's epistemic deliberations is precisely the matter on which Fax's and Prudence's norms differ. And, as such, Prudence's admission is not a dispute-independent reason for Fax to believe that she has the relative advantage. Since, then, as described, Prudence's admission is the only evidence Fax has in her favour, it follows that Prudence is Fax's effective peer.

Notice, however, that Prudence is Fax's effective peer only because the brute facts about the norms they follow make her so. For, were this an ordinary disagreement, Prudence *would not be* Fax's effective peer. After all, were this an ordinary disagreement Fax and Prudence would share epistemic norms. Yet, we have already seen that on Fax's norms, the evidence Fax has that Prudence factored the practical considerations into her judgement is evidence of a significant error on the part of Prudence. Similarly, if the pair were to share Prudence's norms – the evidence Fax has that Prudence factored the practical considerations would be evidence for Fax of a significant error on her part – i.e. she failed to take the practical considerations into account in her reasoning. In either case, then, were the pair to share norms, Fax would have independent reason to believe one of the pair had a significant advantage – and so Prudence would not be her effective peer. Though Prudence would not be Fax's effective peer if they share the same norms, then, the fact that they do not – i.e. the fact that it is a deep disagreement – is sufficient to make her so.

This points to an important and more general feature of deep disagreement. For if the other parties to deep disagreement can be one's effective peers merely as a result of the brute facts about the norms they follow, it follows also that one cannot recognise that the other parties are one's effective peers by reflection alone. Instead, to recognise that the other parties are one's effective peers in this kind of case, one will have to acquire additional evidence – either sufficient to establish that any consideration one might suppose to be in one's favour will be disputed, without establishing that the disagreement is deep. Which, in a case like POLITICIANS, would appear to require Fax to subject a vast range of her background beliefs to the scrutiny of Prudence – something that is likely to be practically impossible. Or, one would have to acquire evidence sufficient to establish that the disagreement is deep. Again this may be difficult to achieve in a practical sense. More significantly, though, it would require that one be capable of thinking about epistemic norms – i.e. conceptualising behaviour in terms of them and recognising behaviour as following them — and, as we have already discussed, most of us do not have the theoretical resources at hand to do that.

Given these various points, then, we can recognise three important features of deep disagreement and epistemic peerhood: (i) those party to deep disagreement can be effective peers, (ii) those party to deep disagreement can be effective peers when they would not be, were the disagreement an ordinary disagreement, and (iii) those party to deep disagreement can be effective peers when they are not in a position to recognise that they are effective peers. Given these last two points, I would suggest that we should recognise that effective peerhood in the case of deep disagreement is *metaphysically distinct* from effective peerhood in the case of ordinary disagreement. If what we have

learned from ordinary disagreement is that effective peers cannot simply dismiss each other on realising disagreement, then, we need to consider what this difference in the nature of epistemic peerhood reveals about the normative significance of deep disagreement.

For that reason I take it that any satisfactory normative theory of disagreement should also consider what we might call ‘The Normative Problem of Deep Disagreement’ i.e. How ought one to respond to realising disagreement, when that disagreement is a deep disagreement and those one disagrees with are one’s effective peers.

## Conclusion

In this chapter we have seen that there are two normative problems of disagreement: ‘The Problem of Ordinary Peer Disagreement’ and ‘The Normative Problem of Deep Disagreement’. I have suggested that a satisfactory normative theory of disagreement ought to address both problems. I shall do that in the following chapters. In Chapters 2 and 3, I survey the literature on ordinary peer disagreement and consider the merits of the various extant accounts of that. In Chapters 4 and 5 I offer my own account of ordinary disagreement. In Chapter 6, I extend this account to deep disagreement.

# Chapter 2. The Problem of Ordinary Disagreement: Conciliationism

## Introduction

In the previous chapter, we encountered two problems central to the epistemology of disagreement: ‘The Problem of Ordinary Disagreement’ and ‘The Normative Problem of Deep Disagreement’. In this chapter and the next, I shall look at the various accounts of ordinary disagreement offered in the literature.

In the first section of this chapter, I offer a simple taxonomy of the possible accounts of disagreement one might offer in response to ‘The Problem of Ordinary Disagreement’, distinguishing between: ‘Conciliationism’, ‘Steadfast Views’, and ‘Non-Uniform Views’ of ordinary disagreement. Having introduced this taxonomy, the rest of the chapter considers the case for and against the Conciliationism.

Conciliationism stipulates that, in the face of ordinary disagreement with an effective peer, one ought always to revise those of one’s beliefs that are under dispute. In the second short section of the chapter, I consider two ways in which we might understand the requirement to revise one’s beliefs. In the third and fourth sections of the chapter I discuss and develop what I take to be the two core arguments for Conciliationism: the ‘Truthometer Analogy’ and the ‘Independence’ argument. In the final sections of the chapter, I consider and develop what I consider to be the principal objection to Conciliationism: the problem of ‘Spinelessness’ and the sceptical consequences associated with that.

## 2.1 Framing the debate

‘The Problem of Ordinary Disagreement’ raises the question: how ought one to respond to disagreement with one’s effective peer? Broadly speaking there are two ways in which one might respond in such situations. One might continue to believe what one originally believed, or one might revise one’s beliefs. Those who argue that one ought to revise one’s beliefs in the face of disagreement with an effective peer we can label ‘Conciliationists’. They have included in their number David Christensen (2007, 2011, 2013, 2009), Adam Elga (2007), Richard Feldman

(Feldman 2006, 2007), and Jon Matheson (2015). The view that one is justified in retaining one's original beliefs in the face of disagreement with an effective peer we can label the 'Steadfast View'. Arguments in favour of Steadfast View have been offered by David Enoch (2010) and Karl Schafer (2015)

Both Conciliationist and Steadfast views are best understood as making claims about the default response to finding oneself in disagreement with an effective peer. A third approach to the problem of ordinary peer disagreement is to deny that there is a default response to the problem. Echoing Lackey (2010), we can call views of this kind 'Non-Uniform' views. According to non-uniform views, how one ought to respond to disagreement with an effective peer or peers can differ between cases. In some cases, the appropriate response will be to revise one's beliefs, in others one may be justified in retaining one's original beliefs. Which is the appropriate response in any given case will depend upon the precise details of the non-uniform view in question. Arguments in favour of non-uniform views have been forwarded most notably by Thomas Kelly (2013, 2010) and Jennifer Lackey (2010), as well as by Richard Feldman (2009) and Paul Faulkner (2016).

This rough schema, then, gives us three broad views on 'The Problem of Ordinary Disagreement':

1. Conciliationism
2. The Steadfast view
3. Non-uniform views

As Kelly points out (2013: 35), the debate around ordinary disagreement has tended to treat Conciliationism as 'the view to beat'. In that light, Steadfast and non-uniform views have tended to be framed in terms of arguments against the key aspects of Conciliationism. I will follow that framing device in the following two chapters. In this chapter I look more closely at the conciliationist view and consider two arguments in its favour: the 'Truthometer Analogy' and the 'Independence' argument. I then introduce and develop an objection to Conciliationism that follows directly from the consequences of the view: the 'Spinelessness' objection. In the next chapter, I look at the arguments against Conciliationism that have been forwarded in favour of Steadfast and non-uniform accounts.

## 2.2 The content of Conciliationism

Conciliationism has it that, upon realising that one disagrees with one's effective peer one ought to revise one's beliefs about the disputed propositions. Before engaging with the arguments for

this view, it will be useful to have some picture of what the requirement to revise one's beliefs involves. It will also help to have some illustrative examples at hand. Our two core examples of ordinary peer disagreement from the previous chapter fit the bill nicely. Here is a quick refresher:

Diners-1 – Jack and Jill are members of a party of diners in a restaurant. At the end of the meal the pair volunteer to work out what each diner owes. Jack concludes that each diner's share is \$45; Jill that it is \$43. The correct figure is \$43.

Doctors-1 – Doctors Quinn and Zaius are treating Patient-X for Condition-C. Based on assessment of the relevant clinical and scientific evidence, Quinn believes that treatment option-A would offer the patient the best chance of recovery. Based on assessment of that same body of clinical and scientific evidence, Zaius believes that treatment option-B would be best.

As the two cases were constructed, the disputants in each case are effective peers. That is to say, in neither case does either party have dispute-independent reason to believe that specifically, they or, specifically, their interlocutor has the relative epistemic advantage in respect to the disputed propositions. The conciliationist says that, given their status as effective peers, and whatever they were justified in believing beforehand, both parties in both cases ought to revise their beliefs upon realising the disagreement. Likewise, for any other case of ordinary disagreement between effective peers. There are two models of 'belief' that have been employed in the literature to describe this requirement:

On a 'coarse grained' or 'categorical' model, beliefs and belief revision are modelled on the tripartite schema of 'belief', 'disbelief', and 'suspension of judgement'. On a 'fine-grained' or 'credences' model, talk of belief and belief revision can be parsed in terms of the *confidence* or *credence* an agent has, or ought to have, in the truth of the target propositions. Drawing on the language of subjective probabilities, one's confidence/credence in the truth of a proposition can be described on a scale of 0 to 1. If one's confidence in Q is 1, one is certain that Q is true; if one's confidence is 0, one is sure that Q is false; if one's confidence in Q is 0.5, one is no more confident that Q is true than it is false.

As I will present them in the following two sections, the central arguments for Conciliationism commit its adherents to the specific claim that, upon realising that one disagrees with one's

effective peer, one should give equal weight to one’s own and one’s interlocutor’s beliefs in determining what to believe in the face of the disagreement.<sup>23</sup>

Thus, on the categorical model, Conciliationism says that Jack and Jill each ought to suspend judgement on both propositions ‘each diner owes \$43’ and ‘each diners owes \$45’; and Quinn and Zaius each ought to suspend judgement on both propositions ‘A is best’ and ‘B is best’. On the credence model, Conciliationism is most typically taken to require that the disputants ‘split the difference’ between their original degrees of confidence in each of the disputed proposition.<sup>24</sup> That is, their credence in each proposition after realising the disagreement should equal the average of their credences prior to realising that they disagree. So, if Quinn and Zaius’s original credences were as follows:

	Quinn’s credences	Zaius’s credences
<b>A is best</b>	0.7	0.1
<b>B is best</b>	0.2	0.9
<b>Some other option is best</b>	0.1	0

Then, the ‘split the difference’ rule would require that, on realising the disagreement, Quinn and Zaius both come to have credences:

	Conciliated credences
<b>A is best</b>	0.40
<b>B is best</b>	0.55
<b>Some other option is best</b>	0.05

<sup>23</sup> It is because of this commitment that the view of Conciliationism I refer to has also been labelled at times ‘The Equal Weight View’. The Equal Weight label is first employed by Kelly (2005) in arguments against Conciliationism, but has since been adopted by a number of conciliationists (for instance Elga 2007; Matheson 2015) as well as other critics of the view (for instance Schafer 2015). It is worth noting also that Matheson (2015) distinguishes between weak and strong variants of Conciliationism – where strong Conciliationism corresponds to what I simply call Conciliationism and Weak Conciliationism refers to any view that suggests some weight should be given to the opinions of one’s peers, if not equal weight to one’s own. On this distinction, most views in the literature would count as variants of Conciliationism. For that reason, I prefer the taxonomy offered in 2.1. (for an alternative use of the distinction between strong and weak conciliationism, see Kelly 2014)

<sup>24</sup> Talk of ‘splitting the difference’ appears to come from Christensen (2007), but the basic idea is endorsed by other conciliationists (see Elga 2007; Matheson 2015). It is worth noting that the notion of splitting the difference is typically presented as an informal approximation of how to follow Conciliationism. This approximation will do for present purposes, but for discussion of the problems with the split-the-difference approach see Jehle and Fitelson (2009), and for an attempt to better precisify the demands of Conciliationism, see the ‘The Variable Equal Weight View’ as presented in Rasmussen et al (2018).



In other words, after realising that they disagree, both Quinn and Zaius should be close to agnostic as to which option is better, though ever so slightly favouring B.

Both ‘categorical’ and ‘credence’ models are employed on all sides of the debate. Feldman’s (Feldman 2006, 2007) conciliationist position is articulated in terms of categorical beliefs, as are Wedgwood’s arguments in favour of steadfastness (Wedgwood 2010), and Lackey’s (2010) and Faulkner’s (2016) in favour of non-uniform positions. Conciliationists such as Christensen (2007, 2011, 2009), Elga (2007), and Matheson (2015), on the other hand, have tended to talk in terms of credences, as have critics of Conciliationism arguing in favour of both Steadfast (Enoch 2010; Schafer 2015) and non-uniform views (Kelly 2010).

Ultimately, which is the better model will depend upon how the concepts of belief, credence and confidence relate to one another more generally. Unsurprisingly, there is a substantial body of literature and disagreement on questions concerning just that. So long as these metaphysical questions remain open, however, it seems to me that epistemologists working on disagreement should allow themselves the freedom to employ both models. Depending on the case or point being made, it may be more useful to talk of disputants having reason to reduce or increase their credences in the disputed propositions, or to talk about their having reason to switch from one categorical belief state to another. In some cases, it will make sense to call on both models. In line with this suggestion, for the most part in this thesis I employ the terminology of the categorical model but, where relevant, also lean upon the notion of rational confidence. (Following the worries about attributing overly sophisticated probabilistic reasoning to ordinary epistemic agents in ordinary circumstances that I expressed in the last chapter, however, I shall generally avoid describing cases in terms of precise subjective probabilities – as is typically done by those who employ the credences model.)

So much for the substance of Conciliationism, let’s turn to consider the arguments in favour of the view.

## 2.3 Arguments for Conciliationism

### 2.3.1 The ‘Truthometer’ Analogy

The basic intuitions behind Conciliationism are, I think, quite natural: When one has no prior opinion on an issue, under the right conditions, learning what another person believes about that issue can justify one in believing as they do. Indeed, we are accustomed to treating speakers as authorities, this being our testimonial practice. Given this practice, it seems reasonable to suppose

also that learning what another person believes about an issue can have justificatory significance when one already has an opinion on that issue. After all, the conciliationist might point out, if I find out that you believe Q prior to considering the issue myself, I would quite possibly believe as you do, and quite possibly be justified in doing so. Why think, then, that the discovery of what you believe should cease to have justificatory significance just because that discovery happens at a different point of time in my epistemic practices?

One way that conciliationists have attempted to draw out these intuitions and their purported revisionary consequences is by way of a simple argument from analogy.<sup>25</sup> Consider the following series of cases:

#### THERMOMETERS

Baker does not trust her oven temperature gauge one bit. His baking often requires him to be able to measure the temperature within a couple of degrees Celsius whilst the built-in gauge seems to be entirely inconsistent. Based on past successes and failures, sometimes it does the job nicely, other times it may as well be off by twenty or thirty degrees. For that reason, Baker decided a while ago to invest in a good oven thermometer. Since there was a deal in the shop when he made the purchase, Baker bought himself two new thermometers. Since then, Baker has used each a number of times, and any time he has used either the bake has gone well. Baker's current project is for the auditions of a once popular British TV programme and, in a fit of extra caution, he decides to put both thermometers in the oven. After allowing a few moments for the oven to heat up, Baker checks to see if it has reached 200C as specified in the recipe. He checks thermometer-1 first. It reads 200C. He immediately checks thermometer-2. It reads 220C.

Echoing 'The Problem of Ordinary Disagreement', we can ask about this case: how ought Baker to respond to the conflicting readings from her two thermometers?

There are a few things that we might note about Baker's situation that are relevant to how we answer this question. For one, it makes no difference at all what order Baker checked the thermometers in. If, say the actual oven temperature is 200C – then, other things being equal, the oven temperature would have been 200C if Baker had first checked thermometer-2 and then 1. Thus, we might say, there is (as the case goes) no evidential relationship between Baker's beliefs

---

<sup>25</sup> See for instance Christensen (2007: 196); Feldman (2006: 234); Matheson (2015: 70-71).

about the temperature of the oven and the order in which he checks the two thermometers.<sup>26</sup> For another, it does not matter whether one or other of the thermometers is telling the correct temperature. Baker does not have direct access to the facts about the oven temperature – indeed, at the kind of temperatures in question, he is entirely dependent upon his thermometers to ascertain what that temperature is. Thus, those facts make no difference to what he is justified in believing vis-à-vis the relative functioning of the two measuring instruments. Similarly, Baker can be sure that at least one of the thermometers is on the fritz. Yet, without independent evidence as to whether it is specifically thermometer-1, specifically 2, or indeed both thermometers malfunctioning, that feature of the case affords Baker no reason to be more confident in one reading over the other.

In light of these considerations, then, it is quite clear that Baker ought to be significantly less confident that the oven is at the right temperature after having checked thermometer-2 than before having done so. How much less confident? Given that Baker has no access to information that would allow him to distinguish between the two thermometers, epistemically speaking, he should be no more confident that the oven is at the right temperature than that it is significantly too hot. Or, in other words, Baker should suspend judgement about the oven temperature.

Now consider the following familiar sounding situation:

#### PATIENT-X

Patient-X is being treated for Condition-C by Doctors Quinn and Zaius. Patient-X has no specialist medical knowledge or training but does have all kinds of reason to think that her doctors are both highly competent and trustworthy. Moreover, Patient-X has no reason to think that either Quinn or Zaius is any more competent or trustworthy than the other. Quinn and Zaius have just completed independent assessments of the clinical and scientific evidence pertaining to which of two treatment options would offer Patient-X the best chance of recovery. Quinn informs Patient-X that she believes ‘A is best’ and on that basis Patient-X comes to believe ‘A is best’. A few moments later, Zaius enters the room and informs Patient-X that he believes ‘B is best’.

There are obvious and epistemically significant differences between this case and the last. For one thing, the question of which treatment option is best is significantly more complex than the

---

<sup>26</sup> If the case was such that there was a significant lapse in time between checking the two thermometers things may be different.

question of whether the oven is at the right temperature for a successful bake. Correspondingly, determining the correct answer to the treatment question requires considerably more epistemic sophistication than does the temperature question. In this respect, we might suppose, the general conditions under which it might be appropriate for Patient-X to rely on her doctors' medical opinions will be significantly more demanding than those under which it is appropriate for Baker to rely on his thermometers.<sup>27</sup> For another, where the thermometers upon which Baker relies upon for information about the temperature of his oven are inanimate measuring instruments, the testimony on which Patient-X relies for information about her treatment options comes from living breathing epistemic agents, with all of the flaws, foibles, and idiosyncrasies that come with that. Depending on one's preferred account of the epistemology of testimony, the interpersonal nature of this kind of relationship may generate significant normative differences when it comes to the appropriateness of Baker's reliance on his thermometers and Patient-X's reliance on her doctors (e.g. Moran 2018).

Though these differences are both significant from a broader epistemological perspective, however, it clearly doesn't follow from either that Patient-X will necessarily have any better reason to believe one doctor over the other than Baker has to believe one thermometer reading over the other. Yet, or so the argument goes, it is that comparison between each of Baker and Patient-X's respective sources of information that is crucial. In this respect, there are two important similarities between the cases.

First, just as Baker does not have the sensitivity necessary to judge differences in temperature in the ranges at which he is baking without her thermometers, Patient-X does not have the expertise necessary to judge the bearing of the clinical and scientific evidence upon his treatment options. Thus, just as Baker is epistemically dependent on his thermometers when it comes to determining whether the oven is at the right temperature or not, so Patient-X depends upon her two doctors for her beliefs about the treatment options.<sup>28</sup> So we might say, to borrow a term from David Enoch (2010), Quinn and Zaius are Patient-X's 'Truthometers'.

---

<sup>27</sup> Nonetheless, it is implicit in the statement that 'X has all kinds of reason to think that his doctors are both highly competent and trustworthy' that, whatever the conditions of justified testimonial reliance are, Patient-X has satisfied these with respect to both Quinn's and Zaius's testimony.

<sup>28</sup> Moreover, given his lack of expertise, this would be so even if Patient-X had *informational* access to the relevant evidence. Just because he has looked at his charts and figures, doesn't mean he can justifiably draw any conclusions from that data.

Secondly, just as Baker is not in a position to distinguish between his two thermometers, epistemically speaking, as the case is described, Patient-X lacks any information that would allow her to determine whether either of her doctors has the epistemic advantage vis-à-vis the question of which treatment option is best. Correspondingly, even if it is so that one of Quinn or Zaius has done a better job of assessing the clinical and scientific evidence, since Patient-X has no way of gleaning this, that fact would have no significance for what Patient-X is justified in believing in the face of their disagreement – just as the facts about which, if either, thermometer is on-the-fritz have no significance upon what Baker is justified in believing about the temperature inside his oven.

In light of these similarities, the conciliationist supposes that Patient-X ought to respond to the disagreement between her doctors in the same way as Baker responds to the conflict between his thermometers. I.e. Upon realising that Zaius disagrees with Quinn, Patient-X ought to reduce her confidence in the truth of Quinn's testimony. How confident should Patient-X be? No more confident that Quinn is correct than she is that Zaius is correct. And vice-versa.

It is easy to see where this is going.

The two cases described so far are described from the third-person perspective. Our core cases of disagreement between effective peers are described from the first-person perspective. Nonetheless, the key similarity between Baker and Patient-X's situations is that neither have access to the kind of evidence that would justify believing that the information from either source in particular is more likely to be accurate than the other. And, so the thought goes, two people are effective peers just when they have no evidence that would justify believing that either party in particular is more likely to be right than the other – or at least no evidence that is independent of the disagreement. The conciliationist suggests, then, that the evidential position one finds oneself in when one has a first-person perspective on effective peer disagreement is analogous to the evidential positions of Patient-X and Baker in the third-person cases. Consequently, Quinn and Zaius should respond to their disagreement in just the same way that Baker ought to respond to her thermometers, and Patient-X to his realisation that Quinn and Zaius disagree. That is, neither Doctor should be any more confident in their own assessments of the clinical and scientific evidence than they are in their colleague's assessments – and they should revise their beliefs accordingly. In other words, they should do exactly what Conciliationism tells them to do.

If the analogy between these cases holds, the conciliationist has a powerful argument in their favour. However, the non-conciliationist is likely to raise an eyebrow at the last step in the

argument. As noted, Patient-X is similarly dependent upon Quinn and Zaius's judgement of the clinical and scientific evidence as is Baker on the reading of her two thermometers. Interpersonal views of testimony aside, then, it is easy to suppose that Patient-X views her two doctors in an analogous way to that in which Baker views her two thermometers. Neither Quinn, nor Zaius, though, are epistemically dependent on the other in the same fashion. Both are directly acquainted with the relevant evidence and both are competent and qualified to assess the bearing of that evidence upon Patient-X's treatment options. As such, it is not that Quinn and Zaius have *no* evidence relevant to determining who is more likely right, but that – as effective peers – they don't have any evidence relevant to that question that is independent of the substance of the disagreement. Of course, since neither Baker, nor Patient-X have direct access to *any* evidence relevant to determining which source of information is correct, they also lack any *independent* evidence of that kind. But still, the non-conciliationist might ask, why suppose that it is this feature of the various cases that counts, and not the contrast with Quinn and Zaius' first-person perspective on their disagreement? Similarly, why suppose that it is not the dis-analogous features that count in cases of disagreement between effective peers more generally?

As we shall see in the next chapter, Enoch (2010) and Schafer (2015) argue against the conciliationist on just this point. In principle – they claim – having a first-person perspective on the disagreement can ground justification for retaining one's original beliefs. I shall turn to their arguments in the next chapter. What is clear, though, is that some further principle is required if the 'Truthometer' analogy is to hold. For that, we can turn to the second argument for Conciliationism – the 'Independence' argument.

### 2.3.2 The Independence argument

David Christensen suggests that what separates Conciliationism from other views of disagreement is acceptance of a principle along the following lines:

Independence: In evaluating the epistemic credentials of another's expressed belief about P in order to determine how (or whether) to modify my own beliefs about P, I should do so in a way that doesn't rely on the reasoning behind my initial belief about P (Christensen 2011: 3).

In addition to Christensen, 'Independence' is explicitly endorsed by Jon Matheson (2015: 82-83, 95-96) and is at least implicitly a part of Elga's (2007) arguments for Conciliationism. So, for instance, Elga suggests that: 'Even if in fact you have done a much better job than your friend at evaluating the claims, simply comparing your verdicts to those of your friend gives no evidence that this is so,' (Elga 2007: 487) and 'Upon finding out that an advisor disagrees, your probability

that you are right should equal your prior conditional probability that you would be right. Prior to what? Prior to your thinking through the disputed issue and finding out what the advisor thinks of it...’ (Elga 2007: 490).

Given the focus on this principle, it will be useful to rephrase it in keeping with the terminology and language used in this thesis. Since I have suggested the debate concerns what disputants are justified in believing in the face of disagreement, we can do so by restating Independence as a principle of epistemic justification. Thus, we might restate it as follows:

*Independence*

If  $S_1$  realises that  $S_2$  disagrees with her about  $Q$ ,  $S_1$ 's attitude toward  $Q$  (whatever that attitude might be) will be justified *only if* it is based upon justification that is independent of the substance of the disagreement.

Given this reformulation, it is clear how Independence might lead us to accept Conciliationism as the correct response to ‘The Problem of Ordinary Disagreement’. Consider: ‘The Problem of Ordinary Disagreement’ arises when disputants are effective peers. As the term was defined, if  $S_2$  believes  $Q$  and  $S_2$  believes  $\sim Q$ ,  $S_2$  is  $S_1$ 's effective peer if and only if  $S_1$  has no undefeated reason *independent of the substance of the disagreement* to believe that she or  $S_2$ , in particular, holds the relative epistemic advantage vis-à-vis  $Q$ . That being so, the only reason that  $S_1$  could have for believing that she is more likely to be correct about  $Q$  than  $S_2$  is the disagreement itself. However, the use of this reason is ruled out by Independence on principle. Hence, if Independence is true,  $S_1$  cannot justifiably conclude that she is any more likely to be right about  $Q$  than  $S_2$ . Consequently,  $S_1$  ought to revise her beliefs. How ought she to revise her beliefs? Given that she cannot justifiably conclude that she is any more likely to be right than  $S_2$ ,  $S_1$  ought to be no more confident that  $Q$  than she is that  $\sim Q$ . I.e. She ought to respond just as Conciliationism suggests.

Independence, then, appears to lead to Conciliationism as a matter of course. Additionally, this is precisely the kind of principle required to legitimize the move from the third- to the first-person cases in the ‘Truthometer Analogy’: For as DOCTORS-1 was described, Quinn and Zaius are effective peers. Thus, if Independence is true, Quinn and Zaius would be no more justified in deciding their disagreement one way or other via their having a first-person perspective on the substance of the disagreement, than Patient-X would be in deciding who to favour in that disagreement on the basis of which of his doctors entered the room first. Likewise, for any other case of disagreement between effective peers.

For what reason might we think that Independence is true?

As Christensen sees it, the motivation behind the principle is ‘to prevent blatantly question-begging dismissals of the evidence provided by the disagreement of others.’ (Christensen 2011: 3)

Arguments like this:

Well so-and-so disagrees with me about P. But since P is true, she’s wrong about P. So however reliable she may generally be, I needn’t take her disagreement about P as any reason at all to question my belief (Christensen 2011: 3).

That there are question marks over the legitimacy of an argument of this sort is undeniable. In particular, it is a clear parallel to so-called Dogmatism Paradox, described here by Gilbert Harman (who attributes the puzzle to Saul Kripke):

If I know that h is true, I know that any evidence against h is evidence against something that is true; so I know that such evidence is misleading. But I should disregard evidence that I know is misleading. So, once I know that h is true, I am in a position to disregard any future evidence that seems to tell against h (Harman 1973: 148-9).

This seems obviously problematic. For sure, if I know that Q, under certain circumstances my knowledge that Q *can* ground me in dismissing evidence that  $\sim$ Q. For instance, I know that the Earth is round, and surely this knowledge *is* sufficient grounds for me to dismiss the evidence for believing that the Earth is flat that I acquire from watching the videos of conspiracy theorists on YouTube – even if I do not have available any means of responding to the specific conspiracy theories forwarded in those videos. But, if we embrace the dogmatism paradox, then we don’t just allow that there will be some situations in which my knowing Q can ground my dismissal of evidence against Q. We allow that my knowing Q can ground my dismissing any and all evidence against Q that I may encounter in the future. But, as Sorensen puts it, a policy of dismissing any and all evidence that might count against one’s beliefs is ‘a policy of close-mindedness that strikes us as ‘perniciously dogmatic’ (Sorensen 1988: 434) A policy that is ‘perniciously dogmatic’, though, is clearly a bad policy, hence, it is reasonable to suppose that there must be something wrong with the argument in the dogmatism paradox. Christensen’s further comments on the question-begging dismissal of disagreement that he describes, make it clear that he sees that response as flawed in just the same way. So, he writes:

Used as a general tactic, [the question-begging response] would seem to allow a non-expert to dismiss even the disagreement of large numbers of those he took to be experts in the field.’ (Christensen 2011: 3).

And, indeed, this seems entirely right. If it is permissible to argue from the fact of disagreement to the conclusion that the other parties to the disagreement are wrong *no matter what other information*



*you have about their reliability*, then, just like the dogmatism paradox, it will be permissible to dismiss any and all evidence one might acquire on learning of disagreement as misleading evidence. Thus, a theory of disagreement that allows the question-begging response to disagreement that Christensen describes, will be committed to the claim that one can justifiably stay steadfast in the face of any and all disagreement, no matter the context, or the people with whom one disagrees. Yet, to paraphrase Sorenson on the dogmatism paradox, such a broad policy of responding to disagreement would be perniciously dogmatic. As such, any theory of disagreement that permits this response will be an implausible theory of disagreement.

Independence straightforwardly rules out this kind of response to disagreement. For, the question-begging response to disagreement that Christensen describes is question-begging just because it involves one's taking the substance of disagreement as sufficient reason to stay steadfast. Yet, this is precisely the kind of response that Independence rules out. Thus, in so far as it rules out a policy of dogmatism, Independence is well motivated. Moreover, since Independence leads to Conciliationism as a matter of course, so is Conciliationism. Whether this is *sufficient* motivation for the view, however, will depend upon whether it is necessary to accept Independence to rule out such a policy of dogmatism in response to disagreement. And what is important to note here is that – whilst it does rule out such a policy – this is not the limit of the theoretical implications of Independence. In particular, there do appear to be cases in which it seems unproblematic to respond to disagreement via an argument of the form Christensen describes. The example of my summary dismissal of the flat-earth conspiracy theorists on YouTube, for instance, would appear to fit the bill. And Kripke offers another, when he writes:

I myself have not read much defending astrology, necromancy, and the like (I remember Stephen Weinberg making the same remark). Even when confronted with specific alleged evidence, I have sometimes ignored it although I did not know how to refute it. I once read part of a piece of writing by a reasonably well-known person defending astrology... (Kripke 2011: 49).

In both these examples, I would suggest, far from being dogmatic, it would be entirely reasonable to summarily dismiss the opinions of the opposing parties on the basis of the disagreement alone. After all, even if I do not *base* my response to the disagreement with the flat-earthers on independent considerations, someone who believes the Earth is flat is clearly not my peer, effective or otherwise, when it comes to that question. Similarly for Kripke's response to astrologists, necromancers, and the like. Crucially, however, if Independence is true, then, these responses are just as problematic and just as dogmatic as it would be to respond this way in case of disagreement with an effective peer or, indeed, one's epistemic superior.

As it happens, Christensen (2010) and Matheson (2009, 2015) do have more to say on the motivations for Independence. Specifically, both leverage the concept of *higher-order evidence* to make the case for the combination of Independence and Conciliationism. Since discussion of higher-order evidence takes us into deeper waters than we are yet prepared for, however, I shall leave this discussion aside until Chapters 4 and 5. For the moment, the important point is that Independence implies a greater theoretical commitment than it first seems. The relevant question, then, is not whether such paradigmatically question-begging defences of one's position as Christensen describes must *always* be permissible – they clearly cannot be – but whether anything as strong as Independence is required to rule them out. Unsurprisingly some non-conciliationists have staked a claim for their preferred view of disagreement on arguing that it is not. I shall consider Kelly's (2010) arguments to this effect in the next chapter.

Discussion of the wider implications of Independence brings us neatly to what I consider the principal criticism of Conciliationism on its own terms, the problem of 'Spinelessness'. Having seen arguments in favour of Conciliationism in the preceding section, then, let us turn to this criticism of the view.

## 2.4 Arguments against Conciliationism

### 2.4.1 The Problem of Spinelessness

One worry that we might have about Conciliationism is that it appears to require, in Elga's (2007) words, a certain kind of *spinelessness*. For any issue, when faced with disagreement with an effective peer, we should shrug our shoulders and roll over on what we believe. This may not bother us so much in bill-sharing cases – indeed, I suspect that few of us now encounter disagreements about restaurant bills that are not quickly resolved by whipping out a phone. But on many issues we deeply care about – philosophical, scientific, religious, and political issues for instance – we disagree both widely and persistently. If Conciliationism requires that we suspend judgement not just on dinner bills but on any controversial question, including those that concern issues we care deeply about, we may have some motivation for preferring an alternative account of disagreement. To see just how deeply Spinelessness cuts against Conciliationism, however, it will be useful to consider first how Conciliationism can be extended beyond the simple one-on-one cases we have seen so far to also include cases of disagreement between more than two people. As I argue in the next section, when Conciliationism is extended this way it is not only open to the charge that we ought to suspend judgement on many issues that we care deeply about, but, more insidiously, that we ought to forego judgement on many of those issues altogether.

#### 2.4.1.1 Spinelessness and multi-party disagreements

Conciliationism says that, whenever you disagree with a peer about some proposition or propositions, you ought to revise your attitudes toward those propositions so as they are in line with the middle ground between your and your interlocutor's positions. Another way of putting this is that you ought to accept only those propositions on which there is *consensus*. To illustrate this thought, think back to Doctors-1. In that case, we might imagine that the pair had reached prior agreement that the best treatment option is one of A or B, even whilst coming to disagree on which of those two options is best. In categorical terms, Conciliationism advises that each continue to believe the proposition on which there is consensus, 'A or B is best', and suspend judgement on each of the component propositions 'A is best' and 'B is best'.

Of course, not all disagreements will involve an explicit pre-disagreement consensus of this sort, but one might often be extrapolated, nonetheless. So, we would certainly not expect Jack and Jill in Diners-1 to have reached a prior consensus that each diner's share would be '\$43 or \$45'. Nonetheless, given the features of that case, it seems they might justifiably assent to that proposition upon realising the disagreement. After all, both are typically competent when it comes to simple bill-sharing calculations, thus, without evidence that *both* are prone to some kind of performance error, it seems reasonable for them to take it as unlikely that both have made a mistake (though that does remain a possibility). Thus, despite having no prior consensus on what answers might be in contention, Jack and Jill could justifiably come to believe that the answer to how much each diner owes is either \$43 or \$45. Thus, Conciliationism would prescribe that they each come to believe that proposition, but neither of the specific component propositions 'Each diner owes \$43' or 'Each diner owes \$45'.

This feature of Conciliationism is important in two respects. Firstly, it allows that in cases of this sort – where there are relatively few rival positions in play – the sceptical consequences of following the tenets of Conciliationism are limited in scope. For, whilst Conciliationism would preclude disputants justifiably believing any of the disputed propositions, under the right circumstances, it does not preclude their justifiably believing that at least one of a narrow range of positions is likely to be correct. That being so, those party to such cases can at least remain confident that they have some and (perhaps some good) grasp of the issue under dispute – even if they ought not to be so confident as to settle on one of the positions in play. At least in respect to isolated cases, then, perhaps it would be fair – as some conciliationists do – to suggest that the view calls not for a dispiriting *spinelessness*, but only a modicum of *intellectual humility*. As Christensen captures the thought:

After all, I want my own beliefs to be those best supported by the evidence. So if the beliefs of other decent, yet imperfect, inquirers turn out to serve as partial checks against my falling short of this goal, that strikes me as being pretty good news (Christensen 2007: 216).<sup>29</sup>

Unfortunately for the conciliationist, however, the sceptical consequences of Conciliationism are not necessarily limited in this way. I shall explain why in short order. Before doing so, however, we need to consider the other aspect of Conciliationism that the point about consensus brings out.

Secondly, then, the emphasis upon consensus helps us to see how Conciliationism neatly generalises from simple cases of one-on-one disagreement to cases in which an individual agent disagrees with a number of other parties. Call these *multi-party disagreements*.<sup>30</sup>

Of course, very many of the disagreements we actually encounter are multi-party disagreements. Thus, it is a strength of Conciliationism that it extends to these cases quite straightforwardly. For, when we put its recommendations in terms of consensus, Conciliationism offers much the same advice in multi-party disagreements as for one-on-one disagreement: I.e. If you disagree with several peers/effective peers about some issue, you ought to accept the group's consensus judgement on that issue.<sup>31</sup>

In cases where there is a clear consensus this does not appear to be unreasonable advice. Imagine, for instance, that the rest of Jill and Jack's party weigh in on the bill calculation, all ten of them agree with Jack, and each one of them is Jill's effective peer. Call this Diners-2. In this case, whatever we think about the original case, I take it that the pressure on Jill to revise her belief is intuitively greater than in the simple one-on-one situation. In particular, if we presume that the various parties reach their positions independently, and given that, as her effective peers, Jill has reason to believe the other parties are all competent in the relevant ways, the fact that the other

---

<sup>29</sup> On similar lines, Feldman writes: '[P]erhaps the conclusion is not so distressing. It calls for a kind of humility in response to the hard questions about which people so often find themselves in disagreement [...] When compared with the intolerant views with which we began, this is a refreshing outcome.' (Feldman 2007: 213) and Matheson: '[E]pistemic humility is an intellectual virtue as well. One thing that the existence of disagreement brings out is that we are limited and fallible epistemic agents [...] Given that we are fallible epistemic agents, agnosticism in the face of controversy seems to be the appropriate response, and not simply the result of lacking convictions or courage.' (Matheson 2015: 140).

<sup>30</sup> I call these *multi-party* rather than *group* disagreements since the latter term might also refer to disagreement *between* groups. Questions about how groups ought to respond to disagreement with other groups (or indeed with individuals) would appear to be entangled with questions about how we ought to view *group agents* from an ontological perspective. For that reason, I take *group disagreements* to fall outside the purview of 'The Problem of Ordinary Disagreement' in a way that multi-party disagreements do not.

<sup>31</sup> Discussions of this aspect of Conciliationism can be found in (Christensen 2009; Elga 2007; Matheson 2015: 125-45)

diner's all agree is powerful evidence in favour of their position. This is just a straightforward application of the results of Condorcet's Jury Theorem. That being so, the intuition that Jill is under considerable pressure to accede to the collective opinion is something that any plausible view of disagreement ought to accommodate and explain. Conciliationism, fares well on both fronts. For, what it says about this case is that, even if she is the only member of the group to have correctly calculated each diner's share, since the consensus strongly favours Jack's answer, Jill ought to accede to that majority view and revise her belief accordingly.

Whilst this easy generalisation to multi-party cases is a strength of Conciliationism, however, it also presents problems. As noted, the humility-not-spinelessness defence does not seem entirely unwarranted if we are only concerned with cases in which the range of disputed positions is relatively narrow. Since Conciliationism so easily extends to multi-party disagreements, however, it also extends easily to cases in which the range of rival positions is anything but narrow. In these cases, however, I would suggest Conciliationism is committed to saying that one ought to take no stance at all upon the disputed issue. Or, in other words, in cases where there is no real consensus and none is likely to arise, Conciliationism will recommend that one *foreswear judgement* altogether on that issue. To illustrate, consider the following case.

#### A DIVIDED STATE

Earlier in the evening of the dinner party from Diners-2, the group turn to discussing politics. By and large the group all agree that there are clear economic and social issues in the country that need to be addressed, for instance, they all agree that rising economic inequality is a problem and that something needs to be done to make society less divided than it has become. As it happens, there is an election coming up and there are six parties ranging across the political spectrum with a serious chance of winning seats. An avowedly socialist party and a less radical social-democratic party on the left. A left-of centre party espousing progressive social positions and light touch regulation on the market. A right of centre party endorsing de-regulated free-market economics and slowing moving further right on social issues. And on the far-right an ethno-nationalist free-market party. Finally, there is a party fronted by a popular comedian, with no policies but a slogan: 'We can't do any worse than that lot!'

As far as Jill knows all of the diners are *au fait* on the relevant issues as well as the various party policies, characters, and manifestos. In other words, all of the other diners are Jill's effective peers on the political stuff. Jill herself thinks that the party best suited to sorting

out the economic and social issues in the country is the social-democratic party. Somewhat to her surprise, however, opinion in the group is split evenly across the five parties. Jack agrees with Jill that the social democrats are the countries best bet. The other diners are split into pairs supporting the remaining five parties – including the far-right party and the comedians party.

What does Conciliationism say about this case?

Given that there are six parties and the group's endorsements are split evenly amongst them, there is not even a sign of consensus that Jill might favour. Without any consensus to go by, then, and given that an equal number of her peers endorse each party, Conciliationism would appear to recommend that Jill have no more confidence that any party is better suited to addressing the economic and social issues than any of the other. There are significant and unpalatable consequences to this response. For one thing, given that there are six rival opinions in play, were Jill to have equal confidence in each of them, she ought to have very low confidence that any particular one of these opinions is correct. But, if Jill cannot be at all confident that any one of the possible opinions is right (there are after all only six parties with a chance of winning seats), Jill should admit to herself that she has no real clue which party has the best chance of sorting out the country's problems. Perhaps more significant than this is the fact that the rival positions fall across such a wide spectrum of political positions. It is one thing to say that, faced with disagreement from her variously left-leaning friends, Jill should be open to the possibility that maybe one of the other parties on the left may be a better option than the social democrats. It is quite another, to say that someone who supports the social-democratic party ought to be equally open to the possibility that the far-right party (or indeed, the a-political party) is the solution to the country's woes. Even if they are only considering the fact-based policy issues, it seems entirely reasonable that Jill could dismiss from consideration a party just because they are championing a vision of society so far removed from her own. Conciliationism, however, expressly rules out this option. Given that all of Jill's friends are her effective peers, Jill ought to give equal weight to each of their positions – whether to the left, or to the far-right of her (and whether in the realm of the absurd). I would argue, though, that were Jill to respond in this way, she would go far beyond showing a modicum of humility and well into the domain of spinelessness.

In cases where we disagree with our effective peers and our peers take a wide range of positions, then, what Conciliationism recommends looks a lot more like spinelessness than humility. In these cases, encountering disagreement with others does not act as a partial check on hubristic overconfidence, but is rather a barrier to having any justified opinion on the matter at all.

How problematic is this for Conciliationism?

As Feldman (2006) and Matheson (Matheson 2015: 141) point out, the kind of scepticism that follows from Conciliationism is a ‘contingent’ scepticism. If one encounters no disagreement, or if it is always clear that one has the relative advantage in those disagreements one does encounter, following the tenets of Conciliationism will lead to no sceptical consequences. Likewise, if we encounter no multi-party disagreements of the relevant kind, the sceptical consequences will at most be limited to suspension of judgement on disputed issues, and not the thoroughgoing loss of confidence that seems required in a case like *Diners-2*. These points are not without *theoretical* significance. As noted in the previous chapters, the debate about the epistemic significance of disagreement presupposes that justified belief is possible. Since the sceptical consequences of Conciliationism are contingent on the kinds of social epistemic interaction we find ourselves in, it remains possible on this view that justified beliefs are possible. Thus, the view is not, in principle, inconsistent with an important presupposition of the debate.

However, this last observation is not enough to lead Conciliationism completely out of the woods. Given the consequences in multi-position disagreements it remains a real possibility on Conciliationism that *the world we live in* is one in which we ought to accept that we have very little idea about many issues we have formed beliefs upon. As per the above analysis, all that would be necessary for this to be so would be that such disagreements are in fact a common part of our social experience. To refer back to the previous points vis-à-vis humility vs spinelessness, though, it is one thing to conclude that the social world we live in requires a little more caution and humility in our epistemic endeavours than we might have realised, but quite another to conclude that people generally ought to accept that they have no clue about what they are doing in those endeavours. To borrow a Moorean strategy, the latter claim seems less plausible than the claim that it is the conciliationist who has gone wrong at some point. If the conciliationist does not have the resources to block that conclusion, then, I take it that there is a problem with the view.

Importantly, and I think rightly so, conciliationists have not been of a bent to deny that we *do* encounter widespread and persistent disagreement, and particularly so on matters that are important to us (for instance Christensen 2009; Elga 2007: 492-95; Matheson 2015: 141) Likewise, I take it, conciliationists would not deny that many of those disagreements will be ones involving multiple parties and a wide range of contested positions. That being so, for the conciliationist to satisfactorily contain the sceptical consequences of the view as they would apply in the world that we live in, the best defence will be to argue that such cases, though common enough, are unlikely

to be cases of peer disagreement. Adam Elga offers a response to Spinelessness on just these lines. In his own words:

[W]ith respect to many controversial issues, the associates who one counts as peers tend to have views that are similar to one's own. That is why – contrary to initial impressions – the equal weight view does not require one to suspend judgement on everything controversial (Elga 2007: 494).

Whilst, then, Elga appears to accept that the Spinelessness objection would have teeth if scepticism in the world follows from Conciliationism, he denies that it does. This claim is not without some motivation. As Elga points out, controversies in the real world do not tend to be controversial in isolation but are ‘tangled in clusters of controversy’ (Elga 2007: 494). So, for instance, to borrow Elga's own example – the claim that abortion is morally permissible is closely linked to other deeply controversial issues – ‘whether human beings have souls, whether it is permissible to withhold treatment from certain terminally ill infants, and whether rights figure prominently in a correct ethical theory’ (Elga 2007: 494). If, then, two people – Ann and Beth – disagree about the first claim, Elga suggests, they are likely to disagree also on a number of these related issues. That being so, he argues, they ought not to see each other as peers on the original issue. As Elga explains:

[T]he accuracy of an advisor's views on these allied issues indicates how accurate the advisor is likely to be, when it comes to abortion. The upshot is that Ann does not consider Beth an epistemic peer with respect to the abortion claim’ (Elga 2007: 493).

To refer back to the previous chapter, this does not seem an unreasonable position to take when we are talking about *justificatory peerhood*. To be justificatory peers Ann would need to be justified in believing that Beth is her epistemic equal. Yet, as I think Elga is right to suggest, the more issues that two people disagree upon – especially if those issues are interconnected – the less likely that they would believe that they are epistemic equals and, so too, the less likely that they would be justified in believing that. So, Elga is probably right about such cases when it comes to justificatory peerhood.

As we saw earlier, however, the crux of the conciliationist view is the principle of Independence. And the conciliatory force of Independence comes into effect – not only when disputants are justificatory peers – but also when they are effective peers. (Indeed, that is just why it counts as a response to the problem of ordinary peer disagreement!) An important feature of effective peerhood, though, is that two people – or indeed a number of people – can be effective peers even when they can be sure that one of them has the epistemic advantage, just so long as they cannot



independently determine which of them has that advantage.<sup>32</sup> Thus, Ann may well justifiably believe on the basis of their wider disagreements that she and Beth are not justificatory peers with respect to the abortion claim. But she cannot conclude on that same basis that it is she who has the epistemic advantage. Or, more to the point, she cannot do so without contravening Independence. As such, if she has no other independent reason to believe she has the relative advantage, according to Conciliationism, Ann should be no more confident in the truth of her own opinion *on any of the areas of controversy* than she is in the truth of Beth's opinions on those matters.

Given that the sceptical consequences of Conciliationism apply to cases of disagreement between effective peers, the conciliationist cannot ward off Spinelessness through Elga's containment strategy. This applies equally well to the more insidious version of that objection I have described in this section. Moreover, since these problems flow directly from the principle of Independence, it is hard to see how the conciliationist could resist them without otherwise stripping the view of much of its appeal. To deny Independence would be to undermine the core arguments behind Conciliationism. To deny that the relevant kinds of case are a common feature of our social world, given the relatively undemanding conditions of effective peerhood, would be to ground the view on a questionable empirical claim. One that appears to go against the experience of disagreement on many of the issues, e.g. philosophical, political, scientific, etc, that motivate the discussion of disagreement to start with. Either way, then, we have strong motivation to develop an alternative response to the problem of ordinary peer disagreement.

In the next chapter, I shall take a closer look at the chief contenders.

---

<sup>32</sup> To be fair, Matheson makes a similar point in his discussion of how Conciliationism extends from idealised cases (effectively disagreements between justificatory peers) to everyday cases. (Matheson 2015: 116).

# Chapter 3. The Problem of Ordinary Disagreement: Steadfast and Non-uniform views

## Introduction

In the previous chapter, I considered the case for the conciliatory view of disagreement. In this chapter, I consider the case for Steadfast and non-uniform views of disagreement. I consider three positions: A defence of steadfastness that I refer to as the ‘Default Self-Trust View’, variations of which have been forwarded by David Enoch and Karl Schafer; Jennifer Lackey's non-uniform ‘Justificationist View’; and Thomas Kelly’s non-uniform ‘Total Evidence View’.

## 3.1 Self-trust and Steadfastness

As discussed in Chapter 1, at the heart of ‘The Problem of Ordinary Disagreement’ is the question of how an agent ought to respond to the realisation that they disagree with their effective peer or peers. Importantly, this question concerns how the agent in question ought to respond *when they are an active participant in that disagreement* – not when they are a third-party to a disagreement between others. As we saw in the previous chapter, conciliationists argue that having this kind of first-person perspective makes no difference to what one is justified in believing in response to the realisation of disagreement. Instead, for the conciliationist, the first and third person perspectives on peer disagreement are *both* analogous to encountering conflicting readings from apparently equally reliable inanimate measuring instruments. And, just as an agent in the latter situation clearly ought to suspend judgement on the matter of which instrument reading is correct, so the conciliationist claims, an agent who encounters a disagreement between effective peers ought to suspend judgement upon the substance of that dispute – whether they are a third-party to that disagreement, or one of the disputants themselves. One prominent strand of non-conciliationist thinking has been to reject this line of argument – and so too Conciliationism – through recourse to the idea of *epistemic self-trust*. In some way, so the thought goes, it is necessary to have a general kind of trust in the reliability of the practices by which one comes to form and hold beliefs. In contrast, it is not necessary that one place such trust in the practices by which any *specific* other comes to form and hold their beliefs. When one has a first-person perspective on disagreement, this disparity between the need for self-trust and the need (or lack thereof) for trust in one’s specific

interlocutors can license one's staying steadfast in the face of disagreement – even when the other parties to the disagreement are one's effective peers.

Arguments of this kind have been offered by David Enoch and Karl Schaffer. Enoch and Schaffer both argue along the lines that, given a general epistemic *entitlement* to trust your own epistemic practices, it is rationally permissible to treat the fact of disagreement as evidence against the epistemic credentials of someone you disagree with.<sup>33</sup> In cases where you disagree with an effective epistemic peer, this evidence can justify believing that you have the relative epistemic advantage over your interlocutor and so tilt the balance in favour of your original position on the disputed issue. This form of argument can be treated as a straightforward rebuttal of the 'Truthometer' argument for Conciliationism. In so far as this argument relies on the idea that we each have a default entitlement to trust our own epistemic practices – I shall refer to it as the 'Default Self-Trust View'. In the rest of this section I shall reconstruct this position, before assessing how well it fares as a defence of Steadfastness.

### 3.1.1 Default Self-Trust

I will offer a more detailed account of the nature and normative significance of epistemic self-trust (as well as engagement with the literature on that) in Chapter 5. Since, though, neither Enoch nor Schaffer argue as such for the need for epistemic-self-trust themselves, it will be enough for our current purposes to offer a rough characterization of the thesis of self-trust as it appears in their arguments. To differentiate from the view I develop in Chapter 5, let us call this the thesis of 'Default Self-Trust' (DST for short). There are two aspects to DST, a descriptive one, and a normative one. The descriptive aspect is clearly expressed by Enoch in the following passage:

Whenever you try to decide how much trust to place in someone, or indeed, *when deliberating epistemically about anything at all*, your starting point is and cannot but be your own beliefs, degrees of beliefs, conditional probabilities, epistemic procedures and habits, and so on (Enoch 2010: 980. Italics Added).

---

<sup>33</sup> Ralph Wedgwood develops a similar view to Enoch's and Schaffer's in his 2010 paper 'The Moral Evil Demons'. I leave Wedgwood's account out of the current discussion mainly to avoid any complications that might come with extending discussion to also include the kinds of 'moral disagreement' that are Wedgwood's main focus of attention.

Linda Zagzebski and Robert Pasnau, meanwhile, offer defences of steadfastness via self-trust that are quite different in spirit to the approach described above. So, Zagzebski suggests that we resolve disagreement by considering 'which beliefs are more likely to satisfy my future self-reflection' (Zagzebski 2012) and that self-trust plays a crucial role in determining this. Robert Pasnau (2015) argues that self-trust has intrinsic non-epistemic value, in respect to which, one can have non-epistemic reasons to stay steadfast in the face of disagreement, even when one does not have epistemic reason to do so. Interesting though both of these accounts are in their own right, to do them full justice would take us too far from the boundaries of the current discussion.

Schafer captures the normative aspect of the thesis, when he posits that:

[I]t is also quite natural to think that there is a basic sense in which we are *rationally entitled* to rely on all our basic-belief forming faculties, whatever they are, provided that we do not have any positive evidence that they are not reliable (Schafer 2015: 28. *Italics Added*).<sup>34</sup>

Following these comments, then, we might formulate the thesis of Default Self-Trust as it corresponds to the rest of Enoch's and Schafer's arguments as follows:

(DST)

- (i) For any proposition Q, S cannot determine the truth of Q without relying upon her own belief forming practices and faculties.
- (ii) S is *entitled* to rely on her own belief forming practices and faculties, when it comes to determining the truth of Q.<sup>35</sup>

Since we are talking about relations of epistemic dependence, to say that S *relies* upon her own belief forming practices and faculties, roughly, is to say that S *acts or reasons as if* the deliverances of those faculties are reliable guides to the truth. To say that S is *entitled* to rely on those practices, following Burge (2003, 1993) and Wright (2004), is to say that S is rationally permitted to act/reason as if the deliverances of those faculties are reliable guides to the truth – even if S does not have independent justification for believing that they are. Importantly entitlements are defeasible. Thus, depending on the particular context, S's entitlement to rely on her basic belief forming practices and faculties may not hold.

I shall discuss these concepts in more detail in the fuller discussion of epistemic self-trust developed in Chapters 4 and 5. For now, however, let's grant that this thesis is true as described. How, then, does it support a steadfast view of disagreement?

---

<sup>34</sup> See also Enoch: [T]he underlying thought that we are not entitled to trust our own epistemic abilities to a degree greater than that their track-record calls for seems a close relative of the claim that we are not entitled to employ a belief-forming method without first having an independently justified belief in its reliability (Enoch 2010: 991).

<sup>35</sup> It is worth noting explicitly that this formulation drops Schafer's reference to '*basic* belief forming faculties.' As I shall discuss in more detail in chapter 5, one's 'basic' faculties (or epistemic practices, as I prefer) are those which one could not justifiably believe to be reliable unless permitted to presuppose that those very same practices are reliable. Whilst Schafer's formulation of the entitlement to self-trust in terms of 'basic' faculties is a fair representation of the wider literature on epistemic self-trust (e.g. Alston 2005; Foley 2001; Alston 1986; Fricker 2016; Zagzebski 2012), both he, and Enoch for that matter, construct their arguments in favour of steadfastness as if the entitlement to self-trust extends also to one's non-basic practices. For the sake of simplicity, then, I shall proceed as if that is the case and as if that is a reasonable assumption to make.

### 3.1.2 The Default Self-Trust View

As noted, the arguments proffered by Enoch and Schafer can best be understood as arguments against the ‘Truthometer Analogy’. We have mentioned the central thought behind this analogy already, i.e. that it should not matter to how you respond to a disagreement that you are one of the disputants. Enoch and Schafer both reject this claim. At the same time, both appear to accept another presupposition of the ‘Truthometer Analogy’: That whatever one is justified in believing in response to the realisation of disagreement must be probabilistically coherent with whatever one is justified in believing about the respective epistemic credentials of the disputants. Call this presupposition ‘Credentials Only’. We can describe it more formally like so:

Credentials-Only (CO): If S realises that  $S_1$  believes Q and  $S_2$  believes  $\sim Q$ ,  $S_1$  will be justified in believing that Q is more likely to be true than false *iff* S is justified in believing that  $S_1$  has the relative epistemic advantage over  $S_2$  vis-a-vis Q, and vice-versa.

Importantly, neither Enoch, nor Schafer reject CO.<sup>36</sup> Nor do they reject the idea, implicit in Conciliationism, that CO extends to all cases of disagreement – not just those where one is a third-party to a disagreement between others. Given those two assumptions, however, Enoch and Schaffer would also appear to be committed to the following:

No Advantages (NA): If,  $S_1$  believes Q,  $S_1$  realises that  $S_2$  believes  $\sim Q$ , and  $S_1$  is not justified in believing specifically that  $S_1$ , or  $S_2$  has the relative epistemic advantage vis-à-vis Q,  $S_1$  should suspend judgement on Q.

NA is a straightforward extension of CO to the first-person case under the condition that all questions about the normative significance of disagreement itself have been answered. Crucially, NA does not entail any particular answer to ‘The Problem of Ordinary Disagreement’ (even whilst it is clearly consistent with Conciliationism). For, ‘The Problem of Ordinary Disagreement’ concerns the question of how you ought to respond to disagreement with your *effective peer*. And, as defined, someone is your effective peer in case of disagreement if and only if you have no undefeated and *dispute-independent* reason to believe that one of you (specifically) has the relative

---

<sup>36</sup> In Enoch’s words: ‘In such a case, then, it seems clear that you should treat [the disputants] as perfectly analogous to thermometers – as truthometers. Whatever else they are (and whatever else they are to you), they are each a mechanism with a certain probability of issuing a true ‘reading’ of theoremhood (or whatever), and the way to take their views into account – the way, that is, to revise your own views given the evidence of theirs – is exactly the way to take the reading of thermometers into account.’ (Enoch 2010: 960).

Schafer is less explicit on this point but frames his whole discussion in the language of probability and reliability judgements – thus, I take it that something like CO is part and parcel of his approach to disagreement.

epistemic advantage. Yet, CO and, consequently, NA say only that one's response to the disagreement must be coherent with what one is justified in believing about the disputant's comparative epistemic credentials. Neither principle, though, says anything about what kinds of reason – dispute-independent or otherwise – are relevant to determining that. Correspondingly, neither CO nor NA entails any particular answer to the question of how one ought to respond to disagreement with an effective peer.

What Enoch and Schaffer both draw attention to, then, is that – even if we accept CO and NA – the real grist for the mill is the question of what effective peers are justified in believing about their respective credentials *upon* realising that they disagree. So, Schaffer suggests:

[M]uch of the action in the epistemology of disagreement relates, not to how one should treat those one regards as “epistemic peers” in the sense popular in the literature, but rather to whom one should treat as such. (Schaffer 2015: 26).<sup>37</sup>

And, in similar spirit, Enoch states:

In the context of finding out how we should respond to peer disagreement, it is a key issue whether the disagreement itself can be sufficient reason to demote your interlocutor from the status of peer (Enoch 2010: 476).

It is in respect to this question that each (in effect) argues that a default *entitlement* to self-trust can underwrite a steadfast response to disagreement with an effective peer. There are enough similarities between their accounts that I think we can reconstruct this line of argument in general form without doing disservice to either. So, let's do that.

As per the discussion so far, the starting point of this argument is CO, to reiterate:

- (1) If S realises that  $S_1$  believes Q and  $S_2$  believes  $\sim Q$ , S will be justified in believing that Q is more likely to be true than false *only if* S is justified in believing that  $S_1$  has the relative epistemic advantage over  $S_2$  vis-a-vis Q, and vice-versa.

Of course, any beliefs that S forms about the comparative credentials and standing of  $S_1$  and  $S_2$  vis-à-vis Q will be 'her' beliefs, thus, given DST we can also add:

---

<sup>37</sup> Schaffer is using 'peer' to specify the situation in which one is justified in believing that the person one disagrees with is one's epistemic equal (i.e. the situation described in NA) Echoing the points about effective peerhood above, this is a yet stronger conception of peerhood than what I referred to as Justificatory Peerhood in Chapter 1.

- (2) S cannot determine whether either of  $S_1$  and  $S_2$  have the relative epistemic advantage without relying upon her own belief forming practices and faculties.
- (3) S is *entitled* to rely on S's own basic belief forming practices and faculties when it comes to determining whether either of  $S_1$  and  $S_2$  have the relative epistemic advantage.

Given (1) and (2), the relevant question in relation to the Truthometer Analogy is whether S's entitlement to rely on her belief forming practices has any special significance upon what S is justified in believing vis-à-vis the credentials of  $S_1$  and  $S_2$  – and whether the answer to that question might differ depending on what S's perspective on that disagreement is.

To help us address this question, let's suppose that we are discussing the disagreement between Quinn and Zaius over Patient-X's treatment options described in DOCTORS-1 and PATIENT-X from the previous chapters. As was stipulated in those cases, Quinn and Zaius are both highly qualified doctors who have independently assessed a significant body of clinical and scientific evidence relevant to the question of which of the treatment options is better; Patient-X on the other hand is entirely dependent upon Quinn and Zaius when it comes to that question. Furthermore, neither of Quinn and Zaius, nor Patient-X, have any evidence independent of the disagreement that indicates that Quinn or Zaius is more likely to be right on the question than the other.

Now suppose we are looking at the case from the perspective of Patient-X. In this case, it seems clear Patient-X's entitlement to rely on her own belief forming faculties has no special significance upon how she ought to respond to the disagreement between Quinn and Zaius. Crucially, Patient-X's epistemic dependence on Quinn and Zaius is no less after realising the disagreement than it was beforehand. Thus, whatever practice or practices Patient-X might employ in assessing the credentials of Quinn and Zaius, *her* continued entitlement rely on her own belief-forming practices after realising the disagreement will not grant Patient-X access to any information she did not have about the respective credentials of Quinn and Zaius beforehand. Since, then, Patient-X was not justified in believing that either Quinn or Zaius had the advantage beforehand, she is not justified in believing differently after realising that they disagree about those options. Thus, in line with NA, Patient-X, ought to suspend judgement on which of option-A or option-B is better. In other words, the analogy drawn by the conciliationist between the perspective of Baker in the thermometer case and the third-party perspective on disagreement holds even if we grant that DST is true.

Suppose now that one takes the perspective of Quinn and so, in doing so, takes a 1<sup>st</sup> person perspective on the disagreement. In this case, so long as the entitlement referred to in (3) holds, things would appear to play out quite differently. For, given that we are taking the perspective of Quinn, (2) and (3) can be transposed into:

- (4) Quinn cannot determine whether she or Zaius have the relative epistemic advantage without relying upon her basic belief forming practices and faculties.

And, crucially:

- (5) Quinn is *entitled* to rely on her basic belief forming practices and faculties when it comes to determining whether she or Zaius have the relative epistemic advantage.

In other words, as a first-person participant in the disagreement, Quinn is entitled to rely on her own belief forming practices and faculties when it comes to determining whether she or Zaius have the relative epistemic advantage vis-à-vis the treatment options. And, if that is the case, or so the argument goes, Quinn would seem to have a bias in her favour built into her first-person perspective on the disagreement.

How so?

Schafer (2015: 31-34) illustrates by way of a simple analogy. Imagine that we have the papers of two students who have taken an exam consisting of 100 yes/no questions. For any given question, we know the first paper is 95% likely to have the right answer. We don't know how reliable the answers on the second paper are, but we know that 95 of the answers on the paper are the same as the first, and 5 are different. In this case, Schaffer argues, since we have no other way of judging paper-2 and since we know that the answers on paper-1 are reliable guides to the truth, we are entitled to rely on the answers on paper-1 in judging the answer on paper-2. Thus, where paper-2 agrees with paper-1, we can take it that there is a 95% chance the answer on paper-2 is right; and where paper-2 differs, we can take it that there is a 5% chance the answer is right. And so, by summing the probabilities, we can conclude that, for any given answer on paper-2, that answer is 90.5% likely to be right. I.e. Slightly less likely to be right than any given answer on the paper-1.

On Schafer's account, cases such as Quinn's – i.e. cases where one realises that one disagrees with an effective peer – are no different. Quinn is entitled to believe that her basic belief forming practices and faculties are reliable, and she can have no standard by which to judge Zaius's epistemic credentials other than the deliverances of those practices. Just as in the exam case, so



the argument goes, Quinn can and should count any deviation from that standard on Zaius's part as evidence against his reliability. Or, at least she should do if she does not acquire any defeaters for her entitlement to rely on her own basic practices and faculties.

Thus, given that:

- (6) Zaius's beliefs about the treatment options deviate from the deliverances of Quinn's basic belief forming faculties and practices.

And, if we assume that the entitlement to rely on one's practices does hold, it follows that:

- (7) Quinn has evidence that Zaius is less reliable than she is.

And so,

- (8) Quinn has evidence that she has the relative epistemic advantage over Zaius.

In which case, even on the presumption that CO is correct, Quinn would not be in contravention of that principle were she to continue to think that it is more likely that 'A is best' than that 'B is best'.

Along these lines, Schafer endorses the following view of disagreement:

Rational Symmetry View\*: Two individuals who are equally aware of each other's evidence, and neither of whom have any special reason to treat themselves as less (or more) reliable than the other, have a disagreement about some issue, both should give more weight to their own opinion than they do to the others. (Schafer 2015: 38).

And Enoch, echoing the thought, claims:

Often, then, in cases of peer disagreement your way of maintaining (or retaining) probabilistic coherence will be by simultaneously reducing your confidence in the controversial claim and in the reliability of both you and your (supposed) peer, *though reducing it more sharply regarding your (supposed) peer...* (Enoch 2010: 993. Italics Added).

Whether or not the evidence that the realisation of disagreement (on this argument) affords S is sufficient to justify outright steadfastness remains an open question on this argument. So long, though, as the disagreement is between effective peers, that evidence will at least tilt the balance in S's favour – even if only a little. And, in that respect, it seems fair to treat this argument as a defence of the steadfast view of disagreement in spirit, if not entirely in substance.

How good is this defence?

Most obviously in its favour, by allowing that by *default* one has reason to favour one's original position in the case of disagreement with an effective peer, the upshot of the Default Self-Trust View is that in the core case – single disagreement with effective peer – one can be steadfast. Thus, this approach avoids the dispiriting sceptical consequences that we saw in the last chapter to beset Conciliationism. Notably, and although it was proffered as an argument against the Truthometer Analogy, the Default Self-Trust View is directly opposed to the principle of Independence. Nonetheless, as described this approach avoids licensing the kind of dogmatism that we saw in the previous chapter to motivate Independence. As we saw, the conciliationists concern was that any view permitting one to treat the disagreement itself as evidence against one's interlocutor's position, will license steadfastness in all cases of disagreement – not just those with an effective peer. At least as it is stated here, however, the Default Self-Trust View does not have this consequence. As noted, the conclusion of the argument is only that the fact of disagreement can be counted as evidence against the reliability of one's interlocutor – not that that evidence is necessarily sufficient to justify steadfastness. But if that is all the argument entails, then, it remains possible that one may have other evidence about the disagreement such that, on balance, one ought to give equal – or indeed greater – weight to one's interlocutor's views when responding to the disagreement. Cases where we might expect this to be so, include, when one disagrees with a number of one's peers and has evidence that they reached their views independently of one another; and when one has evidence, independently of the disagreement, that the person one disagrees with is one's epistemic *superior*. (I.e. Just the kind of cases that were invoked to motivate Independence.)

Despite these upshots, I think that there are two core problems with the Default Self-Trust View.

Firstly, as I noted earlier, entitlements – including any entitlement to self-trust – are defeasible. The relevant question, then, is not whether the argument at the heart of the Default Self-Trust View is valid – but whether the entitlement that underwrites one's access to that argument still holds in the context of disagreement. If it does not, then, clearly one cannot acquire any evidence against one's interlocutor by way of that entitlement.

In this respect, it is reasonable to suppose that an agent's entitlement to rely on their belief-forming practices *will hold* on realising that they disagree with someone they have *no reason* to believe is competent on the matter. For instance, if Doctor Quinn were on a train giving emergency treatment to the driver, and a random passenger were to exclaim that he disagreed with her diagnosis, it hardly seems that she would need to reduce her confidence in her judgement at that moment. But that is just because the kinds of background assumption one might make about a

random stranger on a train will not be sufficient to justify believing that they have the kind of medical expertise required in the situation. For that, to borrow Schafer's phrase, Quinn would have to have *special reason* to believe that the random passenger was competent in the relevant ways. But, that being so, someone who disagrees with Quinn about medical matters will not be Quinn's effective peer *unless* Quinn has *special reason* to believe that they have the relevant kinds of medical expertise. What is true about the train-case is illustrative of a general feature of effective peerhood. For whilst effective peerhood does not entail that one is justified in believing that one's interlocutors are one's epistemic equals – it does entail that one is justified in believing that the other parties to the disagreement are *generally competent* in the relevant ways.<sup>38</sup>

The problem that this presents for the Default Self-Trust View can be illustrated by returning to Schafer's exam analogy. As far as it goes, Schafer's assessment of the exam case as presented seems to be spot-on. However, what drives that intuition, it seems to me, is the thought that paper-1's 95% chance of being right on any question is exceptional. For, if it is appropriate to presume that paper-1 is exceptional, then, it would clearly be apt to rely upon the answers on that paper when judging the answers on paper-2 – at least in lieu of any evidence that paper-2 is also exceptional. The exam case, thus, is analogous to the situation described in the case of Quinn and the stranger on the train. The analogy does not hold, though, when the person that Quinn disagrees with is not a random stranger but, as in DOCTORS-1, a highly respected colleague. Instead, an exam case that would be analogous to DOCTORS-1 would be one in which:

- A. We know that any given answer on *either* paper is more likely to be correct than not.
- B. We have no reason to suppose that a 95% chance of being right on any given question is particularly low or high for the sample group.
- C. We have no other reason to think that the answers on paper-1 are any more likely to be right than the answers on paper-2

But, in this kind of case it seems to me that it would be entirely inappropriate to assess paper-2 on the basis of the answers on paper-1 – even knowing that any give answer on paper-1 is 95% likely to be right. Rather, once we have dropped the presumption that paper-1 is an exceptional exam script, and without other evidence that the answers on paper-1 are likely to be any better than

---

<sup>38</sup> Presuming, that is, that the agent in question ought not to believe that they are also *incompetent* in the relevant ways. Presuming also that we are discussing cases of ordinary disagreement. As discussed in Chapter 1, the conditions of effective peerhood in case of deep disagreement differ in significant ways.

those on paper-2, we ought to be no more confident that paper-1 has the correct answer on any given question than we are that paper-2 has the correct answer. This, of course, is analogous to what the conciliationist says about the irrelevance of the first-person perspective on disagreement. Thus, if I am right that this revised case is a better analogy for cases of disagreement between effective peers than Schafer's original example, we have good reason to think that the entitlement to self-trust will not hold in just those cases that give rise to 'The Problem of Ordinary Disagreement'.

Now, at this point, Enoch or Schafer might refer us back to the descriptive aspect of the Default Self-Trust thesis. For, isn't the point made there that, whatever additional information might be in play in case of disagreement, disputant's can only process that information by exercising their own belief forming faculties and practices. After all, if Quinn is to factor her awareness of Zaius's qualifications into her response to the disagreement in DOCTORS-1 – she still has to rely upon some or other of her belief forming practices to do so. Thus, when it comes down to it, if there is to be any reasonable response to disagreement, isn't it still the case that one will have to rely on at least some or other of one's belief forming practices to arrive at that response?

This point seems fair but leads straight to the second objection to the Default Self-Trust View. That is, even admitting the inevitability of one's reliance on one's own belief forming practices, we can simply reiterate the original problem by asking *which* of one's belief forming practices one should rely upon? Take Quinn: Should she continue to act as if the practices that led her to believe 'A is best' are reliable; should she act as if practices that led her to believe that 'Zaius is a highly qualified doctor' are reliable; or, indeed, should she continue to act as if the practices that led her to believe that 'Zaius and I disagree' are reliable? And so on. Crucially, whichever of these practices Quinn does rely upon – given CO – she ought also to accept that at least one of the others has let her down in this instance. In other words, however Quinn responds to the disagreement – she will have to 'trust' one of her belief forming practices over others. The relevant question from the first-person perspective is which of those practices that is. The thesis of Default Self-Trust, however, has nothing to say upon this question. And, that being so, it is hard to see how the supposed entitlement to epistemic self-trust can support any default response to disagreement with an effective peer.

This last issue, we might note, is not exclusive to the Default Self-Trust View. Rather, if we presume that something like CO is true, then, any disagreement with an effective peer will require some of one's beliefs to be revised, it is just a question of which beliefs those are. The conciliationist says that the answer to that question is that one ought always to revise one's beliefs

about the propositions under dispute. Any alternative to Conciliationism has to explain how, at least sometimes, it is better to revise some other of the relevant beliefs. Though neither is framed in terms of self-trust, Lackey and Kelly's non-uniform accounts can be understood as attempts to explain just that by reference to features of disagreement that we have not yet discussed in detail: Respectively: the content of the disputed beliefs, and the different kinds of evidence one comes to possess in case of disagreement. Without further ado, then, let us turn to discuss these accounts.

## 3.2 The Justificationist View of Disagreement

Jennifer Lackey (2010) offers an explicitly 'non-uniform' account of disagreement that she labels the 'Justificationist View' of disagreement. Per the Justificationist View, in some cases of disagreement between effective peers, one may be required to revise one's beliefs toward the disputed propositions, in others one may be permitted to retain those beliefs. Cases in which steadfastness is permitted, or as Lackey calls it 'non-conformism', are cases in which disputants have an exceptionally high degree of *justified confidence* in those beliefs. In some such cases, despite first appearances, one can infer a symmetry breaking explanation of the disagreement that favours one's own view and that does so independently of the substance of the disagreement. Cases in which conciliation is required, or as Lackey calls it 'conformism', are cases in which one of those conditions is not met.

Let us look at the kinds of case in which Lackey argues 'non-conformism' or steadfastness is permissible. Slightly paraphrased, here are two of Lackey's own examples:

### PERCEPTION

Lackey is eating lunch in her apartment with Edwin, one of her two long-term roomies. She asks Edwin to pass the wine to their other room-mate Estelle. Edwin replies 'Estelle isn't here today'. Prior to this surprising comment, Lackey has no reason to think that there is anything wrong with Edwin's perceptions or that he is currently in anyway cognitively impaired. Nor does she have any impression that there is any insincerity to Edwin's odd assertion.

### DIRECTIONS

Lackey and her neighbour Jack moved into the same apartment block fifteen years ago. Both are familiar with downtown Chicago and regulars at the restaurant My Thai. One

evening Lackey bumps into Jack and informs him that she is headed to Michigan Avenue to dine at My Thai. Jack replies ‘But My Thai is on State Street’<sup>39</sup>

In both of these cases, Lackey claims, she is justified in retaining her original belief. And what grounds her doing so is the exceptionally high degree of justified confidence that she has for her initial belief in each case. Given that justified confidence, so the thought goes, Lackey can take her colleague/friend’s disagreeing with her as evidence that ‘he is not well, either evidentially or cognitively’ (Lackey 2010: 309). In other words, Lackey suggests that in these cases she is permitted to treat the disagreement as evidence that she has the relative epistemic advantage – even though her interlocutor’s in each case are her effective peers.

Lackey’s position clearly has echoes of the Default Self-Trust View. In place of trust in her belief forming practices, Lackey leans upon the notion of justified confidence – but the role that that plays in her response to the disagreement seems, at first blush to be quite similar. Despite the surface similarities, however, there are two crucial differences between the two arguments.

First of all, Lackey does not claim that justified confidence can underwrite one’s treating the disagreement as evidence against an effective peer in all such disagreements. Rather, her having sufficiently high degree of justified confidence in the beliefs ‘Estelle is sat at the table with us’ and ‘My Thai is on Michigan’ to do that, is a feature of the specific content of those beliefs as they relate to the wider background of other things she believes, evidence she has, the environment she is in, the belief forming practices and faculties she brings to the table, and so on. The crux of all that is that, given her general epistemic situation, both propositions are ones that Lackey would be extremely unlikely to *falsely believe* if her basic belief forming practices and faculties were functioning at all properly.

Secondly, Lackey makes it clear that her justified confidence does not in its own right ground her retaining her original belief in either case. What she still requires is some kind of ‘symmetry breaker’ that is specific to the case at hand. So, Lackey explains:

A symmetry breaker is something that indicates that the epistemic position of one of the parties to the disagreement in question is superior to the other’s...there needs to be a symmetry breaker that justifies my nonconformity. Otherwise, my resistance to doxastic revision is little more than dogmatic egotism. (Lackey 2010: 309).

This of course should sound quite familiar; in effect, it is the same thought that motivates Independence. For where on the Default Self-Trust View you can downgrade your assessment of

---

<sup>39</sup> For Lackey’s original formulation of these examples see her (2010: 308-09).

an effective peer in case of disagreement simply because they disagree with you – Lackey’s suggestion here is that you can do so only if you have a non-question begging argument to the effect that you have the relative epistemic advantage over your peer.

What Lackey suggests is special about cases like PERCEPTION and DIRECTIONS, then, is that her justified confidence in these cases – though not itself a symmetry breaker – can serve to make salient a class of information that she possess about herself, and not about Edwin or Jack, and which can serve as such. That is, *personal information* about the state of her cognitive faculties. So, for instance, in normal circumstances, Lackey knows whether she has been drinking or not, if she has had a good night’s sleep, if she has a recent history of hallucination or memory lapses, and so on. With this information in hand, Lackey can infer an explanation of the disagreement that indicates that she has the relative epistemic advantage. The argument goes something like this:

In each case, one party or other must be mistaken. At the same time, both disagreements are ones in which is it extraordinarily unlikely that either party would be mistaken unless they were somehow seriously cognitively impaired at the moment of disagreement. Lackey has direct access to personal information that provides her with evidence that she is not so impaired. She does not have access to similar information about her interlocutor in either case. On account of this asymmetry, Lackey infers that the best explanation of the disagreement is that her opponent is somehow cognitively impaired. If her opponent is cognitively impaired, and she is not, Lackey has a clear epistemic advantage – thus, if she is justified in believing that that is the case, Lackey is under no rational pressure to reduce her confidence in her beliefs as regards the disputed propositions.

In light of her general account, let us call this the ‘Justificationist Argument for Steadfastness’ – or JAS for short. As Lackey summarises this line of thinking:

An instance of ordinary disagreement regarding the question whether p provides me with a defeater for my belief that p. When I am very highly justified in holding this belief, the personal information that I possess about myself and lack about my interlocutor can provide me with a defeater-defeater. And so long as I do not then acquire a defeater-defeater-defeater, I am thereby permitted to rationally retain my belief that p with the same degree of credence. (Lackey 2010: 310).

The first thing to emphasis here is that this argument respects the principle of Independence in a way that the Default Self-Trust View does not. Whilst it is Lackey’s justified confidence that underwrites her inferring a defeater-defeater, it is the personal information that she has about the state of her own cognitive faculties that grounds that inference – and she comes to possess that

information independently of the substance of the disagreement. In this way, we might say, steadfastness via JAS does not so much contravene the Independence principle, as circumvent it.

The second thing to emphasise is that JAS is only available in cases in which one both has an exceptionally high degree of justified confidence *and* the personal information one has about the state of one's cognitive faculties is sufficient to rule out alternative explanations of the disagreement. So, for instance, if in DIRECTIONS Lackey's opinion about the whereabouts of My Thai was not based on personal experience, but Google Maps, she could not rely upon the personal information about her own cognitive faculties to rule out the possibility that Google Maps was in error – even if she initially had an exceptionally high degree of justified confidence in its accuracy. Clearly enough, then, both of these conditions will not apply in *all* cases of peer disagreement. Indeed, this is just what makes Lackey's account a non-uniform account of disagreement.

The crucial question, then, is just how widely JAS is available. This question becomes more pressing if we keep in mind the Spinelessness objection to Conciliationism. For given Lackey's suggestion that conciliation/non-conformism will be the correct response to disagreement with an effective peer in any and all cases where JAS is not available – much of the appeal of the Lackey's account will depend upon how well it contains the sceptical consequences associated with widespread conciliation.

Unfortunately for the account, JAS does not generalize widely at all. To see why, consider Lackey's response to the classic restaurant case (i.e. DINERS-1). In this case, Lackey claims, the JAS gets no traction and conciliation is required. Drawing on the previous cases she cites three key differences between this case and those like PERCEPTIONS and DIRECTIONS that explain this result:

1. The content of the disagreement in the cases in which the JSA is available make it very likely that one or the other disputants is cognitively impaired. Not so for Jack and Jill in DINERS-1.
2. Jack and Jill in DINERS-1 may be confident in their conclusions, but they ought not to have such a high degree of confidence as Lackey's in cases PERCEPTIONS and DIRECTIONS. Why? Because, Jack and Jill are sufficiently fallible in the circumstances described that, even if there were no disagreement, it is a non-negligible possibility that they have made a mistake.



3. If either Jack or Jill does have such a high degree of confidence as Lackey's, this will not be justified – again, Jack and Jill are sufficiently fallible in the circumstances described that they should recognise the non-negligible possibility that they have made a mistake.

I think Lackey is quite right here – if a disagreement has any of these features, then the JSA will be a non-starter. And, once again, the key feature is the content of the beliefs under dispute. Whilst, say, Jill might be able to rule out more egregious forms of cognitive impairment, it isn't the case she would be *exceptionally* unlikely to believe that the diners each owe \$43 if she was not seriously cognitively impaired. For instance, she might have made a mistake because she was just a little more tired than she thought, or because she rushed her sums, or because that last glass of wine was half a glass too many for mathematical funny games, or, indeed, just because mistakes are easy to make when doing mathematics. These would all be good explanations of this disagreement – and would all be non-negligible possibilities even if there were no disagreement. Yet, judgements of our own cognitive functioning just don't come with the fidelity required for Jill to rule these explanations out on the basis of the personal information she has about the state of her cognitive faculties. And, if Jill cannot rule out explanations of the disagreement that would put her at fault she cannot infer that it is Jack who is impaired. In which case, in Lackey's terms, Jill does not gain a defeater-defeater for the realisation that her effective peer disagrees with her about the restaurant bill.

So, Lackey seems right about the kinds of conditions that will preclude one's access to the JAS in case of disagreement. The problem for her account more generally, is that at least one of these conditions is going to be met in the great majority of disagreements. For, whilst I might be extremely unlikely to believe, unless I am seriously and obviously impaired, that someone is sat in front of me when they are not, or that my favourite restaurant is on a completely different street to the one that it is on – it is not that unlikely that I will make some mistake when forming beliefs about politics, economics, philosophy, the time the movie is on, the street a *new* restaurant is on, and so on... even when my cognitive faculties are operating within standard margins of error.

If though, JAS is not available in a great majority of disagreements and conciliation is the correct response in those cases where it is not – then it seems that the Justificationist View will have almost as wide, and unpalatable, sceptical consequences as we saw Conciliationism to have in the previous chapter. In so far as the Spinelessness objection was motivation to look elsewhere than Conciliationism for a satisfactory account of ordinary disagreement, so too is it motivation to look elsewhere than the Justificationist View.

With that said, let us move on to Thomas Kelly's Justificationist View.

## 3.3 Total Evidence

Like Lackey's Justificationist View, Thomas Kelly's account of disagreement is a non-uniform view. According to Kelly, the conciliationist is quite right to argue that, if the principle of Independence is true, Conciliationism is the correct view of ordinary disagreement (Kelly 2013: 38). But Independence is not true. What determines the appropriate response to disagreement, according to Kelly, is not the evidence that one has independent of the substance of the disagreement, but the *total evidence* one has about the disagreement and the context of that disagreement. This includes both the *first-order evidence* upon which one originally reasoned, and which bears directly upon the disputed issues, and the *higher-order evidence* that one acquires on realising that someone disagrees with one on those issues.

Before turning to the positive component of Kelly's account, it will be useful to rehearse the case he makes against Independence.

### 3.3.1 Against Independence

To remind ourselves, we formulated the principle of Independence as follows:

#### *Independence*

If  $S_1$  realises that  $S_2$  disagrees with her about  $Q$ ,  $S_1$ 's attitude toward  $Q$  (whatever that attitude might be) will be justified *only if* it is based upon justification that is independent of the substance of the disagreement.

As we saw in the previous chapter, the motivation for this principle is that it rules out certain clearly dogmatic arguments to dismiss the potential significance of disagreement. Arguments such as:

#### OVERCONFIDENT

Ten similarly competent individuals who share the same evidence as I have all independently come to believe  $\sim Q$ . But the evidence supports  $Q$ . Therefore,  $Q$ . Therefore, my interlocutors have all made a mistake in their assessment of the bearing of the evidence on  $Q$ .

As noted in the previous chapter, however we theorise its shortcomings, an argument such as OVERCONFIDENT is paradigmatically dogmatic – at least if there are no other relevant

epistemic considerations in play – and clearly an inappropriate response to the realisation of disagreement. At least as far as this argument goes, Kelly would not disagree. Nor, I think, would he disagree with the more general idea that beliefs maintained through dogmatic responses to disagreement will fail to be justified (no matter whether they were justified beforehand). What Kelly does suggest, as I shall explain, is that by invoking Independence to rule out obviously problematic arguments like this we also lose the resources to accommodate cases in which the *right* response to disagreement cannot but involve a question-begging response. Thus, if there are such cases, Independence must be false.

The conciliationist's error lies, for Kelly, in treating dogmatism as a 'formal epistemic vice' (Kelly 2013: 43). A formal epistemic vice, he explains, is an epistemic failing of which we can identify instances without having to make *substantive* judgements about the relevant epistemic features of particular cases. To explain formal epistemic vices, we need special epistemic principles that describe the conditions under which the vice is instantiated. Once we have such principles, though, we need only to check the details of any particular case against those conditions to see if the vice in question is instantiated. Independence, Kelly points out, is a perfectly good explanation of dogmatism, *if* dogmatism is the kind of failing that we can identify independently of making substantive judgements about the details of particular cases. Kelly, however, denies that we can identify examples of dogmatism in this way. To illustrate his point, he offers the following example:

[C]onsider the status of so-called "Moorean" responses to revisionary philosophical theorizing. You present me with certain metaphysical responses that you endorse and the arguments for them entail that there are no tables and chairs [...] I consider your principles and supporting arguments. I then reason as follows: "Well, the principles seem quite plausible, and I don't have an intellectually satisfying diagnosis of where the arguments go wrong (if in fact they do). But of course, there are tables and chairs. Therefore, the theory is false (Kelly 2013: 44).

The Moorean argument here has a similar structure to the paradigmatically dogmatic argument above. For the Moorean admits the sceptical arguments to the conclusion that there are no tables and chairs are intellectually persuasive and that he has no independent reason to reject those arguments, yet, on the basis that tables and chairs do exist, he concludes that his friend's thesis must be false nonetheless: even though he cannot take for granted the existence of tables and chairs without thereby presupposing the falsity of the sceptical argument. Question begging though this response is, Kelly maintains that it is the only reasonable response for the Moorean to his experiential evidence of tables and chairs – even in conjunction with his friends otherwise convincing metaphysical arguments. If dogmatism is the formal vice explained by Independence, however, this defence of common-sense will count as no less dogmatic than the paradigmatically

dogmatic response to disagreement that is OVERCONFIDENT. If, on the other hand, Kelly is right about this case and the Moorean's response is, as the only reasonable response to the disagreement, not dogmatic – then Independence is not true – or at least is in need of revision.

Moreover, even if we disagree with Kelly that this Moorean response is the *only* reasonable response to the conjunction of sceptical arguments and empirical evidence, it is not clear that we should conflate this argument with the paradigmatically dogmatic argument at OVERCONFIDENT. After all, as OVERCONFIDENT was described, I dismiss the possibility that my peers might be right about Q out of hand. In Kelly's example, on the other hand, the Moorean considers the various arguments his friend offers, admits their plausibility, weighs these considerations up against the evidence he has from his senses, and only then concludes that his friend must have gone awry somewhere in his arguments. In character at least, these differences would seem to represent a marked difference between the two responses are. So we might think, even if the Moorean response is dogmatic it is less dogmatic than the response described in OVERCONFIDENT. Again, though, if we treat dogmatism as the formal vice that corresponds to contravention of Independence, we cannot accommodate that line of thought: Whilst we might say informally that the Moorean's response is less dogmatic than mine, in theoretical terms, we would have to conclude that there are no significant differences between the two cases.<sup>40</sup>

Rather than attempt to accommodate these considerations by somehow tweaking Independence, Kelly's suggestion is that we should simply conclude that dogmatism is not a formal vice. And, if dogmatism is not a formal vice, then, we do not need to posit some special epistemic principle to explain it. And, if that is the case, the conciliationist loses the primary motivation for positing the principle (i.e. Independence) that we saw in the previous chapter to form the backbone of that view. Instead, Kelly argues that what we need to consider when judging the appropriateness of an agent's response to the realisation of disagreement – including the question of whether that response is dogmatic or not – is not whether that response violates Independence but whether it is the response that a substantive judgement of the agent's total evidence would reveal to be a reasonable response to that evidence. This brings us to Kelly's 'Total Evidence View' of disagreement.

---

<sup>40</sup> See Batally (2018b, 2018a) for a virtue theoretic account of dogmatism that would better accommodate these kinds of considerations than the principle-based understanding implied by Independence.

### 3.3.2 The Total Evidence View

Kelly distinguishes between two kinds of evidence that, combined, make up an agent's total evidence in respect to a given proposition: *first-order evidence* and *higher-order evidence*. Where an agent's first-order evidence for Q is the evidence that they have which bears directly upon the truth of Q; their higher-order evidence is evidence that bears in some way upon the justificatory relationship between their first-order evidence and their beliefs about Q. What that agent is justified in believing about Q, according to the Total Evidence View, is what they are justified in believing having taken into account both first-order and higher-order evidence. In situations where the agent realises that someone disagrees with them about Q, their higher-order evidence will include their evidence of disagreement, their evidence about the context of that disagreement, and their evidence about the epistemic credentials of the people they disagree with.

Crucially, when the higher-order evidence one possesses counts against one's beliefs it can, to borrow Kelly's phrase, exert a 'downwards epistemic push' (Kelly 2010: 159) on one's justification for those beliefs, as Kelly explains:

Rationality consists in responding appropriately to one's evidence. But one's evidence includes evidence to the effect that one does not always respond appropriately to one's evidence... as well as evidence to the effect that one is more likely to have responded inappropriately when one finds oneself in certain circumstances. When one possesses higher-order evidence to the effect that one is currently in circumstances in which one is more likely than usual to have made a mistake in responding to one's first-order evidence, one has a reason to temper one's confidence – even if that confidence is in fact an impeccable response to the first-order evidence (Kelly 2010: 139).

Adding shortly after:

[C]ases in which one in fact responds impeccably to one's evidence, but one's peer responds inappropriately, are much like cases in which one engages in a flawless piece of practical reasoning despite being inebriated. The fact that a peer has responded to the evidence differently should lead one to temper one's confidence in one's own response, just as the fact that one is inebriated should lead one to temper one's confidence in the conclusion of one's practical reasoning, despite the actual flawlessness of one's performance (Kelly 2010: 140).

Whilst the higher-order evidence acquired when an agent realises that their peer or effective peer disagrees with them about Q can lower the support that their beliefs have in respect to their total evidence however, it does not, in Kelly's words 'make all the difference' (Kelly 2010: 141). After all, even having acquired that higher-order evidence, the agent's total evidence will still include their original first-order evidence. And, if they were initially justified in believing Q on the basis of that first-order evidence, their total evidence will still contain a proportion of evidence that supports their believing Q. Or so it will do on Kelly's account. What the agent is justified upon

discovering the disagreement, then, according to the Total Evidence View, will depend upon the *proportion* of their evidence vis-à-vis Q that is first-order evidence – and how strong the support from that first-order evidence is for their believing Q – and the *proportion* of their evidence that is higher-order evidence – and how strongly that higher-order evidence counts against their believing Q.<sup>41</sup> Importantly, Kelly suggests, on this view there will be no satisfactory and properly general answer to the problem of ordinary peer disagreement. In some cases, one’s first-order evidence will be substantial enough to swamp the higher-order evidence acquired on learning of disagreement – and so steadfastness will be the right response to disagreement. In other cases, the balance of first-order and higher-order evidence will be about equal – and so one ought to suspend judgement on the disputed issue, as per Conciliationism. In other cases, the balance of first-order and higher-order evidence will fall somewhere between these two situations, and the appropriate response to one’s total evidence be to lower one’s confidence in one’s original beliefs, but not so far as to suspend judgement altogether. Finally, in some cases, the higher-order evidence will be substantial enough to swamp the first-order evidence entirely; as Kelly, echoing the discussion of multi-party disagreement in the previous chapter, explains here:

Significantly the point generalizes beyond the two-person case. As more and more peers weigh in on a given issue, the proportion of the total evidence that consists of higher-order psychological evidence increases, and the proportion of the total evidence that consists of first-order evidence decreases. [...] when the number of peers grows large enough, the higher-order psychological evidence will swamp the first-order evidence into virtual insignificance (Kelly 2010: 144).<sup>42</sup>

What should we make of this approach to disagreement?

In its favour, at least if we accept Kelly’s claims about the normative consequences of the view, the Total Evidence View will not be subject to the same kind of sceptical consequences as Conciliationism. As discussed in the previous chapter, the sceptical consequences of Conciliationism are problematic only in so far as a wide enough range of real-world disagreements meet the conditions under which that view requires disputants to suspend or forego judgement on the issue under dispute. Since the Total Evidence View allows the possibility that the reasonable response to disagreement with one’s peer can, in some cases, be to suspend judgement – the Total

---

<sup>41</sup> It is worth noting that higher-order evidence can cut both ways – one might, for instance, have higher-order evidence that one is *less* likely than usual to have made a mistake in responding to one’s first-order evidence. However, since my focus is upon the evidence one acquires on learning of disagreement, unless stated otherwise, references to higher-order evidence in the text should be taken to refer to higher-order evidence against one’s beliefs.

<sup>42</sup> Interestingly, Kelly also accepts something like the diagnosis of multi-party, multi-position disagreements that I suggested in the previous chapter that the conciliationist is committed to (see Kelly 2010: 141 and Kelly 2014: 306-7).

Evidence View is also susceptible to the possibility that a wide range of disagreements in the actual world will require significant revision. Thus, just as Conciliationism does not in principle lead to scepticism, so, the Total Evidence View does not in principle evade scepticism. However, given the fact that we *are* sensitive to evidence and form beliefs on the first-order evidence, it would seem to follow that one could not be systematically mistaken in one's assessment of that evidence unless there were the kind of fault line in one's evidence described by sceptical hypotheses. Yet, the possibility of disagreeing with one's peers is not one of these sceptical hypotheses.<sup>43</sup> So, given that one forms one's beliefs in response to the first-order evidence, it is very likely that one's evidence based beliefs, by and large, will be well supported the first-order evidence. But, then, it is also very likely that one's Total Evidence will, by and large, point in the same direction. And so, following the dictates of the Total Evidence View is unlikely to have the sceptical consequences associated with Conciliationism. In a nutshell, where scepticism in the social world we live in is an inevitable consequence of Conciliationism, it is a possible but improbable consequence of the Total Evidence View.

At the same time as the Total Evidence View avoids Spinelessness; in so far as it allows that the conciliatory response *will in some cases* be the reasonable response to disagreement, it also avoids the kind of issues that we saw to face the Default Self-Trust View – namely, why we should think that the entitlement to rely on the practices by which one came to hold the disputed beliefs will always hold in case of disagreement with an effective peer. Indeed, the distinction between first-order and higher-order evidence would appear to be just the theoretical resource by which we might explain why it is sometimes appropriate to rely on one set of one's belief forming practices, and sometimes appropriate to rely on another. I.e. In cases where it is appropriate to continue to rely on the practices by which one came to believe Q, that is because one's first-order evidence swamps one's higher-order evidence; and in cases where it is appropriate to continue to rely on the practices by which one came to believe one's interlocutors are competent, that is because one's higher-order evidence swamps the first-order evidence.

Of course, both of these advantages to the Total Evidence View come at the cost of its being unable to provide the kind of guidance on how to respond to disagreement that would allow us to programmatically read off the correct response to disagreement from principle, and without considering the substance of the particular disagreement. Yet, this would hardly be the first (or last) time that our best philosophical theories have let us down in this way. Moreover, whilst the

---

<sup>43</sup> At least not independently of any particular view of disagreement.

hope of offering actionable guidance is one of the principal motivations behind the literature on the epistemology of disagreement, it is not necessarily the case that a theory needs to provide top-down, off-the-rails, guidance to have something useful to say. The exhortation to respect one's evidence is rarely not worth repeating. And by highlighting the significance of higher-order evidence, the Total Evidence View reminds us to not ignore an important subset of that evidence. Thus, by following the advice of the Total Evidence View, we might still improve our judgements of particular cases, even if the view does not itself tell us what to believe in those cases.

As presented so far, the appeal of the Total Evidence View lies in the claim that it can sail between the Scylla and Charybdis of scepticism and dogmatism. We have seen already that the view has the resources available to steer clear of the jaws of scepticism. However, what I want to suggest in the rest of the discussion is that, despite initial promise, given the way that Kelly appears to conceive the normative significance of higher-order evidence, the Total Evidence View fails to escape the pull of dogmatism. To see why, it will be useful to offer a more formal encapsulation of how Kelly appears to conceive the relationship between one's first-order and higher-order evidence.

Evidence of this can be found in Kelly's discussion of a generic example of peer disagreement in which he and the reader share a body of evidence  $E$ , yet disagree about its bearing on a theoretical statement  $H$ . As per the case, having assessed  $E$ , Kelly is quite confident that  $H$  is false; having assessed the same evidence, the reader is quite confident that  $H$  is true. As it happens, the reader is right – given  $E$ , it is reasonable to be confident that  $H$  is true. On realising that they disagree, Kelly suggests the disputants total evidence consists of  $E^*$ , where  $E^*$  includes:

- (i) The original evidence  $E$
- (ii) The fact that the reader is quite confident, given  $E$ , that  $H$  is true.
- (iii) The fact that Kelly is quite confident, given  $E$ , that  $H$  is false

As per the points above, there is not enough detail to this case to give a judgement on what Kelly or the reader is justified in believing vis-à-vis  $H$  on the basis of  $E^*$ . Nonetheless, Kelly's comments on the case are revealing as to how he conceives the relationship between higher-order and first-order evidence. So, he states:

Given that you and I are peers, it is plausible to suppose that the two pieces of higher-order psychological evidence ((ii) and (iii)) are more or less equally strong pieces of evidence that point in opposite directions. All else being equal, then, one would expect  $E^*$  to favor  $H$  over not- $H$  inasmuch as it is composed of a substantial body of evidence that strongly favors  $H$  over not- $H$ , supplemented by two additional pieces of evidence of



approximately equal strength, one of which tends to confirm H, the other of which tends to disconfirm H. (Kelly 2010: 142-43).

As Jon Matheson points out, and echoing points made earlier, the suggestion here seems to be that E\* offers less support to H than E because E\* contains a greater *proportion* of evidence that supports agnosticism than does E (Matheson 2015: 86). At the same time, though, E\* continues to support H over not-H because a substantial proportion of that evidence, E, is evidence that supports H. And, crucially, if Kelly's claim about the bearing of E\* is correct, E must continue to support H when combined with the higher-order evidence from the realisation of disagreement. Further support for this reading is offered by remarks Kelly makes later in the same paper, where he suggests that:

[I]n any case in which you hold a given belief because you recognize that this is what your evidence supports, not only is your belief a reasonable one, but you are also justified in believing an epistemic proposition to the effect that it is reasonable for you to hold that belief. Indeed, given that *recognizing that p* entails being *justified in believing that p*, you will be justified in believing that the evidence supports your view for as long as you continue to recognize that it does. (Kelly 2010: 156).<sup>44</sup>

More generally, then, the core idea behind Kelly's conception of the relationship between one's first-order and higher-order evidence seems to be that higher-order evidence has normative significance only in so far as it alters the balance of one's total evidence – but, that it has no bearing upon what one's original first-order evidence supports, nor on what one is rationally permitted to believe about what one's first-order evidence supports.<sup>45</sup> This includes the higher-order evidence acquired by those party to disagreement between effective peers. Given that line of thought, then, we might encapsulate the Total Evidence View in some more detail as follows:

*The Proportional Conception of Total Evidence (PCTE)*

If at  $t_1$ , Q is reason for S to believe Q; at  $t_2$ , S acquires higher-order evidence HOE<sub>1</sub> against her believing Q; and at  $t_3$ , S acquires higher-order evidence HOE<sub>2</sub> against her believing Q, other things being equal:

---

<sup>44</sup> Kelly makes this point in responding to the Truthometer Analogy. Kelly's suggestion is essentially that, given Uniqueness, only one party to a disagreement with shared evidence can have recognized what it is that the evidence supports – thus, only one party to a disagreement is 'justified' in believing that the evidence supports their belief. This asymmetry, he claims, casts doubt upon the Truthometer Analogy just because that analogy depends upon their being symmetry between disputants at this level. I'm not convinced that that is the case but will leave that aside for now.

<sup>45</sup> I prefer 'rationally permitted' here since it seems clear that Kelly does not mean 'justification' in the sense employed in this thesis. Nothing about the main discussion depends upon this point though.

- (i) S's total evidence at  $t_2$  ( $P \& HOE_1$ ) will provide less justificatory support for S's believing Q than S's total evidence at  $t_1$ .
- (ii) S's total evidence at  $t_3$  ( $P \& HOE_1 \& HOE_2$ ) will provide less justificatory support for S's believing Q than S's total evidence at  $t_2$ .

Given that what S is justified in believing at any time is determined by her total evidence, S will not be justified in believing Q to the same degree of confidence at  $t_2$  as she is at  $t_1$ , nor will she be justified in believing Q to the same degree of confidence at  $t_3$  as she is at  $t_2$ . Moreover, if the reading of Kelly's view given above is accurate, then, whilst the higher-order evidence that S acquires at  $t_2$  and  $t_3$  alters the bearing of S's total evidence, the evidence that S possessed prior to acquiring that higher-order evidence can be bracketed and considered discretely as providing the same support that it always had. Thus, bracketing the evidence S acquired at earlier times, the following will also be true under PCTE:

- (iii) At  $t_2$ , P will provide the same justificatory support for S's believing Q as it did at  $t_1$ .
- (iv) At  $t_3$ , ( $P \& HOE_1$ ) will provide the same justificatory support for S's believing Q as it did at  $t_2$ .

Given (iii) and (iv), then, we might say that on PCTE the higher-order evidence changes the bearing of S's total evidence whilst leaving intact the justificatory connection between S's first-order evidence and Q.

At first blush this way of conceiving the relationship between one's first-order and higher-order evidence seems quite reasonable. In particular, it would appear to accommodate the worry behind the Spinelessness objection that there is something problematic with any view that requires us to suspend judgement on many issues that we care deeply about and have invested considerable epistemic effort and resources into investigating. For in these kinds of case (for instance, if you discover that someone disagrees with you on a philosophical topic that you have spent considerable time working upon), we might suspect that at least some of the disputants will, through their efforts, have amassed a considerable body of first-order evidence in support of their view. Thus, since on PCTE that first-order evidence will continue to provide the same degree of support for the relevant parties' original beliefs even when discovering that someone disagrees with them on that topic, they would have to acquire a considerable body of higher-order evidence before the balance of their Total Evidence no longer supported their original belief.

Unfortunately, these good first impressions don't last. For, whilst PCTE easily explains the cases in which steadfastness appears to be the correct response, it is not so clear that it can account for cases in which the conciliatory response is intuitively the correct response. To see why, consider the case of DINERS-2 from the previous chapter. In this variant on the original restaurant case, recall, Jill finds out that all of the other members of the dinner party are Jill's effective peers, and that they all agree with Jack that they each owe \$45, and not \$43 as Jill believes. As it was stipulated, prior to discovering the disagreement, Jill was justified in believing that each diner owes \$43. Nonetheless, intuitively it seems that Jill should not only revise her belief that each diner owes \$43 but she should lend considerably more credence to the possibility that each diner owes \$45 than in the one-on-one case. At first blush, PCTE appears to accommodate this intuition quite neatly. For, whilst the first-order evidence will provide the same degree of support for Jill's original conclusion no matter who she disagrees with, as the number of disputants increases, so too the proportion of her total evidence that consists of higher-order evidence against her believing that increases. Recall, however, that PCTE is motivated partly by Kelly's suggestion that, when one recognizes that P is evidence for Q, one is rationally permitted to believe that 'P is evidence for Q'. Thus, given that Jill, as stipulated, *recognizes* that her calculation upon the bill total, number of diners, and tip amount is evidence that each diner owes \$43, so too Jill is rationally permitted to believe that calculation is evidence that each diner owes \$43. But, if PCTE is correct, then, even after learning of the opinions of the other diners, the first-order evidence can be bracketed and considered discretely as providing the same support that it always had. Thus, it seems that Jill will *also* be rationally permitted to believe that her calculation is evidence that each diner owes \$43 on learning of the disagreement. And, if Jill is rationally permitted to believe that her calculation supports her conclusion that each diner owes \$43, then, she has reason to believe that the higher-order evidence she acquires on learning of the disagreement is misleading – i.e. she can reason as follows: (i) we each owe \$43, (ii) the others all believe that each diner owes \$45, so (iii) they are wrong, thus, (iv) I need not give any weight to their opinions.<sup>46</sup>

Interestingly, the position I am suggesting that PCTE leads to is close to an earlier position of Kelly's, labelled by Adam Elga the 'Right Reasons View' (see Kelly 2005; Elga 2007). On the 'Right

---

<sup>46</sup> To be fair to Kelly, he does consider the possibility that encountering disagreement with a peer or peers might 'strip one of one's prior ability to recognize the bearing of the first-order evidence for what it is.' (Kelly 2010: 158) Nonetheless, Kelly is clearly unconvinced by this idea, stating that 'if you were better justified in thinking that your response was reasonable than I was in thinking that my response was reasonable, then this would break the putative higher-level symmetry and provide a basis for favoring your original belief over mine' (Kelly 2010: 157) and, shortly thereafter, 'the peer whose view more accurately reflects the evidence will typically be better justified in thinking that his view is the one that is favoured by the first-order evidence' (Kelly 2010: 160).

Reasons View’, as the name suggests, the appropriate response to disagreement with a peer will be determined primarily by objective facts about what the disputants first-order evidence supports – whilst the significance of the higher-order evidence in such cases washes out.<sup>47</sup> Typically, this view has been taken to represent a particularly robust steadfast view of disagreement. Whilst the Total Evidence View appears to be an explicit retreat from the Right Reasons View, the suggestion here is, effectively, that the way Kelly appears to conceive of the relationship between first-order and higher-order evidence in the Total Evidence View does not support that move. If this is correct, then despite the intentions, the Total Evidence View is not after all a non-uniform view of disagreement – but is committed to a particularly robust version of the steadfast view. Indeed, one that is dangerously close to mandating a policy of dogmatism in the face of disagreement.

Despite these points, I think the distinction between first-order and higher-order evidence that forms the backbone of the Total Evidence View offers a useful general schema with which to approach ‘The Problem of Ordinary Disagreement’ (and as we shall see later, also ‘The Normative Problem of Deep Disagreement’) – irrespective of whether PCTE is the correct way of conceiving the relationship between first-order and higher-order evidence. I have already alluded to how the distinction between first-order and higher-order evidence neatly dovetails with the Default Self-Trust View, as well as the concerns raised about that view. But it also lends itself naturally to Lackey’s Justificationist View. For, if we drop PCTE, we might describe those cases in which, on Lackey’s view, significant revision is required as cases in which the higher-order evidence defeats one’s first-order evidence; and the case in which revision is not required as cases in which one has positive higher-order evidence in the form of the personal information about one’s own cognitive faculties that defeats the higher-order evidence acquired by way of the disagreement. In similar spirit, a number of conciliationist thinkers have turned to the concept of higher-order evidence to supplement the Truthometer and Independence arguments (for instance Matheson 2015; Christensen 2011; Rasmussen, Steglich-Petersen, and Bjerring 2018; Feldman 2009 also leans on the concept of higher-order evidence to develop a view closer to Kelly’s Total Evidence View than his earlier conciliationist position)

On this note, I would suggest that a fruitful way forward in the discussion will be to consolidate the various theoretical approaches to disagreement under the general schema of first-order and higher-order evidence. Thus, we might approach ‘The Problem of Ordinary Disagreement’ (and later ‘The Normative Problem of Deep Disagreement’) by considering the more focused question:

---

<sup>47</sup> The main difference between the two positions appears to be the emphasis in the Total Evidence View on the proportions of one’s Total evidence that is made up, respectively, of first-order and higher-order evidence.

what is the normative significance of the higher-order evidence acquired on discovering that one disagrees with an effective peer? As Kelly notes, and as we have already seen, higher-order evidence is not confined only to contexts of disagreement but is a general and common feature of our epistemic lives. Thus, it will also be useful to consider how the higher-order evidence acquired in non-disagreement cases interacts with the first-order considerations – and how the disagreement cases differ from or resemble these other cases. By doing so, we might hope to arrive at a satisfactory normative theory of disagreement without needing to posit special epistemic principles. As per Kelly's points about the principle of Independence, I take it that this would be a desirable result.

In the next chapter, then, I turn to consider the topic of higher-order evidence in more detail.

# Chapter 4. Higher-Order Evidence and Defeaters

## Introduction

Chapter 1 introduced two problems of disagreement: ‘The Problem of Ordinary Disagreement’ and ‘The Normative Problem of Deep Disagreement’. In Chapters 2 and 3, I surveyed a number of prominent responses to ‘The Problem of Ordinary Disagreement’ that have been forwarded in the literature. On the basis of that discussion we can recognise two desiderata for a satisfactory response to that problem:

1. It should explain how disagreement can have epistemic significance without that leading to the sceptical consequences associated with Conciliationism.
2. It should explain how we can avoid those sceptical consequences without endorsing a policy of dogmatism in the face of disagreement.

My suggestion toward the end of Chapter 3 was that a fruitful approach to developing an account of disagreement that satisfies these desiderata lies in understanding the normative significance of *higher-order evidence*. Roughly speaking, higher-order evidence in respect to Q is evidence that bears upon the evidence or reasons upon which one has based one’s beliefs about Q. The evidence that you acquire when you realise an effective peer disagrees with you, so the thought goes, is a form of higher-order evidence. As such, the question of how to respond to the realisation that one disagrees with an effective peer is simply a question of what it takes to respect one’s higher-order evidence. Call this a ‘Higher-Order Approach’ to disagreement.

In this chapter, I consider how best to understand the normative significance of higher-order evidence. The chapter is in three sections. In the first sections, I offer some examples of negative higher-order evidence outside of the context of disagreement and suggest that the key to understanding the normative significance of higher-order evidence lies in understanding how, and if, it can be accommodated within a more general schema of defeaters, reasons and justification. In the second section, I lean upon John Pollock’s influential account of defeasible reasoning to introduce the relevant kinds of defeater: *rebutting* and *undercutting*. In addition, I suggest that Pollock’s definition of undercutting is insensitive to an important distinction and offer a further

distinction between *contextual undercutting defeaters* and *general undercutting defeaters*. As a result, three kinds of defeater need to be recognized. In the third section, I consider whether any of the higher-order evidence cases can be explained in terms of rebutting, contextual undercutting, or general undercutting.

## 4.1 What is higher-order evidence

Consider the following two cases:

### BROWNIES

Bones is a forensic anthropologist tasked with determining the biological sex of a decomposed body. Bones exhaustively measures, documents, and records the physical structure of the skull and pelvis, then takes a break before analysing the information collected. A little exhausted, Bones eats a pile of brownies left her by her colleague Brer. Sometime later, feeling extremely relaxed, Bones continues with the analysis part of the examination. On the basis of the information she collected earlier, Bones concludes that the deceased individual was male. She puts this in her report and sends it to her boss. A little while later, Brer explains that the secret ingredient to her delicious recipe is a liberal dose of a mind-altering drug. Peeved to hear this, Bones calls to inform her boss that she will need to conduct the analysis part of her examination again, but she won't be able to do so until the next day (when the effects of the drug will have worn off). As it happens, the deceased was male, what is more and unbeknownst to all, Bones is strongly resistant to the psychotropic agent in the recipe.

### HEAVY MUSIC

Truman is an overworked and underpaid research scientist with a heavy teaching load, a huge stack of papers to mark, and not much time to do it in. Truman puts on a playlist of heavy-industrial, noise music made by a friend of his, gets to work, and finishes the marking in record time. The next day, Truman attends a talk by an eminent neurologist about the effect of different music genres on the brain. The neurologist draws the audience's attention to new and compelling research that there are some genres of music that can, with even the slightest exposure, have permanent debilitating effects upon the listeners ability to concentrate upon certain kinds of cognitive task. Amongst the maligned genres is the genre of heavy-industrial noise music that filled up his friend's playlist, amongst the cognitive tasks purportedly affected are the kinds of thinking one does when

marking university exam papers. Reflecting back on the playlist, and with sinking heart, Truman realizes that he will have to ask someone else to regrade the papers (and that is the least of his worries!). As it happens, Truman's grading was spot-on, what is more the research the neurologist was describing was completely bogus.

That Bones might have had some dodgy dessert has no bearing on the fact as to the sex of the body on her examining table. That Truman's playlist might have compromised his ability to focus on the exam scripts has no bearing upon the facts as to the scripts' merits. Each, though, is evidence that Bones and Truman may have been at greater risk of error when judging the bearing of their original evidence than they might otherwise expect. Bones and Truman would appear, at first blush, to respect that evidence by reducing their confidence in their original conclusions.

In the previous chapter, we introduced a distinction between the *first-order evidence* on which those party to disagreement based their original, disputed, beliefs, and the *higher-order evidence* that they acquire on realising disagreement. Neither BROWNIES, nor HEAVY MUSIC are cases of disagreement – yet, a number of commentators (see below) have suggested that evidence of error acquired in cases like these is of a kind with the higher-order evidence acquired in case of disagreement. Thus, or so the thought goes, by understanding the nature and normative significance of the higher-order evidence acquired in cases like these, we might gain some traction on the problems associated with the higher-order evidence acquired on realising disagreement.

Unsurprisingly, no two characterisations of this sort of evidence are quite the same, but the common thread is that higher-order evidence – in both disagreement and non-disagreement cases – somehow pertains to the practices of the agent that led them to form a certain belief, and only indirectly to the facts. So, Christensen suggests that we call evidence that ‘constrains one’s credence on some subject-matter via bearing on the reliability of one’s thinking about that subject-matter – higher-order evidence.’ (Christensen 2016b: 397); Lasonen-Aarnio that ‘higher-order evidence... works by inducing doubts that one’s doxastic state is the result of a flawed process’ (Lasonen - Aarnio 2014: 314); and King that ‘higher-order evidence... is, evidence about our grounds for belief, our dispositions for responding rationally to those grounds, and our performance in responding to those grounds.’ (King 2016: 133)<sup>48</sup>

---

<sup>48</sup> As noted in the previous chapter, whilst higher-order evidence can be both negative and positive, the literature has tended to focus on negative higher-order evidence. In that light, unless otherwise made clear, the reader should take further reference to higher-order evidence to refer to evidence that indicates a flaw or failing in the agent's epistemic practices.



In cases like those above, the higher-order evidence appears to *defeat* the support that the individual in question would otherwise have for his or her belief. If the higher-order evidence does not bear directly upon the facts, however, there is a question as to just how this defeat occurs.

To start with, it will be useful to discuss the kinds of defeater that are relevant to the discussion and how they relate to the general schema of reasons and justification around which this thesis has been framed.

## 4.2 Forms of defeat

Following John Pollock's influential work on defeasible reasoning, we might describe what it is for something to be an epistemic defeater as follows.

### **Defeater**

If P is a reason for S to believe Q, R is a defeater for this reason if and only if R is logically consistent with P and (P&R) is not a reason for S to believe Q (Pollock 1986: 38).

On Pollock's definition, defeaters defeat one's reasons for belief. This definition, thus, fits neatly with the schema of justification introduced at the top of this thesis. To reiterate the key aspects of that, let's use *justification* to refer to the kind of epistemic support an agent has for believing Q when they have reflective access to an argument with Q its conclusion. Justification, so understood, is the kind of epistemic support that corresponds to having reasons for one's beliefs. Correspondingly, then, let's say that if P is a reason for S to believe Q, provided that there is no undefeated defeater R, S would be justified in believing Q, were S to believe Q on the basis of P. Further to that, to say that P justifies S believing Q only given no undefeated defeater, is to say that P is a *prima facie* reason. To echo a point made by Lasonen-Aarnio (2014: 314-15), if higher-order evidence can have defeating force, it may be that many more reasons are *prima facie* reasons than we might otherwise have thought.

On the picture of reasons described, then, to say that you have an undefeated defeater for your reasons for believing as you do, is to say that

1. Your reasons are *prima facie* reasons.
2. You would be justified in believing as you do on the basis of those reasons in some context.
3. You are not justified in believing as you do on the basis of those reasons in the present context.

In other words, defeaters render beliefs that would otherwise be justified unjustified.

This gives us a generic model of what it is for something to be an epistemic defeater. We can recognise two principal varieties of defeater: *rebutting defeaters* and *undercutting defeaters*.

Sometimes you acquire counter-evidence for things you already believe. If that counter-evidence is as strong as, or stronger, than your original reasons for believing as you do, it is a rebutting defeater for those reasons. Pollock defines rebutting defeaters as follows:

**Rebutting Defeater:**

If P is a prima facie reason for S to believe Q, R is a rebutting defeater for this reason if and only if R is a defeater (for P as a reason for S to believe Q) and R is a reason for S to believe  $\sim$ Q (Pollock 1986: 38).

On this definition, there are two conditions for something to constitute a rebutting defeater. First of all, it must be a reason to believe the negation of the proposition supported by your original reasons. Second, it must be a defeater in the generic sense described previously. That is to say, if a reason R to believe  $\sim$ Q is a rebutting defeater for P as a reason to believe Q, then, you would not be justified in believing Q, were you to believe Q on the basis of the conjunction of R and P. To illustrate, consider the following two cases.

In the first case, imagine that a doctor with the unlikely name General Practitioner (let's call her GP) diagnoses a patient with condition-Z on the basis of a positive result from one reliable clinical test. GP then receives a negative result from a second equally reliable test. GP has no reason to believe that either test is more reliable than the other. In this case, GP acquires evidence against her initial diagnosis that is just as good as the evidence she has for it. Consequently, GP's total evidence no longer justifies her diagnosis. Thus, the second test result is a rebutting defeater for the first. Call this case COMPETITION.

Now, suppose that, in place of a second test result, GP receives testimony from a colleague that the patient does not have condition-Z. And suppose that GP's colleague is familiar with the patient's clinical history, but not with the positive test result – and GP is aware of this. In this case, as in the previous, GP has reason to believe that the patient has condition-Z (the clinical history and test result) and reason to believe that the patient does not have condition-Z (her colleague's testimony). The test result, though, outweighs her colleague's testimony. Moreover, given that GP is aware that her colleague has not seen the test result, she can reasonably expect that he will change his opinion on learning of it. In other words, since GP can take herself to have a clear epistemic

advantage over her colleague, he is not her effective peer.<sup>49</sup> Thus, we might suppose, there is no question that GP's diagnosis remains justified even on learning of her colleague's opinion. Call this case NO-COMPETITION. In cases like NO-COMPETITION, R's being a reason to believe  $\sim Q$  is not sufficient for it to be a rebutting defeater for P as a reason for S to believe Q, since it is not sufficient for it to be a defeater for P as a reason for S to believe Q

Rebutting defeaters characteristically leave intact the connection between your original reasons and the content of your beliefs but shift the balance of your total evidence such that those connections are no longer sufficient to justify your belief. Sometimes, though, we acquire evidence that alters the bearing of our total evidence by undermining those connections. Imagine, for instance, that GP receives the test result and forms her diagnosis accordingly, but, then, receives word that there was a mishap in the lab and her patient's test may have been contaminated. Evidence of contamination is not evidence that the patient does not have condition-Z – he may well do. Hence, this evidence is not a rebutting defeater. Nonetheless, if GP can't be sure that the test is uncontaminated, then, any diagnosis she bases on the result of that test won't be justified. Call this case CONTAMINATION. In this case, GP's new evidence defeats her old, as Pollock puts it, 'by attacking the connection between the evidence and the conclusion, rather than attacking the conclusion itself.' (Pollock 1986: 39)

This is an example of undercutting defeat. Pollock defines what it is for something to be an undercutting defeater as follows:

### **Undercutting Defeater**

If P is a *prima facie* reason for S to believe Q, R is an undercutting defeater for this reason if and only if R is a defeater (for P as a reason for S to believe Q) and R is a reason for S to deny that 'P would not be true unless Q were true.' (Pollock 1986: 39).<sup>50</sup>

---

<sup>49</sup> Nor, since GP is justified in believing that her colleague is unaware of an absolutely crucial piece of evidence, is he her *approximate peer*.

<sup>50</sup> It is clear elsewhere that Pollock uses the term 'reason to deny "P would not be true unless Q were true"' to mean 'reason to believe  $\sim(P \text{ would not be true unless } Q \text{ were true})$ ' (E.g. Pollock 1987: 485) On the current understanding of what counts as a reason for belief, this is too strong a condition. For instance, if GP was told 'there is a 50% chance that the test was contaminated' she would not have reason to believe  $\sim(\text{this test would not have come up positive unless my patient has condition-Z})$  – but rather that there is a 50% chance that that proposition is true. Still, I presume that if, GP is justified in believing that 'there is only a 50% that this test would not have come up positive unless my patient has condition-Z', she will have an undercutting defeater for the test result as a reason to believe her patient has condition-Z. We might say that in this kind of case, the agent would have reason to doubt that 'P would not be true unless Q were true'. In what follows, I shall proceed on the understanding that 'R is an undercutting defeater for P as a reason to believe Q, iff R is a defeater [...] and R is reason to deny or doubt that 'P would not be true unless Q were true'.

In CONTAMINATION, the possibility of Z's test result being contaminated is a reason to deny that 'the test result would not be what it is unless the patient had Condition-Z'. This is because there is now another explanation of the test result being as it is, the contamination. Thus, the defeater undercuts GP's justification by giving her reason to believe that the usual connection between a positive test result and the presence of Condition-Z does not hold in the current context. Another way in which GP's justification could be undercut, though, would be if she had reason to doubt that there was any kind of connection between a positive test result and the presence of condition-Z to start with.

To illustrate, imagine that underwriting GP's use of the first test in her diagnostic practices is her awareness of the results of a large-scale randomized control trial that showed a statistically significant correlation between a positive result in the test and later development of condition-Z. Now, imagine that, having made her diagnosis, GP reads a follow up study that details several problems with the randomized control trial that would invalidate its results if true. Call this case FOLLOW-UP.

In FOLLOW-UP, once she is aware of the problems with the randomized control trial, any justification GP's diagnosis might have had in respect to the test results is defeated. It is not defeated, though, because GP has evidence that the positive test result fails to have significance it would otherwise have. Rather, the follow-up study attacks the support that GP has for believing that there is ever a connection between a positive test and the patient's condition. And, as such, GP also has reason to doubt that the test results would not be as they are unless the patient has condition-Z. The follow up study pointing out the flaws in the trial is an undercutting defeater. But it is of a kind that is importantly different from CONTAMINATION. Pollock's definition is not sensitive to this distinction.

Undercutting defeaters are reasons to doubt that the connections between P and Q that would justify believing Q on the basis of P hold. What CONTAMINATION and FOLLOW-UP illustrate is that there are two ways the justificatory relationship between P and Q can be called into doubt. A defeater R might be evidence that the connection between P and Q fails to hold in the current context, whilst giving S no reason to doubt that P justifies Q in other contexts. Or, R might be a reason for S to doubt that P justifies Q in any context. Let's call the first kind of defeater a *contextual undercutting defeater* and the second a *general undercutting defeater*. We can define each as follows:

### **Contextual Undercutting Defeater**

If P is a *prima facie* reason for S to believe Q, R is a *contextual undercutting defeater* for this reason if and only if R is an undercutting defeater and S could justifiably believe (were it not for R, P would be reason to believe Q) without having a defeater-defeater for R.

### **General undercutting defeater**

If P is a *prima facie* reason for S to believe Q, R is a *contextual undercutting defeater* for this reason if and only if R is an undercutting defeater and S could not justifiably believe (were it not for R, P would be reason to believe Q) without having a defeater-defeater for R.<sup>51</sup>

Just as for rebutting defeaters, what it is for something to count as a contextual or general undercutting defeater on these definitions is derivative of the more general conditions on defeaters given previously. First of all, if R is either kind of undercutting defeater, it must have the kind of content characteristic of undercutting defeaters more generally. That is to say, R must in some way constitute reason for S to deny or doubt that ‘P would not be true unless Q were true’. Second, R must satisfy the generic definition of a defeater. Thus, even if R *is reason for* S to deny or doubt that ‘P would not be true unless Q were true’ it will not constitute an undercutting defeater of either sort if, other things being equal, S could still justifiably believe the conjunction of P, Q, and R.<sup>52</sup> These together, of course, are the conditions of undercutting defeat more generally. What distinguishes the two types of undercutting defeater is what, in light of R, S is permitted to believe about the justificatory relationship between P and Q.

Where R is a contextual-undercutting defeater, S has no reason to doubt that the justificatory relationship between P and Q holds in other contexts – thus, S might justifiably believe that P is a *prima facie* reason to believe Q at the same time as believing R. CONTAMINATION is a case of contextual undercutting defeat. Given the evidence of contamination GP ought not to base her diagnosis upon the test result. She has no reason to doubt, though, that that result would justify her diagnosis were it to turn out to be uncontaminated. Correspondingly, GP could justifiably

---

<sup>51</sup> See below for discussion of defeater-defeaters.

<sup>52</sup> Suppose for instance, that in FOLLOW-UP the putative defeater comes in the form, not of the follow-up study, but testimony from Di’s colleague that ‘that test is not reliable’. So long as it is appropriate for Di to trust her colleague, his testimony would constitute reason to deny that ‘there is unlikely to be a positive test result unless the patient has condition-Z’ – but, similarly to NO-COMPETITION, his testimony is outweighed by Di’s awareness of the results of the randomized-control test that indicate that the test is a reliable indicator of condition-Z. Thus, in this case, Di would have evidence that has the content characteristic of an undercutting defeater (i.e. reason to deny ‘not-P unless Q’) but that evidence would not be sufficient to constitute a defeater for the test result as reason to believe the patient has condition-Z.

believe *both* that the current test result is contaminated and that, in other contexts, a positive test result would justify her believing that the patient has condition-Z.

In cases of general undercutting defeat, R impugns the justificatory relationship between P and Q precisely because, for any context in which S believes (or ought to believe) R, S ought to deny or doubt that there is any context in which P justifies believing Q. At least, that is, unless S has some further evidence that shores up her justification for believing P is a *prima facie* reason to believe Q. FOLLOW-UP is a case of general undercutting defeat. Given her awareness of the follow-up study, GP ought not to base her diagnosis upon the test result. But that is because (i) her use of the test result in her diagnostic practice is underwritten by her justified belief that the test is a reliable indicator of condition-Z, and (ii) by impugning the results of the trial on which that belief is based, the follow-up study undercuts that justification. Without that justification, or some other evidence to that effect, GP cannot justifiably believe that a positive test result is evidence of condition-Z in this, or any other, context.

This completes the brief survey of the relevant kinds of defeater. It leaves us with three explanations of how negative higher-order evidence might defeat an agent's justification for believing as they do to:

1. The higher-order evidence is a rebutting defeater for the agent's reasons for belief.
2. The higher-order evidence is a contextual undercutting defeater for the agent's reasons for belief.
3. The higher-order evidence is a general undercutting defeater for the agent's reasons for belief.

As we saw in the previous chapter, Thomas Kelly effectively treats higher-order evidence against one's justification for believing as one does as a form of counter-evidence to one's first-order evidence. Thus, where one's higher-order evidence is a defeater, it would work as a rebutting defeater. Conciliationists such as Matheson (2009, 2015) and Christensen (2010, 2016b) – as well as Feldman (2009) – have suggested that defeating higher-order evidence works as a form of undercutting defeater.<sup>53</sup> Though not always voiced in such terms, the majority opinion in the wider

---

<sup>53</sup> Comments from Feldman that higher-order evidence undercuts 'not by claiming that a commonly present connection fails to hold in a particular case, but rather by denying that there is an evidential connection at all' (Feldman 2009: 113) suggest that he views defeating higher-order evidence of the kind discussed here along the lines of general undercutting; comments from Christensen that 'there is something right about [Feldman's] suggestion' (Christensen

literature on higher-order evidence appears to be in line with the conciliationists (for instance Dipaolo 2018; Horowitz 2014; Schoenfield 2018; Silva 2017; Sliwa and Horowitz 2015). Having said that, a number of authors have argued that we should accept some form of ‘level-splitting’ view (the ‘level-splitting’ label is owed to Christensen 2013) whereby the rational response to the kind of situation Bones and Truman find themselves in in BROWNIES and HEAVY MUSIC is to apportion one’s beliefs to both the first-order and higher-order evidence. Thus, on a level splitting view Bones should believe both that ‘The body is male’ (in response to the first-order evidence) and something like ‘It is not clear to me that the relevant evidence supports believing the body is male’ (in response to the higher-order evidence). Defences of level-splitting have been forwarded by Coates (2012), Lasonen-Aarnio (2014), and Worsnip (2018). I shall not engage with these positions independently. Rather, I shall just note that all of these arguments rest on some variant of the claim that defeat from higher-order evidence cannot be accommodated in non-ad-hoc fashion within a general schema of defeat and a general theory of justification and reasons for belief.<sup>54</sup> I agree. Thus, I shall take it that: if we cannot accommodate the relevant kinds of case within a more general schema of defeat – and are required to introduce special (read ad hoc) principles of justification to explain the putative cases of defeat – we may have reason to go back to the drawing board when it comes to the general suggestion that negative higher-order evidence (including that acquired on realising disagreement) can have normative significance. On the other hand, if we can accommodate the relevant kinds of case within a more general schema, I shall presume we do not have to entertain the level-splitting view.

With that said, in the next section I shall consider whether cases like BROWNIES and HEAVY MUSIC, as well as the different kinds of ordinary disagreement between effective peers, fit to the structure of rebutting, contextual undercutting, or general undercutting defeat. If they do not, then, we may have reason to entertain a level-splitting view. If they do, then, we shall be half-way to explaining how such evidence can have normative significance and, so, half-way to a higher-order approach to ‘The Problem of Ordinary Disagreement’.

Before doing that, however, it is important to note two further aspects of epistemic defeat.

---

2010: 194) suggest that he accepts this conclusion. Neither explicitly argue for this claim though, nor do they place any great significance upon it in their broader approaches to higher-order evidence – thus, I shall not take these comments as either authoritative or necessarily representative of the respective author’s views.

<sup>54</sup> This idea is perhaps most clearly expressed by Lasonen-Aarnio when she writes ‘assume that I am right that there is no elegant way of accommodating systematic defeat by higher-order evidence. [...] Ultimately, I think the right conclusion to draw is that the desire to accommodate limitless defeat lands one into a paradoxical predicament with no happy resolution’ (Lasonen - Aarnio 2014: 342).

First, defeaters can be misleading. For instance, in FOLLOW-UP, so long as she has no reason to doubt the veracity of the follow-up study, GP would still acquire a defeater if that study was a clever fake produced by a powerful pharmaceutical company to drive business toward their own proprietary testing solution. Second, and related to that, just as you can acquire defeaters for your reasons for belief you can acquire defeaters for those defeaters. For instance, if GP learns that the follow-up study is a clever fake, its significance in respect to the results of the trial will be defeated. In turn, the justification that her diagnosis enjoys in respect to the test result will be restored. Where the follow-up study was a defeater for the test result, then, evidence that the follow-up study is bogus is a *defeater-defeater* for that evidence. Just like defeaters, defeater-defeaters can come in rebutting or undercutting form. And, in principle, just as you can acquire defeater-defeaters, in principle, you can acquire defeater-defeater-defeaters, and so on. Importantly, then, any higher-order approach to disagreement needs to consider, not only what form of defeater one acquires in case of defeat from higher-order evidence, but also what kinds of consideration might work as a defeater-defeater in such cases. I consider defeater-defeaters in Chapter 5.

For the moment, though, let's consider whether any of the cases of putative defeat from higher-order evidence fit the general schema of defeaters we have seen in this section.

## 4.3 Higher-order evidence and defeaters

Let's start by considering whether we can explain the significance of the higher-order evidence in any of the cases in terms of rebutting defeat.

### 4.3.1 Rebutting defeat

As we saw in the previous section, the conditions on R being a rebutting defeater for P as a reason for S to believe Q are just that (i) R is a reason for S to believe  $\sim Q$  and (ii) the conjunction of P and R is not a reason to believe Q. Clearly enough, the higher-order evidence in BROWNIES and HEAVY MUSIC is not a rebutting defeater. That Bones ate some dodgy brownies is certainly not evidence that the body on her examining table is not a male body (as she originally concluded it was). That Truman's playlist may have permanently impaired his ability to concentrate on exam-marking is not evidence in favour of any other set of grades that Truman could have dished out.

What about cases in which justification for believing Q is purportedly defeated when you learn that someone disagrees with you about Q?



Here it is worth returning to a distinction between two core types of disagreement noted in Chapter 1: Cases in which the disputants are effective peers and share the same evidence, and cases in which the disputants are effective peers and have different evidence.

#### 4.3.1.1 Different evidence disagreements

Let's consider 'different evidence' cases first. To remind us, here is the example from Chapter 1 in full again:

SPIES

Agent-M is an US intelligence analyst. Agent-M has collected a large body of data pertaining to the interference of a foreign agent G in the recent US election, on which basis she has concluded that 'G interfered'. Agent-M liaises with Agent-S, a British analyst investigating the same issue, who informs Agent-M that he believes 'G did not interfere'. Agent-M and Agent-S are both justified in believing each other to be competent and trustworthy analysts. However, both Agent-M's and Agent-S's evidence is constituted by information that is classified by their respective governments and neither are in a position to share all of their evidence with the other. As it happens, Agent-M's belief that 'G interfered' is justified by her evidence.

Let's consider this case from Agent-M's perspective. What significance does Agent-M's realisation that Agent-S disagrees with her about G's purported interference in the US election have upon her justification for believing that G did interfere?

Agent-M knows that Agent-S is both trustworthy and competent. Thus, given that:

(1) Agent-S says 'G did not interfere'

It follows that:

(2) (1) is *prima facie* reason for Agent-M to believe that Agent-S has evidence that justifies believing G did not interfere,

And, so

(3) (1) is *prima facie* reason for Agent-M to believe that G did not interfere

Treated evidentially, given the stipulations about competence and sincerity, Agent-S's testimony is *evidence of evidence* that supports 'G interfered'. In the absence of defeaters, were Agent-M to believe 'G did not interfere' on the basis of Agent-S's testimony, she would be justified in believing 'G did

not interfere'.<sup>55</sup> Thus, the discovery of the disagreement fulfils the first condition of rebutting defeat. Does it fulfil the second condition? I.e. Is it a defeater in the more general sense. It does, but in a way that is distinct from a case like COMPETITION.

In COMPETITION, GP initially has the evidence of the first test. On acquiring the evidence of the second test, the *total evidence GP has* consists of the first and second test. Since, as stipulated, these together do not support believing 'the patient has condition-Z', GP has a rebutting defeater. In this kind of case, whether the counter-evidence is a defeater reduces to the question of whether the *first-order evidence the agent has* supports their initial belief or not. Importantly, in this kind of case – given that GP has no evidence of any other evidence – GP can rely upon her judgement of *the evidence she has* as reliable guide to the truth about the patient's condition.

Since, in this kind of case the putative defeater defeats without calling into question the practices by which the agent came to believe Q, let us call a defeater of this kind a *first-order rebutting defeater*. And let's say more generally that:

### **First-order rebutting defeat**

If P is a *prima facie* reason for S to believe Q, R is a *first-order rebutting defeater* for P as a reason for S to believe Q, *iff* R is a rebutting defeater (for P as a reason for S to believe Q) and S is permitted to rely upon her judgement of (P&R) as a reliable guide to the truth about Q.

Now consider Agent-M's position in SPIES. Agent-M initially has the data she collected. On realising the disagreement, Agent-M has as first-order evidence the data and Agent-S's testimony. Agent-S's testimony is *prima facie* reason for Agent-M to believe ~G interfered, but it is also higher-order evidence of evidence. It is in respect to the higher-order aspect of the evidence M acquires on discovering the disagreement that M acquires a defeater. For crucially, given the practical details of the case, i.e. that Agent-M and Agent-S's are prevented from sharing their evidence, Agent-M does not have access to the content of the evidence on which S based his belief that 'G did not interfere'. Thus, given the *evidence of evidence* that supports G did not interfere, M can no longer rely upon her judgement of *the evidence she has* as a reliable guide to the truth about G's interference. (And this is so even if the conjunction of the evidence Agent-M has with the evidence Agent-S has does in fact support Agent-M's original belief.) In lieu of a defeater-defeater, then, Agent-M

---

<sup>55</sup> There is some debate as to whether, as Feldman (2006) coined the term, 'evidence of evidence is evidence'. Given the features of this case, though, we can put aside those worries. As we might, rather clumsily, transpose the slogan Agent-M has 'evidence that justifies believing that there is evidence that justifies believing ~Q'. Evidence that fits that description, I presume is evidence that ~Q (for further discussion of the question of whether 'evidence of evidence is evidence', see Feldman 2006, 2009, 2014; Fitelson 2012).

will no longer be justified in believing ‘G interfered’. Thus, the discovery of the disagreement fulfils the second condition of rebutting defeat.

If a first-order rebutting defeater leaves the agent’s judgement of the *evidence they have* unquestioned, then, it seems apt to call a rebutting defeater that prohibits an agent from relying upon their judgement of the evidence they have a *higher-order rebutting defeater*. Thus, let’s say more generally that:

### **Higher-order rebutting defeat**

If P is a *prima facie* reason for S to believe Q, R is a *higher-order rebutting defeater* for P as a reason for S to believe Q, *iff* R is a rebutting defeater (for P as a reason for S to believe Q), R is a *prima facie* reason for S to believe there is evidence R\* that S does not have and that justifies believing  $\sim Q$ , and S is not permitted to rely upon her judgement of (P&R) as a reliable guide to the truth about Q.

The evidence that one acquires on discovering disagreement with an effective peer who does not share one’s evidence, thus, does afford one a rebutting defeater – it affords one a *higher-order rebutting defeater*.

#### **4.3.1.2 Same evidence disagreements**

As reconstructed in the previous chapter, Thomas Kelly’s ‘Total Evidence’ view effectively treats the higher-order evidence acquired in case of same-evidence disagreements on the lines of rebutting defeat. As we saw in that discussion, this way of conceiving such evidence fails to explain how it could have significance in same-evidence disagreement cases – whether with one’s effective peer or, indeed, one’s superior. Though we have covered some of this ground already it will be worth revisiting these issues in light of the discussion in the previous section. To do so, I will focus upon DOCTORS-1. In this case, recall, Doctors Quinn and Zaius are attempting to determine which of two treatments, A and B, is the best option for a patient they are jointly treating. Based on her assessment of the relevant clinical and scientific evidence, Quinn believes ‘A is best’. Based on his assessment of the same evidence Zaius believes ‘B is best’. As the case was described, Zaius is Quinn’s effective peer and Quinn is initially justified in believing ‘A is best’ on the basis of the clinical and scientific evidence.

Let’s consider the case from the perspective of Quinn.

Similarly as with Agent-S for Agent-M in SPIES, given that Zaius is Quinn’s effective peer, Quinn is justified in believing that Zaius is both sincere and competent when it comes to assessing the

clinical and scientific evidence relevant to the treatment options. Thus, paralleling the reconstruction of SPIES, given that:

(1) Zaius says 'B is best'

It follows that,

(2) (1) is *prima facie* reason for Quinn to believe that 'Zaius has evidence that justifies believing B is best.'

Therefore:

(3) (1) is a *prima facie* reason for Quinn to believe 'B is best'

Where this case differs from SPIES is in that:

(4) Quinn is justified in believing that Zaius based his opinion 'B is best' on the same body of clinical and scientific evidence that she has already assessed.

Given the discussion in the previous section, we might approach the question of whether the evidence Quinn acquires on learning of the disagreement is a rebutting defeater by considering whether it can be accommodated under the heading of either first-order rebutting or higher-order rebutting.

Consider higher-order rebutting first. As we defined higher-order rebutting, for R to be a higher-order rebutting defeater-defeater for P as a reason for S to believe Q it must be evidence of evidence that justifies believing  $\sim Q$ . More specifically, it must be evidence of evidence that S does not have. Given (4), however, this is not the case for Quinn. At most, Zaius's testimony is evidence *that the evidence Quinn has* is evidence that justifies believing  $\sim Q$  – it is not, though, evidence of evidence that Quinn does not have. Thus, this case – and same-evidence peer disagreements in general – are not cases of higher-order rebutting defeat.

The question, then, is whether they are cases of first-order rebutting defeat? I would suggest not. Why? Because first-order rebutting defeaters leave intact the connection between one's original first-order evidence and one's original conclusion – instead, they work by changing the balance of the first-order evidence one has. That being so, the evidence that Quinn acquires from Zaius's testimony will not be a first-order rebutting defeater. For, if Zaius's testimony does not give Quinn reason to doubt the connection between the clinical and scientific evidence and her conclusion that 'A is best', then, given that, as stipulated:

- (5) Prior to learning (1), Quinn justifiably concluded that the clinical and scientific evidence establishes that ‘A is best’,

Quinn will have access to the following argument:

- (i) The clinical and scientific evidence establishes that ‘A is best’
- (ii) Zaius believes that ‘B is best’
- (iii) Zaius is wrong
- (iv) Zaius has made a mistake in his assessment of the clinical and scientific evidence
- (v) I need not give any weight to Zaius’s opinion that ‘B is best’

And if Quinn has access to that argument, then, the higher-order evidence she acquires on learning of the disagreement with Zaius will not be a first-order rebutting defeater – for just the reason that it will not be a defeater at all.<sup>56</sup>

It is worth being a little careful here. As far as Pollock’s general conditions for rebutting defeaters go, since Zaius’s testimony is *prima facie* reason for Quinn to believe ‘B is best’, the evidence Quinn acquires on learning of the disagreement will be a rebutting defeater in the formal sense – just so long as it is a defeater. I.e. Zaius’s testimony is a reason for Quinn to believe that ‘B is best’, What I am saying here, though, is that Zaius’s testimony does not fit either of the specifications of how rebutting defeat *works* – i.e. as first-order or higher-order rebutting. Thus, if the realisation of disagreement in which the disputants share the same evidence affords any party a defeater, that defeater must do so via some form of undercutting – just as for Brer’s testimony to Bones in BROWNIES and Truman’s newfound awareness of the debilitating effects of his music preferences in HEAVY MUSIC.

With that said, then, let us turn to consider whether we can account for any of these cases in terms of undercutting defeat – and if so, what kind of undercutting defeater that might be.

### 4.3.2 Undercutting defeat

In the previous section, we saw that the higher-order evidence acquired in different-evidence peer disagreements satisfies the conditions of rebutting defeat. We also saw that, if negative higher-order evidence acquired in the non-disagreement cases *or* in same-evidence disagreement has defeating force, it must do so in the way characteristic of undercutting defeat. If it can, then, we

---

<sup>56</sup> This analysis, of course, echoes the discussion of the Dogmatism Paradox and Independence in Chapter 2 and what I called the PCTE conception of higher-order evidence in Chapter 3.

have the makings of a case that such evidence does have defeating force. If it cannot, then, in line with the suggestions made by Lasonen-Aarnio (2014), Worsnip (2018), and others, we should consider the possibility that defeat from higher-order evidence cannot be accommodated within a standard model of justification.

Earlier, we saw that undercutting defeaters work, in general, by calling into question the justificatory relationship between your reasons for belief and the content of your beliefs. We have already seen that this is the only way in which the higher-order evidence acquired in cases like BROWNIES and HEAVY MUSIC, as well as the same-evidence disagreement cases, can have defeating force. At first blush, it also seems an apt description of what happens in the various cases. Consider Bones and her suspicious brownies.

In this case, Brer's testimony about the brownies is not evidence that the deceased was male. It is also not reason for Bones to doubt that the specific skull and pelvis measurements on which she based her conclusion are correct – after all, she collected and recorded that evidence before she ate the brownies. Brer's testimony is, though, reason for Bones to believe that the *calculation* she made with those measurements was much more likely prone to error than it would usually be. And, if she has reason to believe that, it seems reasonable to suppose that she ought also to doubt that the connections between her evidence and her conclusion are as she took them to be.

Similar points apply to Quinn's discovery of the disagreement with Zaius. That her otherwise competent and trustworthy colleague believes that 'B is best' may well afford Quinn *prima facie* reason to believe that 'B is best'. At the same time, however, the realisation that Zaius came to believe 'B is best' on the basis of the same clinical and scientific evidence that she took to establish 'A is best', raises the possibility for Quinn that her judgement of the bearing of that evidence may have somehow gone awry. The discussion in the previous section established that, whilst Quinn's discovery of the disagreement does afford her *prima facie* reason to believe 'B is best', its defeating force must lie in its casting doubt on the justificatory connection between the clinical and scientific evidence and Quinn's conclusion that 'A is best'. That Quinn has evidence of the possibility that she may have gone awry in her judgement of the clinical and scientific evidence would seem to explain why. For, if Quinn has reason to doubt her judgement in that way, then, it seems reasonable to suppose that she ought also to doubt that the connections between her evidence and her conclusions are as she took them to be.

The higher-order evidence in both the same-evidence disagreement and non-disagreement cases, then, does look like an undercutting defeater. To paraphrase David Christensen (2010: 198) – and

echoing the discussion of SPIES – the higher-order evidence in each case appears to put the agent in a position where they cannot *trust* the practices by which they came to treat P as a reason to believe Q. In section 4.2, however, I suggested that Pollock’s general definition of undercutting is insensitive to a significant distinction between two kinds of undercutting defeater: *contextual undercutting defeaters* and *general undercutting defeaters*. So, we also need to ask if any of these cases fit the model of contextual or general undercutting defeat – and what determines which case fits which model if that is the case.

Per the earlier discussion, a contextual undercutting defeater R for P as a reason for S to believe Q, casts doubt upon the connection between P and Q in the context of the case at hand – but in such a way that R’s significance upon the relationship between P and Q is limited to the case at hand. The significance of a general undercutting defeater, on the other hand, carries across from the current context into any other context in which S’s evidence relevant to Q is P.

Let’s start with BROWNIES. Presuming that the psychotropic agent Brer included in the brownie recipe has no permanent effects, Brer’s testimony is not evidence for Bones of a *general* shortcoming in her skills as a forensic anthropologist. Instead, it is evidence of a higher than usual likelihood of performance error in the specific context following her consumption of the spiked brownies (and for a relevant period thereafter). So, we might say, Brer’s testimony about the Brownies is reason for Bones to doubt that she was in a state of mind to properly appreciate the bearing of the physical evidence she collected earlier *when* she assessed that evidence. But it is not reason for Bones to distrust her judgement either before ingesting the psychotropic agent or in future examinations made after the drug has worn off.

In a case like BROWNIES, then, what the higher-order evidence indicates about the agent’s epistemic practices is specific to her performance of those practices in the context to which that evidence pertains. Thus, the higher-order evidence is evidence that Bones may have made an error in judging the bearing of the physical measurements on the sex of the body. And so, Brer’s testimony is evidence for Bones that the skull and pelvis measurements may not, as she judged them to be, be evidence that the body is male. Importantly, though, just as the doubts raised about Bones performance are confined to the moments after she consumed the brownies, so too is Brer’s testimony status as a reason to doubt ‘the physical measurements are evidence that the body is male’ confined to those moments.

Or, in other words, the higher-order evidence in this kind of case is a *contextual undercutting defeater* for the first-order evidence.

Interestingly, however, this need not be the case for all instances of putative defeat from higher-order evidence. Compare HEAVY MUSIC. As this case was described, Truman acquires evidence that the kind of music he was listening to whilst marking the exam scripts may have had debilitating effects upon his ability to concentrate on just that kind of cognitive task. If Truman is unable to concentrate when marking the exam papers, he is presumably more likely to make an error when marking those papers than if he were concentrated. Thus, similarly to Brer's testimony for Bones, Truman gets evidence of an increased likelihood of performance error after listening to the music than he would have expected beforehand. Unlike that case, however, since the research paper was evidence that the debilitating effects of heavy-industrial music are *permanent* (with even the slightest exposure) the significance of that evidence is not confined to a single context. Instead, it is evidence for Truman that – since listening to his friend's playlists – he will be unable to properly concentrate *anytime* he is marking exam scripts in the future. Thus, in so far as the research paper was compelling evidence and Truman has no defeater-defeaters for that evidence, the doubts that Truman ought to have about the practices and performance by which he reached *this* set of exam marks are not specific to the context in which he initially came to those marks – but are quite general for any other relevantly similar context (from that point onwards).

In other words, the higher-order evidence in this kind of case is a *general undercutting defeater*.

Higher-order evidence of the kind acquired in the non-disagreement cases, then, can constitute either a contextual or general undercutting defeater. As per the discussion of the two cases above, which category of undercutting defeater any particular case corresponds to will depend upon whether the doubts induced by the evidence are specific to the performance of the agent in the particular case or are more general in scope. Now let's consider the same-evidence disagreement cases.

As we saw earlier, it is reasonable to suppose that the higher-order evidence acquired in the same-evidence permutations of peer disagreement generate the same kinds of doubts about the agent's assessment of their first-order evidence as do the non-disagreement cases. Likewise, that kind of doubt seems to raise questions about the justificatory relationship between an agent's first-order evidence and the conclusions they reached on that evidence that are characteristic of undercutting defeat in general. The question, then, is whether those doubts are (i) specific to the context of disagreement or (ii) general in nature.

In the case of BROWNIES and HEAVY MUSIC we were able to determine which kind of doubts were in play by paying attention to the scope of the potential errors indicated by the higher-order



evidence (i.e. a temporary impairment of Bones's performance; and a permanent impairment of Truman's concentration). To return to comments in Chapter 1, it is possible that those party to a disagreement could be effective peers and have evidence of error that is just as specific as the evidence that Bones and Truman acquire. For instance, we might add the additional wrinkle to DOCTORS-1 that Quinn and Zaius both pulled a 36-hour shift prior to assessing the evidence pertaining to the treatment options for their shared patient (and both are aware of this). That being so it would seem reasonable for Quinn (and Zaius) to conclude that the most likely explanation of their disagreement is that one of them made a mistake in their assessment of the clinical and scientific evidence as a result of extreme fatigue and tiredness. In which case, any doubts Quinn ought to have about the practices by which she came to believe 'A is best' would be specific to the context. Thus, when this detail is added to the case, the higher-order evidence Quinn acquires on realising the disagreement would correspond to contextual undercutting defeat. (Likewise for Zaius) We might just as easily add details to this case (or any other hypothetical case) such that the doubts about the agents' practices and performance would be more general in nature – and so, the disagreement-evidence would correspond to general undercutting defeat. We shall consider just such a class of disagreements in Chapter 6. Though, for now, I shall leave it to the readers to guess which class of disagreement this is (hint: Chapter 6 is about deep disagreement).

Similarly, to the non-disagreement cases, then, the higher-order evidence acquired in cases of same-evidence peer disagreement may come in the form of a contextual undercutting defeater or a general undercutting defeater. And, which that is will depend upon whether the evidence the disputants have about the content and context of their disagreement indicates a context-specific or general source of error. On the other hand, where the evidence the disputants have about the content and context of their disagreement *does not* allow them to narrow the possible explanations of the disagreement down to either context-specific or general sources of error – the higher-order evidence will be a general undercutting defeater. Why? Because, when S has higher-order evidence R of some kind of error or flaw in the practices by which she came to believe Q on the basis of P – and she cannot rule out a general source of error, then, S could not justifiably believe (were it not for R, P would be reason to believe Q) – or at least not without some form of defeater-defeater for R.

It is important to note, however, that the kinds of case in which disputants are effective peers, the disagreement is an ordinary disagreement, and the disputants cannot rule out general sources of error, will be relatively uncommon. This is because, in ordinary disagreement (or at least the kinds that are theoretically interesting) interlocutors will only be effective peers if they are justified in

believing that they are all generally competent. Unless there is specific reason to think a general error is more likely than usual in the particular circumstances, satisfying this competence condition will be enough to rule out more general errors. So, for instance, in the original DOCTORS-1, none of the evidence Quinn has about the content and context of her disagreement with Zaius indicates anything specific about how the disagreement arose. However, Quinn *is* justified in believing that she and Zaius generally conduct their medical practices with a high degree of expertise and competence. Thus, without a defeater for her justification for believing *that*, Quinn might justifiably infer that the disagreement is most likely explained by some factor or feature of the case that is specific to the context of the disagreement. In which case, as per the discussion above, the higher-order evidence Quinn acquires on realising that Zaius disagrees with her will be in the form of a contextual undercutting defeater – even whilst Quinn has no evidence of any particular context-specific flaw or shortcoming in the practices by which she came to hold the disputed belief.

In principle, then, cases of same-evidence disagreement between effective peers can be instances of both contextual and general undercutting defeat – in practice, however, where there is no positive evidence of a general flaw or shortcoming as explanation of the disagreement, such cases are most likely to be cases of contextual undercutting defeat.

## Conclusion

This completes the first half of the study of higher-order evidence, ordinary peer disagreement, and defeaters. Significantly, we have found that we can accommodate all of the various cases within a more general structure of reasons and defeaters. As noted in section 4.2., however, a full account of ‘The Problem of Ordinary Disagreement’ also needs to say something of the kinds of consideration that might constitute a defeater-defeater for the higher-order evidence in the relevant cases. I shall address this topic in the next chapter. To do so, I shall return to a suggestion that I glossed over earlier: that defeating higher-order evidence works by placing the agent in a position where they cannot rely upon or trust the epistemic practices by which they came to hold the relevant beliefs.

# Chapter 5. Higher Order Evidence, Defeater-Defeaters, and Epistemic Self-Trust

## Introduction

In the previous chapter we saw that the cases of putative defeat from higher-order evidence can be accommodated within a standard framework of defeaters. This includes cases of ordinary disagreement between effective peers. In this chapter, I turn to consider the kinds of consideration that might count as defeater-defeaters in such cases. Echoing the distinction between first-order and higher-order forms of defeater forwarded in the previous chapter, I suggest in the first section of the chapter that, where one has independent evidence that rebuts or undercuts the defeating higher-order evidence, one will have a *first-order defeater-defeater* for that higher-order evidence. In cases of disagreement between effective peers no such evidence is available. Thus, I suggest that, unless there can also be *higher-order defeater-defeaters*, conciliation will always be the appropriate response to the realisation of such disagreement. Roughly, if defeating higher-order evidence puts the agent in a position where they cannot, in lieu of defeater-defeaters, *trust* their appreciation of the first-order evidence, a higher-order defeater-defeater would be reason in favour of such trust. In the rest of the chapter I address the question of whether there can be higher-order defeater-defeaters that are genuinely distinct from first-order defeater-defeaters. To do so, I explore what we might mean by epistemic self-trust, what role it has to play in the structure of justification and reasons, and what forms such self-trust might take.

## 5.1 Forms of defeater-defeater

In this section, I make some general points about defeater-defeaters as they apply to cases of higher-order defeat. These points lead to the suggestion that, just as there can be first-order and higher-order defeaters, so too we might consider the possibility that there are both first-order and higher-order defeater-defeaters.

Following the discussion of defeaters at 4.2 there are two general points that shape how we theorise defeater-defeaters.

First, defeaters work by giving an agent access to reason to believe ‘ $\sim Q$ ’ (rebutting) or reason to doubt or deny ‘ $\sim P$  unless  $Q$ ’ (contextual and general undercutting). For convenience, let’s refer to these propositions as ‘defeat propositions’. Thus, we can say, when  $R$  is a rebutting defeater, the corresponding ‘defeat proposition is  $\sim Q$ ; when  $R$  is an undercutting defeater the corresponding defeat proposition is ‘ $\sim P$  unless  $Q$ ’. The discussion of the various kinds of defeating higher-order evidence in the previous chapter showed that defeaters can give an agent access to reasons to believe, deny, or doubt such propositions in two ways. Where  $R$  is a *first-order* defeater for  $P$  as a reason for  $S$  to believe  $Q$ ,  $R$  does so because the defeater pertains directly to the relevant defeat proposition. Where  $R$  is a *higher-order* defeater,  $R$  does so indirectly – by placing  $S$  in a position where, in lieu of a defeater-defeater,  $S$  cannot *trust* the practices by which she came to believe  $P$  on the basis of  $Q$ .

Second, defeater-defeaters are just defeaters. What are they defeaters for? They are defeaters for defeaters as reasons to believe the corresponding defeat proposition.

Thus, taking these two points together, for any defeater-defeater  $DD$  for  $R$  as a defeater for  $P$  as a reason for  $S$  to believe  $Q$ , and where  $(_{\text{def}}P)$  refers to the relevant defeat proposition, we can say:

- (i) If  $DD$  is a defeater-defeater for  $R$  as a defeater for  $P$  as a reason for  $S$  to believe  $Q$ , then,  $(DD\&R)$  is not a reason for  $S$  to believe  $(_{\text{def}}P)$ .

i.e.  $DD$  must be a defeater for  $R$  as a reason to believe, doubt, or deny  $(_{\text{def}}P)$  in the general sense.

- (ii) If  $DD$  is a defeater-defeater for  $R$  as a defeater for  $P$  as a reason for  $S$  to believe  $Q$ , then,  $DD$  is a reason for  $S$  to believe, doubt, or deny  $(_{\text{def}}P)$ ,

i.e.  $DD$  must be a rebutting, contextual undercutting, or general undercutting defeater for  $R$  as a reason to believe, doubt, or deny  $(_{\text{def}}P)$ .

These constraints simply reiterate the general definitions of defeaters given in the previous chapter. Importantly, since both first-order and higher-order defeaters have defeating significance for  $S$  only in so far as they bear in some way upon a relevant defeat proposition, these conditions hold whether  $R$  comes in the form of a first-order or higher-order defeater. For that reason, just as we saw that higher-order defeaters fit to the general taxonomy of defeaters, so we only need to introduce a single taxonomy of defeater-defeaters for first-order and higher-order defeat. Whilst it is not entirely necessary in the current discussion, I have included such a taxonomy at fig. 1 at the end of the thesis.

The different classes of defeater-defeaters, then, are simply iterations of the different kinds of defeater we discussed in the previous chapter, indexed to the particular kinds of doubt thrown up by each category of defeater. That being so, just as we described *first-order defeaters* as defeaters that pertain directly to the relevant defeat propositions for P as a reason to believe Q, so we might describe *first-order defeater-defeaters* as defeaters that pertain directly to the defeat propositions relevant to R's status as a defeater for P.

To illustrate this idea, suppose that in SPIES, after talking to Agent-S, Agent-M acquires evidence that agents of G coerced Agent-S into telling Agent-M that he believed G did not interfere in the US election. Given this evidence, Agent-S's testimony would no longer be reason for Agent-M to believe that G did not interfere in the US election. As such, this new evidence would be a defeater-defeater for the higher-order evidence Agent-M acquired on learning of the disagreement originally. More specifically, it would be a *first-order contextual undercutting defeater-defeater*. Or, suppose that in DOCTORS-1, Quinn acquires new evidence that Zaius has faked his entire qualification history and track-record – as it turns out Zaius is not a highly qualified medical doctor but a Doctor of Philosophy, and a middling one at that. In this case, Quinn's new evidence would be reason for her to deny that Zaius's testimony that 'B is best' was ever reason to believe 'B is best' or that 'the medical and clinical evidence supports believing B is best'. Thus, Quinn's new evidence would be a defeater-defeater for the higher-order evidence she originally acquired on learning of the disagreement. More specifically, it would be a *first-order general undercutting defeater-defeater*.

That one's defeater comes in the form of higher-order evidence about the practices by which one came to believe Q, is in principle no barrier to having a first-order defeater-defeater for that higher-order evidence. Crucially, though, for that to be so, the putative defeater-defeater cannot fall within the scope of the doubts about the agent's practices induced by the higher-order evidence. Where defeating higher-order evidence is acquired through the realisation of disagreement, this amounts to the more specific requirement that any first-order defeater-defeater an agent has must be *independent* of the substance of the disagreement. If not, it will fall under the scope of the doubts about the agent's practices induced by the higher-order evidence acquired on realising the disagreement. Moreover, any first-order defeater-defeater must also be independent reason for the agent to believe that they have the relative epistemic advantage over their opponent vis-à-vis the substance of the disagreement. For, if it is not, then it is not a defeater-defeater. The significance of these points in respect to 'The Problem of Ordinary Disagreement' should be quite clear. 'The Problem of Ordinary Disagreement' arises in cases where disputants are effective peers. Yet, someone you disagree with is your effective peer if and only if you have no *independent* reason to

believe that you or they have the relative epistemic advantage vis-à-vis the substance of the disagreement. In other words, ‘The Problem of Ordinary Disagreement’ concerns just those cases in which disputants do not have a first-order defeater-defeater for the higher-order evidence the realisation of that disagreement affords them. This implies that, if the only kinds of defeater-defeater one might have for defeating higher-order evidence are first-order defeater-defeaters, Conciliationism will be the correct view of ordinary disagreement. Given the sceptical consequences of Conciliationism discussed in Chapter 2, however, this would be a dispiriting conclusion indeed.

For that reason, I would suggest that we have motivation to consider the possibility that, just as there can be higher-order defeaters, there might also be *higher-order defeater-defeaters*. To explore this possibility, I want to spend some time discussing a thought alluded to in the previous chapter. That is the idea, to paraphrase David Christensen, that higher-order evidence works by putting the agent in a position where, in lieu of a defeater-defeater, they cannot trust the practices by which they came to believe Q on the basis of P (Christensen 2010: 198). To echo a point made by Silva, commenting on Christensen, the thought that misleading higher-order evidence gives an agent reason to *distrust* their epistemic practices suggests that in other circumstances it can be appropriate for the agent to *trust* their epistemic practices (Silva 2017). In the remaining sections of this chapter, then, I want to return to the topic of epistemic self-trust touched upon in Chapter 3 to consider: what is the nature of epistemic self-trust, what role does it play in the structure of justification and reasons, and what kinds of consideration might make it appropriate to have epistemic self-trust.

## 5.2 Epistemic self-trust

### 5.2.1 Reliance

As we saw in Chapter 3, discussions of epistemic self-trust centre around the thought that one cannot achieve one’s fundamental epistemic aims without relying upon one’s epistemic practices. Before discussing the nature of epistemic self-trust, it will be worth going over this thought and its motivations in some more detail.

In what way, then, must each of us rely upon our own epistemic practices?

The first thing to consider is the way in which our chances of fulfilling our fundamental epistemic aims depend upon the reliability of our epistemic practices.

As I have been using the term, ‘epistemic practice’ refers to the broad gamut of ways in which we come to acquire beliefs. Thus, forming beliefs based on the deliverances of my perceptual faculties

is amongst my epistemic practises; but so too are inquiry, inference to the best explanation, trusting others, and so on. In short, I form beliefs by way of my epistemic practices, and any belief that I might come to hold, I must have done so by way of one or more of my epistemic practices.<sup>57</sup>

Say, though, that none of my epistemic practices reliably yield truths over falsehoods (i.e. all of my practices are unreliable). That being so, the beliefs I come to hold by way of those practices are no more likely to be true than false. Presumably, though, my fundamental epistemic aims include the aims of acquiring true beliefs and avoiding false beliefs. But, if any beliefs I might come to hold are no more likely to be true than false, then, I am no more likely to achieve those aims than not to achieve them. Similarly, if my practices are all anti-reliable, i.e. they reliably yield falsehoods over truths, I will be less likely to achieve my fundamental epistemic aims than I am to achieve them. It is only if my practices reliably yield truths over falsehoods that my chances of acquiring true beliefs and avoiding false beliefs will be higher than not. In this way, my chances of achieving my fundamental epistemic aims are dependent upon how reliably my epistemic practices yield truths over falsehoods.<sup>58</sup>

Broadly speaking, there are two ways in which my having unreliable epistemic practices can prevent my achieving my epistemic aims. One is if they include practices that are generally unreliable. For example, wishful thinking is (presumably) not a reliable way of forming beliefs. Thus, if I come to hold a number of beliefs because I wish the relevant propositions true, those beliefs will no more likely be true than false. Correspondingly, other things being equal, I will be less likely to achieve my fundamental epistemic aims than if wishful thinking were not amongst my practices. Similarly, for practices such as counter-induction, forming beliefs via my biases and prejudices, and so on. These are all ways in which I might come to form beliefs, but in so far they are generally unreliable ways of doing so, beliefs formed via such practices are no more likely to be true than false (and depending on the practice, may be more likely false than true). Thus, if any of these are amongst my epistemic practices, other things being equal, I will be less likely to achieve my fundamental epistemic aims than if they are not.

A second way in which my practices may prevent me achieving my fundamental epistemic aims is if I employ otherwise reliable practices in unreliable ways. For instance, assume that induction as

---

<sup>57</sup> Discussion of self-trust is sometimes framed in terms of reliance on one's own cognitive faculties (see Foley 2001; Fricker 2016; Zagzebski 2012). I prefer 'epistemic practice', since, it is not only our cognitive faculties that we rely on but also the various methods, rules, and norms that we follow when exercising those faculties.

<sup>58</sup> This is not to say anything about whether I might have other epistemic aims that can be fulfilled if my practices fail to be reliable.

is generally practiced by people in our world is reliable. Even if that assumption happens to be correct, my own practice of induction can still fail to be reliable. I might tend to over or underestimate relations of inductive support, make my inductions on false premises, only on grue/bleen like properties, consistently infer from non-representative samples, and so on. Take another example – our practices of forming beliefs via testimony. Assume that the world we live in contains trustworthy and untrustworthy speakers, but that testimony is in general a reliable way of acquiring beliefs. Even if that assumption is correct, if I am truly terrible at discriminating between trustworthy and untrustworthy people to the extent that I am more likely to believe the untrustworthy than the trustworthy, beliefs that I form on the basis of other peoples' testimony are unlikely to be true. On the assumptions made, then, both of these are cases in which a generally reliable practice is unreliably employed. This is the second way in which my practices may prevent me achieving my fundamental epistemic aims.

To take stock. We rely on our epistemic practices in the sense that our chances of achieving our epistemic aims depend upon those practices reliably yielding truths and not falsehoods. And, we can be let down by our practices if (i) they are generally unreliable, or (ii) we fail to employ otherwise reliable practices competently.

The second feature of our reliance on our own epistemic practices I want to draw attention to is the way that – as believers – we *act* or *reason as if* those practices are reliable. Imagine, for instance, that I come to believe that there are two orange birds in the garden on the basis of my experience of seeing two orange birds in the garden. We might say that, in forming that belief, I *act as if* my perceptions have yielded a truth about the avian contents of my garden. Similarly, if I suppose that the best explanation of there being two orange birds in my garden is that it is summer in Africa and so come to believe that it is summer in Africa, I *reason as if* my practice of inferring to the best explanation has yielded the truth about the season in Africa. Importantly, such acting or *reasoning as if* need not result from any explicit consideration of my epistemic practices, nor reflect any explicit attitude toward their reliability. Instead, we might say, in acting or reasoning as I do, I *presuppose* that the deliverances of those practices are reliable indicators of the way things are in the world. Following Wright (2004) and Gerken (2013), we can understand such presupposing as a kind of underlying normative commitment. That is to say, to take the current example, it would be epistemically irrational for me to believe that it is summer in Africa, if I were to doubt the reliability of my practice of IBE or my eyesight when I took myself to see two orange birds in the garden. More generally, we might describe this kind of commitment in terms of the following principle:



Presupposition of Reliability (PR): For any practice K and belief Q formed via K, other things being equal, S would not be justified in believing Q were S to doubt that ‘K is reliable’.<sup>59</sup>

So stated, and as per the previous paragraph, this principle does not require belief in the reliability of one’s practices, only an absence of doubt. As such, PR is a clear corollary to what I called the ‘Objective Attitude Condition’ (OAC) on justification in the Introduction. As I shall discuss shortly, the negative nature of this commitment is crucial to the place that epistemic self-trust has in our lives. What is important at this point, though, is that this kind of presupposition underlies any attitude of belief that I might come to hold. In this sense, so long as I hold any beliefs at all, I cannot avoid relying on at least one of my epistemic practices.

Note that this is not to say that one cannot avoid taking for granted the reliability of specific practices in certain circumstances. I can, for instance, avoid relying upon my eyesight to yield the truth as to the state of my garden by keeping my eyes closed whenever I am in, or even turn my head towards, the garden. In this case, I would avoid relying on my garden-vision by not employing the practice at all. It is harder to see, though, how I could lead any kind of life without ever engaging my perceptual faculties. To do that, I would have to look at nothing, listen to nothing, feel nothing, and so on. Similarly, even if I had relied upon my garden-vision when initially forming my bird-beliefs, I could allay that reliance by, say, snapping a photo of the garden and showing it to my neighbour. Having double-checked the truth of my beliefs about the contents of my garden with my neighbour, I would no longer presuppose the reliability of my garden-vision. Nonetheless, I only allay my reliance on my garden-vision by, instead, relying on my judgement of other people’s trustworthiness/testimony (as well as my perception of utterance). In respect to these points, then, it is worth making a further distinction between epistemic practices. In his seminal paper ‘Epistemic Circularity’, William Alston defines epistemically ‘basic sources of belief’ as follows:

O is an epistemologically basic source of belief – df. Any otherwise cogent argument for the reliability of O will use premises drawn from O (Alston 1986: 8).<sup>60</sup>

Alston’s definition fits neatly with the internalist notion of justification that we have been working with and is easily transposed into the terms of the current discussion. Along these lines, let’s say that:

---

<sup>59</sup> Similarly, Wright states: ‘P is a *presupposition* of a particular cognitive project if to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project.’ (Wright 2004: 191)

<sup>60</sup> In his later work on epistemic circularity Alston switches to the term ‘doxastic practices’ instead of ‘sources of belief’ (See Alston 2005)

K is a basic epistemic practice for S, *iff* K is amongst S's practices and S *could not* justifiably believe 'K is reliable' without presupposing that K is reliable.

K is a non-basic epistemic for S *iff* K is amongst S's epistemic practices and S *could* justifiably believe 'K is reliable' without presupposing that K is reliable.

So defined, my basic practices are those that I cannot allay reliance upon by way of any of my other practices. These include perception, as well as fundamental forms of reasoning such as induction and deduction. Practices like my garden-vision are non-basic – I can legitimately allay my reliance on these by employing other practices to establish their reliability.

One further point of importance regarding this distinction is that any general attempt to allay reliance on one's non-basic practices will ultimately lead back to reliance on one's basic practices. For instance, I might be astute enough to make sense of some inference to the best explanation to the effect that testimony is reliable – and so come to believe that testimony is reliable without presupposing that it is. In doing so, however, I presuppose that my inferences to the best explanation are reliable. Inference to the best explanation, though, is either basic for me, or it is not. If it is basic, I have hit bedrock, so to speak. If it is not, my attempt to allay reliance on my non-basic practices will not be successful, and I shall have to dig deeper. For that reason, we can make a parallel point about each of our standing vis-à-vis the sum of our practices as to our basic practices. That is to say, it is not possible to believe 'My epistemic practices are reliable overall' without presupposing that at least of some of those practices are reliable.

With these various point in mind, then, it appears that the only other way that I could entirely avoid reliance upon my epistemic practices would be to withhold judgement on each and every output of those practices. Such withholding of belief, I assume, is practically impossible. Thus, it is also practically impossible to avoid relying on one's epistemic practices altogether.

What I now want to suggest is that, not only can one not avoid relying upon one's epistemic practices, but that the possibility of having justified belief depends upon the rational permissibility of doing so. The argument is simple enough.

Suppose that:

- (1) S believes Q

From the foregoing discussion, it follows that:

- (2) There is some practice K such that S cannot believe Q without relying upon K.

And, to this we can add:

- (3) If it is rationally impermissible for S to rely on K, it is rationally impermissible for S to believe Q

And, since:

- (4) If it is rationally impermissible for S to believe Q, S is not justified in believing Q

We can also infer that:

- (5) If it is rationally impermissible for S to rely on K, S is not justified in believing Q, and
- (6) If S is justified in believing Q, it is rationally permissible for S to rely on K

A necessary condition of having justified beliefs, thus, is that it is rationally permissible to rely on the practices by which we come to hold those beliefs. Two further lines of thought allow us to draw additional conclusions from this argument.

First, as noted in previous chapters, it is implicit in debates about the epistemic significance of disagreement that it is possible to have justified beliefs. Likewise, for the debate over how higher-order evidence might serve as a defeater for beliefs justified by first-order evidence. For the sake of the current discussion, then we can table sceptical worries about the possibility of justified belief (or at least those that do not follow from discussion of disagreement and higher-order evidence). With that said, let us stipulate that:

- (7) S has acquired a large number of justified beliefs via K

I.e. K is the kind of practice that can produce justified beliefs and S has been a competent employer of K.

The second line of thought picks up the point from earlier, that we often form beliefs without having considered the practices by which we came to hold those beliefs or the reliability of those practices. For, whilst we might suppose that we are sensitive to facts about our epistemic practices up to a point – and indeed PR requires this – the project of establishing the reliability of those practices is a distinct epistemological undertaking of the sort that we do not ordinarily engage in. At the same time, it is hard to see why we should think that beliefs formed without having considered the reliability of the relevant practices would generally represent any kind of error in performance of those practices. For example, it does not seem that my chance of avoiding performance error in coming to believe that I saw two orange birds in the garden particularly

depends upon my having a conscious, reflective attitude to my garden-vision. Or, at least not if those circumstances are what we might think of as normal (e.g. good light, a clear view, not under the influence of psychotropic substances, not fanatically hopeful of seeing a hoopoe, and so on.)

What this suggests, though, is that for many of us, a large number of our beliefs will be formed and held (i) without having reflected upon the reliability of the practices by which we came to hold those beliefs, and (ii) without having committed any performance error in the ways that we employed those practices to do so. But if that is so, I would suggest it is reasonable to suppose that the possibility of justified belief does not depend upon one's having reflected upon the reliability of the relevant epistemic practices. For, if it does, we would have to allow that someone could be in normal circumstances, have evidence P that justifies believing Q, assess the bearing of P upon Q via the competent employment of an appropriate (justificatory) practice, and on that basis believe Q – yet not be justified in believing Q. But, except in cases where the person in question also has undefeated defeating higher-order evidence – that seems to describe a paradigmatic instance of justified belief.

Following these points, then, and presuming that S is much like the rest of us, K is much like her other epistemic practices, and given the stipulation at (7) – I would suggest that we can assume the following:

- (8) S has acquired a large number of justified beliefs via K, without having any conscious attitude toward the reliability of K

And, if that is so, we can also infer that:

- (9) In lieu of defeaters, it is rationally permissible for S to rely on K without having any conscious attitude toward the reliability of K.

Following Burge (2003, 2013a) and Wright (2004) we can refer to this in terms of *entitlement*. In Burge's words:

[E]ntitlements 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. The unsophisticated are entitled to rely on their perceptual beliefs. Philosophers may articulate these entitlements. But being entitled does not require being able to justify reliance on these resources, or even to conceive such a justification. (Burge 2013b: 230).

Thus, the more general suggestion is that we are each *entitled* to rely on our epistemic practices without having any conscious attitude towards their reliability – or at least those of our practices by which we come to hold justified beliefs.

This argument and its conclusion represent a more detailed formulation of the comments on self-trust in Chapter 3. A key point I want to bring out here, however, is that this is not an argument for self-*trust* in anything but name. ‘Trust’, in contrast to reliance, is typically understood to include some conscious attitudinal component taken by the trusting toward the trusted (see Baier 1986; Faulkner 2011; Holton 1994; Jones 1996) So understood, the entitlement to rely on one’s epistemic practices without any attitude toward that reliance is not an entitlement to self-trust. What I want to suggest next, then, is that there are certain contexts in which this entitlement does not hold and in which it is not permissible to rely on one’s epistemic practices without having some kind of positive attitude toward that reliance. As we shall see, these are contexts in which justified belief requires a genuine form of epistemic self-trust.

What kind of contexts do I have in mind?

Contexts in which an agent realises both that they came to believe Q via K and that beliefs formed via K are unlikely to be true if K is not reliable; and contexts in which an agent ought to realise both that they came to believe Q via K and that beliefs formed via K are unlikely to be true if K is not reliable.

To illustrate, let’s suppose that S is in the kind of context described. That being so, we can say

- (10) S realises, or ought to realise, that she came to believe Q via K
- (11) S realises, or ought to realise, that beliefs formed via K are no more likely to be true than false unless K is reliable.

Let’s say that, given (10), S is in a *context of realisation* vis-à-vis her belief Q. We have established that, in other contexts, S is entitled to rely on K without any conscious attitude toward K’s reliability. Being in a *context of realisation* is not, on its own, sufficient to bring this entitlement into question. (Simply realising that I came to believe there are ‘two birds in the garden’ via perception is not a reason to doubt my perceptions). However, when S is in a context of realisation *and* also realises that the likely truth of beliefs formed via K depends upon how reliable K, S is confronted directly with the question: Is K reliable? And, once that question is raised, if S cannot give a positive answer to it, S ought to rely on K no longer. Why? Because, to be faced with the question ‘Is K reliable’, yet have taken no attitude toward the reliability of K that would involve the presumption

that K is, or is likely to be, reliable – just is to doubt that K is reliable. And, if S doubts that K is reliable, or ought to doubt that K is reliable – then S cannot continue to rely on K without contravening PR. So, we might say given (11), the *context of realisation* at (10) becomes a *context of doubt*. And, when in a *context of doubt* vis-à-vis her reliance on K:

(12) S ought to take a positive attitude toward the reliability of K, or not rely on K.

In respect to that suggestion, then, more generally, we can introduce the following principle:

Positive Attitude Requirement (PAR): For any practice K and belief Q formed via K, if S is in a *context of doubt* vis-à-vis Q and K, other things being equal, S will not be justified in believing Q unless S has a positive attitude toward the reliability of K and that attitude is a defeater for the doubts that K is reliable.

PAR complements both PR and the OAC condition on justification suggested in the Introduction to the thesis. Given that PAR comes into play in S's case, the question is what kind of close the question as to whether K is reliable and so underwrite S's continued reliance on K?

The obvious answer is that S might close the question via belief – more specifically, belief that 'K is reliable'. The combination of reliance and belief in the reliability of the thing relied upon, though, describes a genuine form of trust. Following Faulkner (2011), we can refer to this form of trust as *predictive trust*. Correspondingly, were S come to believe that K is reliable, and so continue relying on K, we might reasonably describe S as having *predictive self-trust* (in respect to K). Once we have in mind the idea that S's situation requires a form of epistemic self-trust, however, the question of whether there are forms of self-trust other than predictive self-trust comes into play. For, just as predictive trust is only one among various forms of trust more generally – so it may be that predictive self-trust is only one among forms of epistemic self-trust.

What would be the significance of this for the current discussion?

In short, when one acquires higher-order evidence that impugns the practices by which one came to believe Q – including that acquired in case of disagreement with an effective peer or peers -- one is placed in a *context of realisation* vis-à-vis Q precisely because one ought to consider the question 'Is K reliable?' In other words, higher-order evidence generates *contexts of doubt*. Thus, in lieu of some independent reason to dismiss the significance of the higher-order evidence acquired on realising that one is engaged in such a disagreement, one will retain any justification one has for believing Q if and only if one has the kind of positive attitude that can properly underwrite one's continued reliance on the relevant practices. The possibility that there may be forms of epistemic

self-trust other than predictive self-trust, however, raises the possibility that one may have such an attitude even if one would not be justified in believing that those practices are reliable. And, that being so, it may also be so that there can be higher-order defeater-defeaters available to those party to peer disagreement that are significantly distinct from what we referred to as first-order defeater-defeaters.

In the next section, I shall consider in more detail how predictive self-trust works. As I shall argue, there are limitations to the extent that predictive self-trust can be justified. These limitations require us to seriously consider the possibility that there are other forms of epistemic self-trust.

### 5.2.2 Predictive self-trust

As noted in the previous section, the obvious attitude through which one might close the question of whether one's epistemic practices are reliable is belief that those practices are reliable. I described this combination of belief and reliance as *predictive self-trust*. Let's say, then, that:

Where S is amongst S's epistemic practices, S has **predictive self-trust** *iff*:

- (i) S epistemically relies on K
- (ii) S believes that K is reliable

As I explain below, the belief component of predictive self-trust can underwrite reliance on one's epistemic practices in contexts of doubt by affording one reason to believe that the beliefs one comes to by way of those practices are true. Consider:

Given the stipulation that S is in a context of doubt vis-à-vis K, S either realises or ought to realise that the likely truth of her belief Q depends upon the reliability of K. Now, suppose that S has predictive self-trust (in K). This entails that:

- (13) S believes that K is reliable

Given (13), however, so long as S has understood the context in which she finds herself, S could reason to 'Q' as follows:

- (a) My belief 'Q' was formed via K
- (b) Beliefs formed via K are more likely to be true than false
- (c) My belief 'Q' is more likely true than false
- (d) Q is more likely true than false

S could make this inference either by simple induction or inference to the best explanation. Thus, so long as S is permitted to rely on induction or inference to the best explanation, the following will also be true:

- (14) If S is justified in believing that K is reliable, then, in lieu of defeaters, were S to believe Q on the basis of her belief that K is reliable (in conjunction with the realisation that she came to believe Q via K), S would be justified in believing Q

I.e. S's belief that K is reliable is, in the context, *prima facie* reason for S to believe Q. Importantly, whilst the argument from (a) to (d) gives S access to a new source of justification for believing Q, it does not allay her reliance on K entirely – as might be the case if S has access to an argument for Q that makes no reference at all to K or its reliability. For, were S to believe 'Q' *on the basis of her belief* 'K is reliable', S would still rely on K as a source of information, no matter the source of justification for that inference. That being so, it follows from (13) that:

- (15) If S is justified in believing that K is reliable, S is rationally permitted to rely on K

How might one acquire justification for that kind of belief?

Broadly speaking, there are two types of case to consider. There are cases in which one comes to predictive self-trust through practices other than those under review. This would be the case, for instance, if I were to come to believe that my practice of trusting the testimony of others is reliable by way of inference to the best explanation. There may also be cases in which one comes to, or can only come to, predictive self-trust through the very same practices that are under review. This would be the case, for instance, if I were to come to believe that my practice of trusting the testimony of others is reliable on the basis of testimony to that effect from someone that I trust.

The first kind of case is, in principle, unproblematic. Consider S again and suppose the following:

- (16) L is amongst S's epistemic practices  
(17) S has acquired justified beliefs via L  
(18) S comes to believe 'K is reliable' via L

That being so, then, as per the earlier argument, we can say that:

- (19) In lieu of defeaters, S's belief 'K is reliable' is justified



In which case, for any beliefs that she forms via K, S will have access to the relevant variations on the inference pattern at (a)-(d) and so, in lieu of defeaters, will have justification for those beliefs. And, that being so, it follows that:

(20) S is rationally permitted to rely on K

I see no reason to think that there will not be a wide variety of situations in which it would be possible to acquire predictive self-trust (relative to specific practices) in this way. Nonetheless, there is an important limitation to how far this gets us. Following Alston (1986), we can describe instances in which predictive self-trust is acquired by way of the same practices in which that trust is placed in terms of *epistemic circularity*. As I shall discuss in more detail in the next section, justification is epistemically circular if one cannot justifiably believe the target proposition without presupposing that that proposition is true. As per the discussion of basic and non-basic practices earlier, however, one cannot justifiably believe that one's basic epistemic practices – or the sum of one's practices – are reliable, without presupposing that those practices are reliable. Reflecting these points, the limitation to predictive self-trust is that one cannot have justified predictive self-trust when it comes to one's basic practices, or the sum of one's practices that is *not* epistemically circular justification. In the next section, I discuss the implications of this point.

### 5.2.2.1 Epistemic Circularity and Predictive Self-Trust

Epistemic circularity, as is widely noted, is distinct from logical and rule circularity.<sup>61</sup> The following epistemically circular track-record argument illustrates this well:

#### **TR-K**

At  $t_1$ , I came to believe  $Q_1$  via K,  $Q_1$

At  $t_2$ , I came to believe  $Q_2$  via K,  $Q_2$

[...]

At  $t_n$ , I came to believe  $Q_n$  via K,  $Q_n$

Conclusion. K is reliable (i.e. K reliably yields truths over falsehoods).<sup>62</sup>

Any shortcomings TR-K might display cannot be chalked up to other forms of circularity. The conclusion does not appear in the premises, so the argument is not logically circular. The conclusion does not describe the form of inference instantiated by the argument, so the argument

---

<sup>61</sup> See Alston (1986), Battaly (2012), Bergmann (2004).

<sup>62</sup> This is a reconstruction of an argument Alston (1986: 9) lays out to illustrate epistemic circularity with reference to perception.

is not rule-circular. Rather, what seems circular about this kind of argument is that I cannot believe the premises of the argument without in some way presupposing that the conclusion is true. Specifically, the problem lies with the second conjunct of each premise. Say that  $Q_1$  is the belief that ‘my curtains are red’ and  $K$  is *perception*. It seems that, to establish the second conjunct of the first premise, I need to somehow verify my belief that ‘my curtains are red’. I might do so directly, say, by going back to look at the curtains again. In that case, though, I will depend upon the reliability of my perceptions when verifying my initial belief, just the same as I did when I first formed it. Alternatively, I might try to verify my initial belief indirectly, say, by asking you whether you believe the curtains are red. In that case, I might indeed verify my initial belief without relying upon my perception of the curtain’s redness. However, in accepting your testimony I would still presuppose that my perceptions are reliable more generally. To paraphrase Bergmann, it is only by relying on my perceptions that I believe that there is someone with me and that they confirmed what my senses told me about the colour of my curtains (Bergmann 2004: 739).

It is in this respect, then, that Alston and others following him describe this kind of argument as epistemically circular. Whether epistemically circular arguments can ever confer justification is the matter of some debate.<sup>63</sup> There is consensus, though, that they carry no persuasive force. As already discussed, epistemically circular routes to belief work by presupposing the truth of the belief in question. What this means, though, is that epistemically circular arguments cannot rationally compel belief in a person who is not in some way committed to the truth of the conclusion of that argument. For instance, to call back to one of our core disagreement cases, if Zaius believes that Quinn’s judgement of medical evidence is unreliable, Quinn could not rationally compel Zaius to believe that her judgement of that evidence is reliable by way of an epistemically circular track-record argument of the sort we saw before. To quote Bergmann:

[A]n argument is commonly evaluated in terms of how useful it would be in convincing someone who initially doubts its conclusion. An epistemically circular track record argument fails abysmally by that standard. It is of no use whatsoever to anyone who begins by questioning its conclusion. (Bergmann 2004: 720).

To say that epistemically circular arguments carry no persuasive force unless the audience for that argument is already committed to the truth of the conclusion to that argument is, in effect, to say that epistemically circular arguments cannot close questions – when they are genuinely open. This applies equally to cases where one is presenting such an argument to another person and when

---

<sup>63</sup> See Alston (1986, 2005), Bergmann (2004), Foley (2001), Sosa (2009) for qualified endorsements of epistemically circular justification; Battaly (2012) and Brown (2004) respectively for criticisms of Sosa and Bergmann’s positions; Cohen (2002), Fumerton (1995), and Vogel (2000) for rejections of the idea of epistemically circular justification.

one is considering that argument for oneself. For instance, if S treats the question of whether ‘K is reliable’ as a *genuinely* open question, then, S *should not* be persuaded by consideration of the argument at TR-K. Why not? Because S can only reason from the premises of TR-K if she presupposes that ‘K is reliable’ – and, if S reasons in a way that presupposes that ‘K is reliable’, S does not treat the question of whether ‘K is reliable’ as genuinely open. Whilst, then, S’s consideration of TR-K may bring to the surface her presuppositions about the reliability of her own epistemic practices – S should not make any new commitments of that sort on such a basis.

Now, this on its own does not mean that S cannot come to have justified predictive self-trust via an argument such as TR-K, even in a context of doubt. After all, if K is already amongst S’s practices, then, S is already committed to the presupposition of K’s reliability. So, *perhaps* it is enough that S can use an epistemically circular argument to bring her commitments to the surface – even if she cannot make any new commitments on that basis.<sup>64</sup> I don’t want to rule this out – yet, it seems to me that there is an important limit to this kind of reasoning. Namely, it will not deliver justified predictive self-trust *if* belief is the only kind of attitude that can underwrite reliance on one’s epistemic in a context of doubt. To see why, it will be useful to return to an earlier point in our general argument. Thus, let us move back a few steps and stipulate that (contrary to (16)-(20)):

(16’) K is a basic practice for S

Given (16’), S could *only* justifiably come to believe ‘K is reliable’ by relying on K. Thus, S could only come to have justified predictive self-trust via an argument such as TR-K, which moves from premises of the kind ‘At t<sub>1</sub>, I came to believe Q<sub>n</sub>, Q<sub>n</sub>’, to the conclusion ‘K is reliable’. Given (10) and (11), however, S is in a context of doubt vis-a-vis Q and K and, as per PAR, in such a context S will not be justified in believing Q via K unless S has a positive attitude toward the reliability of K. These, together with (16’), entail that:

(17’) If S does not have a positive attitude toward the reliability of K, then, S will not be justified in believing the premises of TR-K (or any similar argument)

Now, suppose that the only attitude that could underwrite S’s reliance on K in a context of doubt is belief (i.e. reliance in a context of doubt requires predictive self-trust). Given this assumption, it follows from (17’), that:

---

<sup>64</sup> Foley refers to this as the project of putting ‘one’s intellectual house in order’ (Foley 2001: 13); it also recalls what Salmon(1991), drawing on Reichenbach, refers to as a ‘vindication’ of one’s reasoning.

- (18') If S does not believe that K is reliable, then, S will not be justified in believing the premises of TR-K.

But, if S cannot justifiably believe the premises of the argument without believing 'K is reliable', S cannot justifiably *come to believe* 'K is reliable' on the basis of that argument. Nor, even if she did so beforehand, can S justifiably *continue to believe* that 'K is reliable' if her original justification came via an argument like TR-K. (Which, given K is basic, it must have done). For, given the demand that the implicit presupposition of K's reliability be made explicit in a context of doubt, 'K is reliable' would in effect come to feature amongst the premises of TR-K. In which case, the argument would no longer be epistemically circular, but logically circular. And, whether there are contexts in which epistemically circular arguments can confer justification, I assume there are no contexts in which logically circular arguments can do so. Thus, S cannot have justified predictive self-trust within a context of doubt – at least not if predictive self-trust is necessary for S to be rationally permitted to rely on K in such a context. Given, then, that we are working under the assumption that one can sometimes be rationally permitted to continue to rely on one's basic epistemic practices in a context of doubt – there must be some combination of attitude and reliance other than predictive self-trust that makes it so. In other words, there must be some form of epistemic self-trust that does not involve belief. I shall explore what that might be in the next section.

### 5.2.3 Self-trust without belief

The idea that there is some form of epistemic self-trust that does not involve belief is a widespread one, both in the literature explicitly focused on self-trust and the literature on epistemic circularity. Richard Foley, for instance, suggests that since the kinds of sceptical doubt associated with epistemic circularity cannot be eliminated by any degree of inquiry, at some point inquiry must stop and self-trust step in. At times, Foley describes self-trust in terms of an 'intellectual faith' in one's practices – as when he claims that: 'Significant inquiry requires an equally significant leap of intellectual faith. The faith need not, and should not, be unlimited; that is the path to dogmatism and irrationalism. But there does need to be such faith.' (Foley 2001: 20). And at other times, it is clear that self-trust as Foley conceives it has a cognitive element that is in parallel with, if not equivalent to, belief. Thus, he reiterates at various points that 'trust in one's intellectual faculties, methods, and practices ought to be proportionate to the degree of confidence one has in their reliability and to the depth of this confidence' (Foley 2001: 25). But, whilst these remarks are suggestive, Foley does not develop them sufficiently to help us answer the question of just how

self-trust can underwrite reliance on one's epistemic practices when it does not include (justified) belief.<sup>65</sup>

Similar points apply to Crispin Wright's discussion of the *rational trust* we each must have in the truth of any number of propositions that we cannot but presuppose in our various intellectual projects (Wright 2004, 2014). These presuppositions include one's about the reliability of our epistemic practices and faculties, and so we might take this 'rational trust' to encompass what I have referred to as epistemic self-trust. Echoing Foley, Wright is careful to emphasise the cognitive component of self-trust. Thus, and in line with the earlier discussion of whence the need for self-trust arises, he argues that 'if acceptance of such a presupposition [about the reliability of one's epistemic practices] is to be capable of underwriting rational belief in the things to which execution of the project leads, it has to be an attitude which excludes doubt' (Wright 2004: 193) and elsewhere, 'it needs to be a rational attitude to take for reasons other than our possession of evidential support for the effectiveness of the methods concerned.' (Wright 2014: 227)

As the latter quote makes clear, Wright is of the view that the conditions of this rational trust must be distinct from the justification conditions of predictive self-trust. I agree. Again, though, whilst these points help us to mark out the boundaries between self-trust without belief and predictive self-trust, they tell us little about the kind of attitude that might accompany the former.

In that respect, it is worth briefly considering how the attitude of trust has been discussed in the broader corpus of work on the topic, and especially in work upon relations of interpersonal trust. Here we do find more substantive accounts. Particularly promising is the suggestion made by a number of writers that there is an *affective* dimension to certain forms of epistemic trust. So, Karen Jones suggests that:

Trusting is composed of two elements, one cognitive and one affective or emotional [...] Roughly, to trust someone is to have an attitude of optimism about her goodwill and to have the confident expectation that, when the need arises, the one trusted will be directly and favourably moved by the thought that you are counting her (Jones 1996: 5).

And Richard Holton, that:

In cases where we trust and are let down, we do not just feel disappointed, as we would if a machine let us down. We feel betrayed. [...] betrayal is one of those attitudes that Strawson calls the reactive attitudes. These are attitudes that we normally take only towards people. We feel hurt or resentful when they let us down; grateful, perhaps touched, when they help (Holton 1994: 66).

---

<sup>65</sup> This is not a criticism of Foley's discussion of self-trust, it is just not a question that comes up in his discussion.

Though not as such in conflict, these quotes illustrate two ways in which the affective nature of interpersonal trust can be conceived depending on which aspect of that relationship is emphasised. On Jones's account, trust is affective primarily because the attitude of the trustee *when trusting* involves an affective, as well as, cognitive component. Holton, on the other hand, emphasises the kinds of reactive – and so affective – attitudes that the trustee is liable to experience *when the trusted acts or fails to act* in line with the trust placed in them. Paul Faulkner draws upon this latter approach to identify a mode of trust that he explicitly labels *affective trust* and which stands in contrast to predictive trust. He defines this as follows:

A **trusts** S to  $\Phi$  (in the **affective** sense) if and only if (1) A depends on S  $\Phi$ -ing; and (2) A expects (1) to motivate S to  $\Phi$  (where A expects this in the sense that A expects it of S that S be moved by the reason to  $\Phi$  given by (1)) (Faulkner 2011).

Like predictive trust, 'affective trust' so defined implies a relationship of reliance and an expectation that the trusted party will make good on that relationship. Where they differ is in the nature of that expectation. As Faulkner puts it, with predictive trust the trusted expects that '*something will happen*'; with affective interpersonal trust the trusted expects '*something of someone*' (Faulkner 2011: 148). Crucially, the expectation associated with predictive trust is impersonal – its emphasis is upon the likelihood of the desired outcome, not the person relied upon to bring about that outcome. As such, and as we have seen, the notion of predictive trust can be extended to relationships other than those between people – as for instance the relationship between an agent and their own epistemic practices. In contrast, the expectation associated with affective trust, as Faulkner defines it, is inherently personal – when we trust in this way, we not only rely on the people we trust to do certain things, but we expect them to do those things *because* they are the kind of person to recognise and be moved by the fact that we rely on them. Thus, echoing the quote from Holton, whilst we might be disappointed, angry, or frustrated if events don't go as we expected, the feelings of resentment and betrayal that accompany affective trust are reserved for the people who break our trust. In that respect, the model of affective trust understood in terms of the reactive attitudes associated with that stance does not naturally extend to the domain of self-trust.

More promising is Jones's emphasis upon the affective nature of the attitude held *when trusting*. For whilst, on Jones's account, the attitude of optimism takes the other person and their goodwill as its object in the context of interpersonal trust, it need not in other contexts. Instead, just as one may *believe* that X will 'reliably yield truths and not falsehoods' when X is another person or when X refers to one's own epistemic practices; so one may be *optimistic* about relying on X to 'reliably yield truths and not falsehoods' when X is another person *or* X refers to one's own epistemic

practices. In this respect, then, I would suggest it is worth exploring whether there is a mode of self-trust that can be characterised in terms of an agent's affective attitude towards their reliance on their epistemic practices.

Precedent for this move can be found in Jones's own work on self-trust, where (echoing her work on interpersonal trust) she argues that 'self-trust is best understood as a domain-relative attitude of optimism about our cognitive competence' (Jones 2012a: 243). Along similar lines, McCraw suggests that trust – including self-trust – is best characterised as an attitude of *confidence in the trusted* involving 'an affectively loaded way of seeing the one trusted (McCraw 2015: 418); and Zagzebski that self-trust involves, in addition to reliance and belief, 'a trusting emotion towards [one's] faculties in that respect' (Zagzebski 2012: 38-39). I think Jones's 'optimism' gives us the best grip on the notion of an affective mode of self-trust, so I will follow her in that usage.

### 5.2.3.1 Self-trust without belief: Affective self-trust

Echoing our earlier definition of predictive self-trust let's define *affective self-trust* such that:

Where K is amongst S's epistemic practices, S has **affective self-trust** *iff*:

- (i) S epistemically relies on K
- (ii) S is *optimistic* about relying on K

With this concept on the table, we might ask: What does optimism involve? And, how might it underwrite one's continued reliance on one's epistemic practices in a context of realisation?<sup>66</sup>

As I conceive it, optimism is a primitive attitude that manifests in an agent's willingness to act in certain ways and willingness to think well of the person, practice, or relationship to which that action corresponds. In the context of trust and reliance, optimism manifests in the willingness to rely on X to  $\Phi$  and a willingness to think well of X and one's reliance on X. To think well of X in the sense implied involves making a series of *presumptions in the context of action* that correspond to having a positive affective and cognitive attitude toward X's  $\Phi$ -ing.

What I would like to suggest first, then, is that for S to be optimistic about her reliance on K is for S to manifest a willingness to rely on that practice *under the presumptions* that (i) it will be beneficial

---

<sup>66</sup> To be clear the aim in this section is not to identify some definitive interpretation of 'optimism'. Rather, to borrow the words of Karen Jones, we should see 'optimism' as discussed in reference to trust and self-trust as a 'a term of art serving as a place-holder [...] It draws from our ordinary usage of the term "optimism" connotations of confidence and security, but it is not meant to imply a generalized tendency of the kind attributed when we say a person is optimistic' (Jones 2012a: 243).

to her epistemic aims and interests if K reliably yields truths over falsehoods and (ii) that S will accrue those benefits if she acts as if K is reliable, and so (iii) K will reliably yield truths over falsehoods.

These presumptions allow affective self-trust to play a similar role in the rational structure of the agent's attitudes as belief does in the case of predictive self-trust. Crucially though, they do not involve or imply belief – explicit or otherwise. Instead, we can think of optimism here in terms of one's having an affectively loaded way of *perceiving* or *experiencing* the world that is structured by a commitment to act in accordance with the three presumptions described.<sup>67</sup> In the words of Guido Möllering (who conceives trust in a way that fits to the current discussion of optimism):

If trust's destination is a state of favourable expectation (from which various functional consequences can follow), then its point of departure is the *experiencing* [...] of our life-world which we interpret as the reality to which our trust relates. (Möllering 2001: 412. Italics Added).

And, similarly Faulkner,

[I]f any of these propositions were not accepted, it would cease to make sense to say that A trusted because accepting this set of propositions is an expression of A's attitude of trust and so a commitment of A's decision to trust. This is not to suggest that trusting someone to do something involves explicitly committing to these propositions in one's reasoning. The claim is rather that the acceptance of these propositions partly defines how it is that the attitude of affective trust involves *seeing things in a certain light* (Faulkner 2011: 152. Italics Added).

To illuminate this initial sketch, it can be useful to think of epistemic self-trust as it stands in relation toward the inherent 'riskiness' of each of our reliance on our own epistemic practices.<sup>68</sup>

As we saw earlier, in relying upon our epistemic practices we subject ourselves to a kind of epistemic jeopardy. I.e. If S relies upon K and K reliably yield truths over falsehoods, S will stand to fulfil her fundamental epistemic aims by relying on those practices; if K is unreliable, S will stand to have those aims frustrated. But, so too S will have those aims frustrated if S somehow refrains from that reliance. In a context of doubt vis-à-vis K, S realises, or ought to realise, both the fact that she relies upon K and the vulnerability that comes with that. Thus, we might say that in such contexts S becomes (or ought to become) aware of how vulnerable her reliance on K makes her.

---

<sup>67</sup> I return to and discuss further literature on the idea that there are normative ways of engaging with the world that can be described on a broadly perceptual model in the next chapter.

<sup>68</sup> I talk of 'riskiness' and vulnerability to avoid the technical connotations of 'risk' and 'uncertainty' as those terms are used in economics and other fields (for the classical statement of the distinction, see Knight 1921: 19-20)



Correspondingly, in continuing to rely on K in such contexts by way of epistemic self-trust, S takes a stance towards the riskiness of her doing so.

As defined earlier, the attitude of predictive self-trust is belief – specifically belief that the relevant practices are reliable. As noted, then, it will only be permissible for S to continue relying on K by way of predictive self-trust if her belief that ‘K is reliable’ is justified. And, at least as we have been treating justification for belief in this thesis, S will not be justified in believing that ‘K is reliable’ if S doubts that belief is objectively justified. Thus, and to call back to chapters 2 and 3, we might say that: for S to have justified predictive self-trust in K is for S to take a *third-person theoretical* stance toward K and the riskiness of her continued reliance on that practice.

Another stance that S might take toward her reliance upon K, however, is as a relationship that she is already invested in and one –when seen positively through the light of her broader interests and aims – worth investing in despite how doing so leaves her vulnerable. In contrast to predictive self-trust, for S to take this stance toward her reliance on K, is for her to take what we might call a *first-person phenomenal* stance toward that reliance. To take this stance is for S to view her relationship to K, not through a calculation of the odds of benefitting or losing, but through a kind of affectively felt experience of the importance and value of that relationship to herself.

It is this first-person stance that S takes when she has an attitude of optimism about her reliance on K – and it is in seeing things this way that, at some level of cognition, S makes the presumption that she will benefit from relying on K if K is reliable. Again, though this presumption of conditional benefit is not belief. It is not a response to S’s assessment of the evidence or critical reflection on what is and is not to her benefit, but rather a constitutive feature of S’s own personal, subjective, and affective experience of what comes with acting as if K is reliable. At the same time as this, since the attitude of optimism is a positive attitude, for S to see her reliance in this way, involves not only the presumption of conditional benefit but also the presumption that that benefit will come her way. I.e. The presumption that if she acts as if K is reliable, she will fulfil her epistemic aims. This presumption of expected benefit, in turn, rests upon the further presumption that K is reliable. For if S is optimistic about *relying on K*, then, it cannot be so that S presumes that that benefit will come any way other than through K’s proving to be reliable. E.g. If the presumption is that luck will do the job or that by acting as if K is reliable, someone else will step in to help her out, S is not optimistic about *relying on K*. Again, though, it is not that S has come to this conclusion via cost-benefit analysis, or assessment of the evidence that K is reliable. Instead, as Jones puts it, given S’s positive and affective investment in her reliance on K: ‘the situation is experienced as *cognitively safe*’ (Jones 2012b: 5. Italics Added).

So much for the attitude of affective self-trust. How does this relate to our broader discussion, i.e. how might affective self-trust underwrite reliance on one's practices in a context of doubt. There are three important points to bring out.

First, in so far as optimism manifests in a willingness to act, affective self-trust entails a propensity for action. More specifically since we are talking about *epistemic* self-trust, affective self-trust entails that, other things being equal, if K in some way indicates that Q, one will believe Q. In this way, if S believes Q because S has affective self-trust in K and K in some way indicates that Q, S does not believe Q *for the reason that* she has affective self-trust. She believes Q because believing Q in the circumstances is constitutive of what it is to manifest a willingness to rely on K – or in other words, it is constitutive of what it is to have affective self-trust. Thus, whilst we shall see that affective self-trust has an internal structure of reason, affective self-trust does not provide S new justification for believing Q as we saw justified predictive self-trust to do. Rather, we might say, affective self-trust can restore one's access to whatever justification one previously had for the beliefs formed via the relevant practices.

Second, in so far as optimism involves making the presumption that K is reliable, affective self-trust precludes doubting that K is reliable. To refer back to Jones's description, to be optimistic about relying on K is to experience the situations in which one's reliance on K comes to the fore as *cognitively safe* – and so to put aside or bracket the riskiness and vulnerability that are otherwise salient in contexts of doubt. This is not to say that the self-trusting person is ignorant or blind to their vulnerability – that way would be foolishness, arrogance, or over-confidence in one's practices, not optimism. Instead, to borrow from Möllering, this way of seeing 'can be defined as the mechanism that brackets out uncertainty and ignorance, thus making interpretative knowledge momentarily 'certain' and enabling the leap to favourable (or unfavourable) expectation.' (Möllering 2001: 414).<sup>69</sup>

The self-trusting person, thus, is not unaware of the riskiness of their relationship, it is just that these features of the predicament become less salient when that relationship is seen through the lens of optimism. Attention is directed toward the value of one's continued reliance; value that can only come under the presumption that the practices on which one relies on are reliable.

In respect to this aspect of affective self-trust, we can observe that S might have affective self-trust in K whilst in a context of doubt without contravening PAR. For confronted with the question:

---

<sup>69</sup> Möllering refers to this dimension of trust as 'suspension'.

'Is K reliable?' S's optimism suffices to close that question, just because that attitude involves the presumption that K is reliable. Moreover, since it also involves the presumptions of benefit, S's optimism provides a practical explanation of why S continues to rely on K in the context. Affective self-trust we might say is self-rationalizing. Faulkner's description of the rationality of affective interpersonal trust illustrates this thought nicely:

In trust one lowers one's guard because the possibilities of one's trust being betrayed are excluded by what must be accepted in holding an attitude of trust. This background of acceptance then specifies a way of thinking about the trust situation, which yields the presumption that the trusted is trustworthy that provides a reason for trusting. (Faulkner 2011: 153).

Much the same can be said of affective self-trust.<sup>70</sup> Since affective self-trust is self-rationalizing, however, it follows that it is not subject to the problems of epistemic circularity that we saw were associated with predictive self-trust in at least some contexts. This is because those problems arise when one attempts to justify belief in the reliability of the relevant practices – in the sense of finding reason to form such an attitude, or reason to continue to hold such an attitude. With affective self-trust there is no such act of justification, or practice of finding reasons. Rather, affective self-trust comes as a package – to have affective self-trust is to have reason for that trust built in. The question that remains, then, is not whether one can have reason to have trust oneself in the affective sense in a context of doubt – one can – but whether one has good reason to trust oneself?

Which brings us to the third and final point: we do not and should not have affective self-trust *come-what-may*. Where predictive self-trust requires justification, whether one has good reason to trust oneself in the affective sense is a matter of whether such trust is *available* and whether it is *well-calibrated* to one's epistemic experiences – or more simply, whether it is *well-placed*.

Whether affective self-trust is *available* will depend upon the cognitive and affective character of the agent in question, as well as the circumstances they find and have found themselves in. It depends on the character of the agent, because some people are not inclined psychologically – whether by nature or as a result of experience – to see things in the positive affective light that affective self-trust involves. The availability of affective self-trust also depends on the agent's character because to *see* or *experience* one's reliance on one's own practices in the way characteristic of affective self-trust requires a degree of sensitivity to epistemic considerations, and not all people

---

<sup>70</sup> Though, as noted earlier, it is questionable whether the attitude of betrayal has any relationship to self-trust – frustration or disappointed might be better terms in the current context.

show that sensitivity in all circumstances. (Just as the availability of predictive self-trust depends upon the agent having a certain capacity for critical and reflective reasoning.) The availability of self-trust depends on *circumstance* just because circumstances make a difference to how we see and experience the situations we are in. Whilst the optimism that is characteristic of self-trust can shape one's experience of a situation such that one's attention is elsewhere than its riskiness, as circumstances change, one's experience of that situation as 'cognitively safe' *may* recede and one's attention shift toward the sources of doubt and vulnerability. This may happen, for instance, if one is faced with persistent and fervent disagreement – then, again, it may not – that same situation may be the catalyst for someone else to find new levels of self-trust. This points to a crucial point about affective self-trust. For, whilst it is to be expected that one should be sensitive to such changes in situation, how that will play out in a particular case, I would argue, is not something that we can predict in abstraction from that case. In short, whilst we can point to its broad determiners, the ways in which character and circumstance can interact to render affective self-trust available or unavailable are too various to give a principled account affective self-trusts availability.

This brings us to the question of why we *should* not have affective self-trust *come-what-may*. The answer to this question is simple enough. We should not have affective self-trust come what may for the simple reason that the optimism that is characteristic of such trust can be misplaced – as it will be if one trusts K when K is not reliable. Since the availability of self-trust is dependent on an irreducible commixture of character and circumstance, however, we cannot be expected to trust only when K performs as we hope it too. Instead, what we can hope for and expect is that one's trusting experience of situations are, over time, *well-calibrated* to one's actual experience of the functioning of K in those situations. In general, well-calibrated self-trust will track one's actual experience of the functioning of K – but what this amounts to will depend upon the nature of those experiences as well as the agent's own sensitivity to the various features of that experience relevant to their trusting. As with availability, what it takes for self-trust to be well-calibrated is too dependent on the features of an agent's particular history of experience and circumstance to be amenable to any general, principled account of that.

I shall finish the discussion of affective self-trust here, the question that remains is how this relates back to our discussion of higher-order evidence and disagreement. In line with the points above, I shall not attempt to offer a principled account of this. Instead, I shall attempt to elucidate the role of affective self-trust in contexts of disagreement and doubt by way of an example of ordinary peer disagreement in which such considerations are particularly pertinent.

## 5.2.4 Affective self-trust and disagreement: an example

### THE TOXIC BULLROOM

Kazimira has recently been promoted to the senior editorial team on a major international newspaper. One aspect of this new role is to participate in meetings where the paper's editorial positions are set out, argued for, and decided upon. Though experienced, well-qualified, and respected in the wider journalistic field, Kazimira is finding many of her otherwise well-founded assertions knocked back by the other members of the team. 'That's the wrong take', 'I think you've misread the climate', 'your biases are showing' are not unheard-of reactions. All in all, Kazimira's suggestions are met with disagreement an inordinate amount of the time.

All of the other members of the editorial team have similar or greater experience and qualifications to Kazimira. From prior experience Kazimira has developed a keen sense of respect for everyone on the team. As it happens, though, Kazimira is the only woman and the only non-English journalist currently sitting in on the meetings. And, unbeknownst to her and everyone else, the environment is amplifying the underlying biases and prejudices of the men in the room regarding the journalistic credentials of women and non-English speakers. However, Kazimira has no obvious evidence of this aside from her disagreements. Indeed, the paper is a progressive, left-leaning publication and the various writers in the room have reported on and contributed to numerous progressive movements, including the broadly feminist stance the newspaper has cultivated in recent years.

Initially, Kazimira's response to the situation is to accept the 'criticisms' of her colleagues and stoically return to the metaphorical drawing board to come up with better ideas, research the issues, seek out other viewpoints and so on. As these negative experiences pile up, however, she begins to have a change of heart and approach to the situation. What started as a dispiriting and frustrating sense of her own inadequacy, morphs into a more defiant sense of anger and injustice. Instead of doubting herself, and with increasing confidence, Kazimira becomes increasingly steadfast in her own opinions and doubtful of her colleagues'.

In this case, Kazimira is subject to an instance of *testimonial injustice* that is manifest in the context of persistent disagreement.<sup>71</sup> What is particularly pernicious about this case is that, as far as Kazimira can tell, the people who disagree with her so persistently are either her effective peers or indeed superiors. Since, though, the culture of disagreement that is aimed at Kazimira is shaped by bias and prejudice, the apparent authority with which the others speak is anything but. Again, though, Kazimira has no evidence of this aside from her experience of persistent disagreement itself. Nonetheless, I would suggest that given the features of the case it is (i) not implausible that Kazimira would respond in this way (though it would also be understandable were she never make the leap to trusting herself) and (ii) that Kazimira remains entirely within the limits of epistemic rationality in responding in the way that she does. Even if she never acquires any evidence of the real source of the reactions her opinions receive, apart from the experience of disagreement itself, it seems to me that there comes a tipping point where Kazimira should stop being conciliatory and begin to trust herself instead.

Without recognising the role of affective self-trust, we cannot explain how this could be so. Since the group consists of her effective peers and superiors, Kazimira's experience of disagreement affords her a significant body of higher-order evidence. And of the kind in light of which, in lieu of a defeater-defeater, Kazimira ought to not only revise her beliefs but accede to the group entirely. What is more, in so far as her opinions are consistently subject to disagreement from the group, in lieu of a defeater-defeater, Kazimira ought at some point to start questioning her general competence in the relevant fields. To refer back to the discussion of kinds of defeater, each single disagreement may be a non-contextual undercutting defeater. But the cumulative higher-order evidence at some point shifts into being a general undercutting defeater as a more general explanation of the persistent disagreement becomes more salient. As the case goes, however, the only evidence that Kazimira has in her favour is the history of disagreement itself. And that being so Kazimira will have no first-order defeater-defeaters for the higher-order evidence, nor will she have any kind of defeater-defeater by way of justified predictive self-trust (for just the reason that, without affective self-trust, she won't be justified in having predictive self-trust). Thus, without recognising the possibility for affective self-trust to play a role in one's response to disagreement,

---

<sup>71</sup> Testimonial injustice is one of two kinds of epistemic injustice that Fricker describes in her seminal work on the topic 'Epistemic Injustice: Power and the Ethics of Knowing' (2007). Testimonial injustice occurs when a speaker is granted less credibility than they merit because of the operation of some kind of identify-prejudice on the part of the hearer. Important contributions to the growing literature on this and other varieties of epistemic injustice include Dotson (2011, 2012, 2014) and Medina (2013), as well as Jones's (2012a, 2012b) work on self-trust that I have been drawing on in my discussion of affective self-trust. For recent collections on epistemic injustice and related topics see Kidd and Medina (2017) and Barker, Crerar, and Goetze (2018)

the theorist is forced to accept that in this case rationality demands that Kazimira both accede on the matters under dispute *and* become significantly less trusting of herself in general. This is so, even though it is Kazimira's colleagues who are in the wrong, epistemically and morally, and even though she has first-hand – if *evidentially* hidden – experience of how their biases and prejudices play out. This diagnosis of the case, it is worth noting is not entirely off base, as Karen Jones points out, epistemically unjust encounters are often characterised by such effects:

[U]njust social relations cause epistemic injustice, which undermines self-trust among the underprivileged; unjust social relations create excessive self-trust among the privileged, which perpetuates epistemic injustice, which further undermines the self-trust of the disadvantaged in a vicious feedback loop. The pressure epistemic injustice places on self-trust, however, is not only causal, it is also a *rational* pressure. In our shared practices of inquiry we are counting on each other not only for information with respect to whether or not p, but also for ongoing mutual epistemic calibration [...] The social calibration of our knowledge-seeking strategies puts rational pressure on us to conclude we have made a mistake when we cannot explain divergent judgment. (Jones 2012a: 247).

In line with these points, we should recognise that it is quite possible that Kazimira would react to the experience described with deference and doubt – that is a part of the nature of epistemic injustice. Moreover, it is possible that she would be rational in doing so – be that a pernicious and destructive consequence of rationality, again, it is a part of the nature of epistemic injustice. However, in recognising the role that affective self-trust can have in responding to disagreement we can make sense of how Kazimira could also respond in the way this case is described – and how her remaining steadfast in the face of unjust disagreement might be rational, even when she has no independent evidence that that is the case.

First, the concept of affective self-trust allows us to explain Kazimira's gradual shift from deference to steadfastness in terms other than stubbornness or dogmatism. For, Kazimira does not make a sudden shift from basing her responses on the evidence to dogmatically resisting her colleague's protestations. Instead, her perceptions of herself, her colleagues, and the situation undergo a gradual recalibration; from one of an externally imposed humility to a defiant optimism-in-resistance. Once that optimism begins to manifest, Kazimira eventually comes to *see* the situation in a way that is structured by the presumption that her own practices are reliable. That quasi-perceptual experience in turn can ground her reasoning about the situation in a way that brackets out the doubts she might otherwise have about her own practices. The shift from doubt to optimism (about herself), thus, opens up new rational possibilities for Kazimira as it provides an internally consistent platform from which she might explore explanations of the history of disagreement that take the fact that her colleague's positions are often wrong and poorly supported

as its starting point. In that way, we might say, as she comes to have affective self-trust Kazimira acquires a kind of *higher-order defeater-defeater* for her experiences of disagreement – one that allows her to rely upon the facts of the disagreement itself as the rational grounds for her staying steadfast.

Whether this defeater-defeater is good, will depend, not only upon whether Kazimira has such self-trust, but whether it is well-placed. In this case, I would suggest that it is. Kazimira's colleagues obstinance to her ideas is fuelled by prejudice and bias, not by critical reflection on the issues under debate. Kazimira's own opinions, as stipulated, are well-founded and the produce of her own competence. If Kazimira *did* have independent evidence of these features of the disagreement, she would clearly be justified in downgrading her assessments of her colleagues' competence. In that respect, her eventual trust in herself over her opponents is well-placed. What is more, as we saw, when Kazimira does manifest that trust in herself that shift is not the result of a sudden and inexplicable shift from deference to dogmatism. Rather, as noted, it is the result of a gradual re-calibration of her epistemic sensibilities. Moreover, that re-calibration reflects epistemically significant features of the situation that Kazimira is directly subject to and of which she has immediate experience – even if they are initially hidden beneath the structure of explicable reasons. Even if the only evidence that Kazimira has in her favour is the history of disagreement, then, it seems to me entirely apt, that as a competent and sensitive epistemic agent, Kazimira would come to trust herself over her colleagues in this situation.

## Conclusion

In discussing the example of testimonially unjust disagreement in the previous section, I hope to have shed some light on how affective self-trust as I have described it can relate to disagreement and the broader discussion of higher-order evidence. I hope also to have illustrated how the presence of such self-trust can help to explain how the steadfast response can be a rational possibility in case of disagreement between effective peers – as well as how an absence of such self-trust can take that possibility off the table. In this light, I alluded to the idea that affective self-trust might be seen to give one a kind of *higher-order defeater-defeater* – of the kind that we were looking for at the start of this discussion. Given this, it seems apt to refer to this approach to 'The Problem of Ordinary Disagreement' as the 'Higher-Order Trust' approach to disagreement. In the conclusion to this thesis I shall lay out the Higher-Order Trust approach in more schematic form. Before doing that, what remains to be seen is whether this approach might also be extended to 'The Normative Problem of Deep Disagreement'. I shall consider that in the next chapter.



# Chapter 6. The Higher-Order Trust Approach and Deep Disagreement

## Introduction

In the final chapter of the thesis, I extend the Higher-Order Trust approach to disagreement developed in Chapter 4 and 5 to ‘The Normative Problem of Deep Disagreement’. In the first part of the chapter, I re-introduce ‘The Normative Problem of Deep Disagreement’ and discuss the possible theoretical responses to that problem. In the second part of the chapter, I situate the notion of ‘epistemic norms’ within the broader understanding of epistemic practices and self-trust forwarded in the previous chapter. In the final section of the chapter, I consider in more detail what the Higher-Order Trust account has to say about how those party to deep disagreement between effective peers ought to respond to that in two types of case: cases in which disputants are not in a position to recognise the nature of their disagreement, and cases in which disputants are in a position to recognise the nature of their disagreement.

## 6.1 Re-introducing The Normative Problem of Deep Disagreement

As per the discussion in Chapter 1, ‘The Normative Problem of Deep Disagreement’ concerns the question of how, epistemically speaking, one ought to respond to the realisation of disagreement when the other parties to the disagreement are one’s effective peers and when the disagreement is deep.

‘The Normative Problem of Deep Disagreement’ is distinct from ‘The Problem of Ordinary Disagreement’ in so far as deep disagreements involve more than a difference in beliefs, they also involve a difference in epistemic norms. More specifically, I characterised deep disagreements as occurring when a disagreement at the level of belief is explained by the disputants following different epistemic norms in coming to hold those beliefs. We glossed epistemic norms as corresponding to rule-like statements of the following form:

*Where  $\Phi$  designates some normative epistemic act such as believing, inferring, trusting, and so on,*

Under condition C, one ought to/ought not to/is permitted to  $\Phi$

Or, imperatives of the form:

If C,  $\Phi$ ! / don't  $\Phi$ !

And we said that, for someone to follow an epistemic norm is, roughly, for that person to regulate the relevant epistemic actions in accordance with what they perceive to be the prescriptions of that norm in different circumstances. Following a norm, I suggested, then, is (i) something an agent does, not necessarily a matter of their holding any beliefs about that norm and (ii) constitutively connected to what an agent treats as epistemic reasons. I shall return to the topic of what epistemic norms are and what is involved in following an epistemic norm in 6.3, but this rough characterisation will make do for now.

To remind us of what a deep disagreement might look like in practice, here again is the example from Chapter 1:

#### POLITICIANS

M and Prudence are politicians in rival political parties. A recent glut of polls from reputable pollsters indicates that the public have an extremely low level of trust in the political class. Both Fax and Prudence have sincerely professed that they think it crucial to their political roles to act in such a way as to engender just the kind of public trust the polls refer to. Additionally, both Fax and Prudence recognise that, were they not personally trusted by the public, engendering such trust would require considerable effort and expenditure on their behalf. Shortly after the release of yet another poll indicating the dearth of public trust, Fax asks Prudence what she thinks about the public-trust problem. Prudence replies 'What problem?' Surprised, Fax asks Prudence if she has seen the recent polls. Surely, Fax says, given the numbers 'It is quite clear that the public do not trust either of us.' Prudence responds: 'Of course I've seen the polls and they definitely spell trouble for you. Fortunately for me, the public trust me implicitly,' before adding, with no apparent irony or insincerity, 'I mean, if they didn't, I would have to change the whole way I go about doing politics!'

As it happens, Fax based her beliefs about who the public trust on the polls alone; Prudence treats both polls and practical considerations as relevant to determining who the public do and don't trust. Since Prudence would incur a greater cost were she to act as if the public do not trust her, Prudence determined that the poll results must not be evidence of how the public feel about her. What is more, this difference in how the pair came to their beliefs represents a quite general feature of their epistemic practices – Prudence sees practical considerations as generally relevant to determining what is true, Fax does not.

In this case, Fax and Prudence disagree about the proposition ‘The public do not trust either of us.’ Fax and Prudence are both aware of the polls that show the public’s general trust in politicians has declined, as well as the practical costs that they will incur if the evidence of the polls applies to themselves. Moreover, they concur that the poll results are relevant to the question of who the public trust. Where the two differ, and what makes this a case of deep disagreement, is in the way they perceive the relevance of the practical considerations. As I put it in Chapter 1, Fax distinguishes between the practical and epistemic reasons, whereas, for all intents and purposes, Prudence treats practical reasons *as* epistemic reasons. Given that this difference represents quite general features of their epistemic practices, then, I suggested that we could characterise this disagreement as one in which:

Fax believes ‘the public do not trust either of us’ *because* Fax follows a norm such that:

If the only considerations that support your believing Q are practical considerations, then, for the sake of acquiring true beliefs, you ought not to believe Q.

And, Prudence disagrees with Fax *because* Prudence follows a norm such that:

If the only considerations that support your believing Q are practical considerations, then, for the sake of acquiring true beliefs, it is permissible for you to believe Q.

Having illustrated the idea of deep disagreement, I suggested that it was also important to consider the question of whether and when those party to a deep disagreement might be considered effective peers. I suggested that there are two important points to recognise in respect to this question. First, someone might be one’s effective peer in case of deep disagreement where they would not be so if it were not for the difference in norms. Second, and following that, I suggested that, since one may not be aware of the difference in norms, someone might be one’s effective peer in case of deep disagreement without one having the kind of evidence that would allow one to recognise that they are one’s effective peer. *POLITICIANS* illustrates both points.

Were *POLITICIANS* an ordinary disagreement, and so Fax and Prudence shared the same norms, Prudence would not be Fax’s effective peer. This is because when Fax asks Prudence how she could not think that the pair *both* face a public trust problem, Prudence close to as admits that she has been swayed by the practical considerations. On Fax’s norms, however, Prudence’s ‘admission’ would afford Fax independent reason to believe that Prudence had made a mistake in her

reasoning. On Prudence's norms, on the other hand, Prudence's admission would afford Fax independent reason to believe that she had made a mistake in her reasoning (i.e. she ignored the practical considerations). Thus, in either case, Fax's possession of the evidence about Prudence's reasoning would be sufficient to prevent Prudence from being Fax's effective peer.

Since POLITICIANS is a deep disagreement, however, things are quite different. For, not only do Fax and Prudence follow different epistemic norms, but since their respective norms determine the kind of considerations that each sees as providing reasons for belief, that difference cannot be separated from the substance of the disagreement. That being so, any justification that Fax might otherwise have for believing that she has the relative advantage, which would not count as such on Prudence's norms, will fail to be independent of the substance of the disagreement (similarly for any justification Fax might otherwise have believing that Prudence has the advantage). Thus, since the only evidence Fax has in her favour is the evidence that Prudence was swayed by the practical considerations, and the difference in Fax's and Prudence's norms concerns just whether it is permissible to be swayed by practical considerations; Prudence is in fact Fax's effective peer – even though she would not be were this an ordinary disagreement,

This brings us to the second point. For, Prudence is Fax's effective peer only because the brute facts about the norms they follow makes her so. And as a consequence of this, Fax cannot recognise that Prudence is her effective peer solely by reflection on the evidence she possesses as the case is described. Instead, she would have to acquire some additional evidence that would indicate to her that any consideration that she might suppose to favour her is likely to be disputed by Prudence. Given how fundamental is the difference in Fax's and Prudence's norms, this would appear to require either that Fax submit a wide range of her background beliefs to the scrutiny of Prudence, or that Fax have more direct evidence about the fundamental nature of her disagreement with Prudence. Neither, I would suggest, are necessarily easy to acquire, even with a good deal of further dialogue between the pair. In particular, recognising that the disagreement is deep would require that Fax be capable of thinking about epistemic norms – i.e. conceptualising behaviour in terms of them and recognising behaviour as following them – and it is doubtful that Fax could do this. Not only is Fax not currently in a position to recognise that Prudence is her effective peer – but, quite possibly, such recognition may be beyond what it is reasonable to expect her to achieve.

POLITICIANS, thus, illustrates the core points about the relationship between deep disagreement and the status of effective peerhood: (i) those party to deep disagreement can be effective peers; (ii) those party to deep disagreements can be effective peers when they would not be were they to

share the same epistemic norms, and (iii) those party to deep disagreement can be effective peers even when they are not in a position to recognise that they are effective peers.

In section 6.4, I shall consider what significance these features of deep disagreement have upon ‘The Normative Problem of Deep Disagreement’. Before doing that, it will be useful to summarise the potential theoretical positions on the question of how those party to deep disagreement ought to respond when they are effective peers. And, following that, it will be useful to say a little more about the nature of epistemic norms.

## 6.2 Theoretical positions on the problem of deep disagreement

Since deep disagreements are constituted by differences at two levels (i.e. the level of norms and the level of beliefs), the ways in which disputants might respond to deep disagreement are complicated in a way that they are not in ordinary disagreement. In the case of ordinary disagreement, the only theoretical question is whether disputants ought to revise or stick to their respective beliefs about the disputed propositions. In the case of deep disagreement, there are two (initial) theoretical questions: Should the disputants revise or stick to their original beliefs? And should the disputants continue to follow their respective norms? In contrast to the two possible responses to ordinary disagreement, then, there are four possibilities for how disputants might respond to deep disagreement:

- (a) Stick by norms and retain beliefs
- (b) Stick by norms but revise beliefs
- (c) Abstain from norms but retain beliefs
- (d) Abstain from norms and revise beliefs.

As far as theoretical positions go, we can immediately rule out (c) – in lieu of some other source of justification for believing as they do, it will not be rational for someone engaged in a deep disagreement to abstain from their norms yet retain the beliefs formed in accordance with those norms. A similar line of reasoning to that underpinning the discussion of self-trust in the last chapter will help us to see why. Consider Fax in *POLITICIANS*. Any justification Fax has for believing ‘the public do not trust either of us’ is dependent on her being permitted to follow the norms by which she came to hold that belief. Fax’s being justified in holding the belief, thus, entails that she is permitted to follow the relevant norms. Correspondingly, if Fax ought to abstain from the norms by which she came to believe ‘the public do not trust either of us’ Fax will lose any

justification she had for believing as such by way of those norms. Thus, in lieu of acquiring justification for believing ‘the public do not trust either of us’ from some other source, it cannot be the case that Fax ought to abstain from her norms and retain her beliefs. In other words, (c) is not a theoretically viable response to the disagreement. More generally, this helps us to recognise the following principle:

Total Revision (TR): If, in light of deep disagreement, S ought to abstain from her epistemic norm N, then, for any belief Q that S came to hold via N, other things being equal, S ought to revise her belief Q.

I shall return to the significance of this principle shortly. First, let’s consider the plausibility of (b). Like (c), (b) calls for a different response at the levels of norms and beliefs (i.e. stick by norms, revise beliefs). As I shall argue in the later sections of this chapter, however, unlike (c), (b) will be appropriate in some cases of deep disagreement. I will say more on which cases those are later in the chapter, for the moment the important point is to recognise that the possibility of this response rests on the fact that epistemic norms properly understood are defeasible. Indeed, one of the lessons of the possibility of defeat from higher-order evidence is that, no matter how competently one may have formed one’s beliefs, no justification is inviolable. To say that norms are defeasible, though, is just to say that one can be in situation where one has no reason to abstain from one’s norms in any general fashion, but one ought to withhold from following the prescriptions of those norms in the specific context. The mixed response at (b) should be understood in this way. And, to jump ahead in the discussion, it would be rational for me to respond to deep disagreement in this way, if I have reason to distrust the practices by which I came to hold the disputed beliefs, but not specifically to doubt the norms by which I came to hold those beliefs. As I shall argue later in the chapter, this will sometimes – and perhaps often – be the case with deep disagreement. More specifically, this will be the correct response in cases where the disputants are not in a position to recognise that the disagreement is deep yet would be effective peers if the disagreement were not deep.

Given these points, then, (a), (d), and (b) – properly understood – are all theoretically viable responses. Let’s call these the *steadfast response* to deep disagreement, the *conciliatory response* to deep disagreement, and the *mixed response* to deep disagreement. As with ordinary disagreement, we can categorise the possible theoretical views of deep disagreement in terms of their attitude to these different responses. Corresponding to uniform positions on ‘The Problem of Ordinary Disagreement’, one might offer a ‘uniform-steadfast account of deep disagreement’ or a ‘uniform-conciliationist account of deep disagreement’ (i.e. the appropriate response is always to stick by

norms and beliefs, or always to abstain from norms and revise beliefs). Corresponding to non-uniform positions on ‘The Problem of Ordinary Disagreement’, one might offer a ‘non-uniform account of deep disagreement’, whereby, which of the steadfast, conciliatory, or mixed responses is appropriate will depend upon the circumstances of the particular disagreement. In the rest of the chapter I shall argue that the Higher-Order Trust approach to disagreement developed in the previous two chapters – which is a non-uniform account of disagreement – can be extended to deep disagreement.

Before doing so, it is worth considering some motivations we might have for endorsing a non-uniform account of deep disagreement independently of the merits of the Higher-Order Trust approach.

Consider the principle of Total Revision. On a uniform-conciliationist account of deep disagreement, this principle would apply in all cases of deep disagreement between effective peers. In a case like POLITICIANS, however, this would have sceptical consequences at a near global level. Why? Because, whilst the disagreement between Fax and Prudence occurs in the political domain, the norms that explain that disagreement are not specific to that domain. As described, Fax’s norms forbid her from placing weight upon the practical considerations in any and all epistemic deliberations; Prudence’s norms permit her to place weight on the practical considerations in any and all epistemic deliberations. Thus, if Conciliationism is the right response, both Fax and Prudence should withhold judgement on any and all beliefs in which they, respectively, omitted or included practical considerations in their epistemic deliberations. Given that the vast majority of our beliefs will relate in some way or other to our practical interests, we might expect that this would require Fax and Prudence to revise the vast majority of their beliefs.

Of course, it is not a surprise that a case like POLITICIANS could have such global consequences. Not only are the differences between Fax’s and Prudence’s norms fundamental, but they are such that a follower of Prudence’s norm is likely to disagree widely with the follower of the Fax’s norms. After all, except for a lucky few, the world we live in does not consistently track our practical desires and interests – thus, someone who provisions their beliefs according to what is in their interests is likely to disagree widely with someone who takes, dare I say, a more realistic view on things. Given that we might expect there to be wide swathes of disagreement between Fax and Prudence, it is not a surprise that a conciliationist response to the disagreement between the pair would have similarly wide sceptical consequences.

I make these points here, since much of the discussion in the literature on deep disagreement has focused on cases similar to POLITICIANS. As we saw in Chapter 1, common examples include disagreement between practitioners of alternative and conventional medicine, those who follow scientific principles of inquiry and those who follow the authority of religious texts, people on opposite sides of the conservative/progressive political divide, and so on. All these, though, are examples in which we might expect the rival norms to be both general in scope and cover a range of issues wide enough that any sceptical consequences of the disagreement will be near-global. An additional feature that is common to these cases, including POLITICIANS – and one that I think stems from just the kinds of sceptical worries at issue – is that intuitively at least one party in these cases *ought to* retain both their norms and their beliefs. I.e. These are cases in which, for at least one party, the steadfast response to deep disagreement is intuitively the correct response. By focusing primarily on such cases, however, there is a danger that we are too easily tempted into endorsing, not just the steadfast response in the particular case, but a uniform-steadfast account of deep disagreement. Too readily giving into that temptation, I think, would be a mistake. For, the differences between those party to deep disagreement are not always so extreme as in POLITICIANS or the various other examples listed. Indeed, it strikes me that many of the most interesting examples of deep disagreement will be ones in which the disputants are in many ways, so to speak, following the same rulebook, yet are embroiled in deep disagreement nonetheless. In these cases, it is far less clear that either side to the dispute should stay steadfast. Nonetheless, such disagreements can be just as intractable and resolving them just as problematic as the cases typically discussed. To illustrate those thoughts, consider the following variant on the by now familiar peer disagreement between Doctors Quinn and Zaius:

#### DOCTORS-2

Doctor Quinn is treating a patient with Condition-X. Quinn is trained and experienced in the relevant field and has assessed the available case-evidence – including patient history and test results. Quinn has also checked and doublechecked the relevant medical guidelines. Following her assessment of the case-evidence and guidelines, Quinn believes that the patient is more likely to recover under treatment option-A. Quinn then comes to realise that a second doctor in the case, Doctor Zaius, having considered the same case-evidence and guidelines, believes that the patient is most likely to recover under treatment option-B. Quinn has every reason to believe that Zaius is as well trained and experienced as she.

Whilst Quinn and Zaius based their conflicting beliefs on the same body of clinical and scientific evidence, however, there are some curious differences in their underlying approach



to that evidence. Quinn and Zaius agreed that the evidence for this case closely fits the conditions under which the medical guidelines for Condition-X recommend treatment option-B. However, Quinn felt that there was something about the evidence in this case that did not sit well with the advice in the guidelines. It was this subjective judgement that guided Quinn to her conclude that, given the case-evidence, treatment option A was best. Interestingly, Zaius shared Quinn's feeling about the case-evidence but dismissed this in favour of more strictly following the medical guidelines.

As it happens, Doctor Zaius trained at a top medical college in which students were taught to base their medical beliefs only upon 'objective and high-quality scientific and clinical evidence'. In adhering strictly to the recommendations contained within the 'evidence-based' medical guidelines, Zaius was following his training to the letter. Doctor Quinn, on the other hand, trained at a top medical college in which students were taught to base their medical beliefs primarily upon consideration of high-quality scientific and clinical evidence, but to also factor their 'clinical judgement' into assessment of each case. Crucially, Quinn's clinical judgement sometimes, though not often, leads her to forming beliefs that differ from the relevant guidelines. This is one such case.

In DOCTORS-2, as per the original example DOCTORS-1, Quinn and Zaius disagree about the truth of a narrow set of propositions, i.e. 'A is best' and 'B is best'. Here, though, the disagreement is not explained by their having different evidence about the case, referring to different guidelines, or by any personal error in performance the assessing the case-evidence and guidelines. Instead the roots of the disagreement lie in the difference in how each weighted the case-evidence in light of the information in the medical guidelines and their own *clinical judgement*. What is more, in doing so, each was following the general understanding of how medical beliefs ought to be formed that they had been exposed to during their medical training. In this respect, I would suggest, the difference in how the pair came to their beliefs can be characterized in terms of epistemic norms.

So, we might say, Quinn follows a norm that can be characterised as:

If the available clinical and scientific evidence and your clinical judgement together indicate that Q, you ought to believe Q; if not, not.

Zaius follows a norm that can be characterised as:

If the available clinical and scientific evidence indicate that Q, you ought to believe Q; if not, not.

Since it is the difference in these norms that explains the disagreement, similarly to POLITICIANS, DOCTORS-2 counts as a deep disagreement as I have defined them.

Unlike POLITICIANS, however, Quinn's and Zaius's norms do not prescribe how one ought to or ought not to form beliefs in all circumstances – but only when deliberating on the likely truth of propositions that fall within the medical domain. In this respect, then, we might say that this is a *domain specific deep disagreement*. Also, in contrast to POLITICIANS, Quinn and Zaius would be effective peers were this an ordinary disagreement. After all, apart from the additional wrinkle concerning the difference in norms, this case is identical to DOCTORS-1. Furthermore, I would suggest, given the content of their respective norms, this is much as we would expect. There are two aspects to this. Firstly, though allowing space for clinical judgement, Quinn's norm still mandates that one place weight on the clinical and scientific evidence. Secondly, and more significantly, for Quinn to competently follow her norm, she will have had to have honed her clinical judgement through medical training and practice. More specifically, she will have honed it – even if not explicitly – against the standards and norms operative in the medical community. These will include the standards explicitly formulated in medical guidelines, as well as those explicitly or implicitly instantiated in the practice of other qualified medical professionals, including someone like Zaius. Thus, one would *not* expect a competent follower of Quinn's norms to widely disagree on medical matters with a competent follower of Zaius's norms.

What is the significance of these differences?

I would suggest that the possibility for deep disagreements that are (i) domain specific and (ii) between interlocutors who we might otherwise expect to be effective peers, undercuts the motivations that cases like POLITICIANS provide for a uniform steadfast response to deep disagreement. This raises the question: Is it possible to that the appropriate response for Fax in POLITICIANS is to be steadfast, whilst the appropriate response for Quinn and Zaius in DOCTORS-2 is to be conciliatory?

Lending credence to this possibility are two further features of this case. For one thing, a conciliationist response to a domain specific deep disagreement – even granting the principle of Total Revision – would not have the global sceptical consequences of a case like POLITICIANS. The anti-sceptical motivation for a uniform steadfast view of deep disagreement stems from the thought that, if all cases of deep disagreement are like POLITICIANS, then, given the global sceptical consequences of conciliating in such cases, allowing that *any* deep disagreements require a conciliatory response would be to cede the mountain to the sceptic. However, the possibility of

domain-specific deep disagreements undercuts this anti-sceptical motivation. Granted the consequences of conciliating for, say, Quinn and Zaius would still be extensive and would still entail revising all beliefs formed via the respective norms – but they would not be global. Thus, we might allow the possibility that the conciliatory response is appropriate in some cases without that entailing that all cases of deep disagreement between effective peers have globally sceptical consequences.

For another thing, the possibility that the normative perspective of those party to a deep disagreement might be sufficiently close that they could view each other as effective peers undercuts the intuition fuelled by cases like *POLITICIANS* that at least one party to the disagreement ought not to harbour any doubts about their practices. In cases like *POLITICIANS*, it is appealing to think that someone in Fax's positions should not doubt her practices for the reason that, by her lights, Prudence is unlikely to seem in anyway a competent agent. But such intuitions do not hold up in a case like *DOCTORS-2*. As noted, in this case, despite their conducting their medical practices according to conflicting epistemic norms, there is no reason to suppose that either party (in lieu of specific evidence about the differences in their epistemic practices) would view the other party as incompetent. But that being so, the intuition that, say, Quinn ought not to question herself in respect to the practices by which she came to believe 'A is best' simply because Zaius follows different norms has little traction at all.

In respect to cases like these, then, I think we need to take seriously the thought that the steadfast response will not be appropriate in all cases of deep disagreement. Instead, as I shall argue in the next two sections, we would do better to endorse a non-uniform account of deep disagreement. More specifically, I will argue that we can extend the Higher-Order Trust based approach to ordinary disagreement developed in the previous two chapters to also accommodate deep disagreement. To do so, it will pay first to consider in more detail what it means to say that a disagreement is explained by the disputants following different norms.

## 6.3 Epistemic norms as social practices

In this section, I discuss two aspects of epistemic norms and norm-following. First, I suggest that the epistemic norms that an agent follows can be understood in terms of the discussion of epistemic practices and self-trust developed in the previous chapter. Second, I forward an explanation of how most of us engage in such practices.

### 6.3.1 Epistemic norms and epistemic practices

As described in the previous chapter, the term ‘epistemic practices’ refers to the gamut of ways in which we come to acquire beliefs. For any belief that an agent holds, we saw, there will be some practice or practices upon which that agent relied in coming to hold and continuing to hold that belief. As such, any epistemic practice is apt for appraisal in terms of how reliably it leads to true over false beliefs. As I shall discuss in more detail below, to say that an agent follows a certain epistemic norm is to say that that agent somehow regulates the relevant epistemic actions in accordance with that norm. For our purposes, epistemic actions are actions that, in some way fundamental to their nature, contribute to achieving epistemic aims – the most fundamental of which are acquiring true beliefs and avoiding false beliefs. What this means, however, is that an agent’s following a certain epistemic norm can be considered amongst that agent’s epistemic practices. This is most obvious for norms of belief. In so far as some norm encodes conditions under which an agent ought, ought not, or is permitted to form beliefs in a given domain, following that norm will be a good epistemic practice just in so far as following the prescriptions of that norm will reliably yield true and not false beliefs. Though less directly, we can say similar things about other kinds of epistemic norms. Norms of trust will be epistemically good in so far as agents who follow those norms reliably trust reliable sources of testimony. Norms of inquiry will be epistemically good in so far as agents that follow those norms reliably follow strategies of inquiry that reliably yield truths over falsehoods. And so on.

Following these points let’s say that:

If N is an epistemic norm and S follows N, then, N is amongst S’s epistemic practices.

If, though, following-N is amongst S’s epistemic practices, the general considerations about reliance and epistemic self-trust in the previous chapter will apply to S’s following-N. So, we can add:

- (1) If S forms justified beliefs via N, then, outside of contexts of doubt, S is entitled to rely on N

And,

- (2) If S is in a context of doubt vis-a-vis N, S will be rationally permitted to rely on N, *iff* N has some form of epistemic self-trust in respect to her following-N.

As we saw in the previous chapter, epistemic self-trust comes in two forms: *predictive self-trust*, which is the combination of reliance and belief that the relevant practices are reliable; and *affective self-trust*,

which we characterised as the combination of reliance and optimism about one's reliance on the relevant practices. Crucially, one ought not to trust one's practices in either way come-what-may. Thus, we can say, in so far as the attitude of predictive self-trust is belief:

- (3) If S is in a context of doubt vis-à-vis N, S will be rationally permitted to rely on N via **predictive self-trust**, *iff* S justifiably believes N is reliable.

Importantly, we saw that in contexts of doubt, one cannot justifiably come to predictive self-trust, or justifiably continue to hold predictive self-trust, by way of epistemically circular justification. Thus, we can also add, as for other epistemic practices:

- (4) If S is in a context of doubt vis-à-vis N, and S has no non-circular justification for believing N, S will be rationally permitted to continue relying on N, *iff* N has well-placed **affective self-trust** in respect to her following-N.

I will return to the topic of what it might mean to say that one has well-placed affective self-trust in the face of doubts about one's epistemic norms in 6.4. For now, however, I want to consider the question of how it is that most of us come to subscribe to and follow the norms that we do.

### 6.3.2 Epistemic norms as social practices

What I want to suggest in this section is that ordinarily – though, not necessarily – we can understand an agent's practice of following some epistemic norm in terms of their internalization of socially sanctioned ways of regulating their epistemic activities. In other words, I want to suggest that norm-following should be understood not only as an epistemic practice, but as a *social* epistemic practice.

The understanding of what it is to follow an epistemic norm that I shall forward in this section owes much to responses to the so-called 'rule-following problems' found in the various writings of Wittgenstein that posit a social or communitarian dimension to rules or rule-following. This includes positions forwarded by Robert Brandom (1998), Susan Hurley (1989), Saul Kripke (1982) (1982), John McDowell (1984), and Crispin Wright (1980).<sup>72</sup> Rule-following problems, we might note, have the potential to afflict many domains. However, the bulk of contemporary discussion has followed Wittgenstein's concern with the metaphysics of linguistic meaning. Before proceeding, then, I want to be entirely clear that I am not aiming in this section to forward any

---

<sup>72</sup> For Wittgenstein's own remarks, see especially (Wittgenstein 1997 §138-242). For an excellent summary of this debate see McGinn (2013) For alternatives to communitarian responses to the rule-following remarks see Blackburn (1984) and Boghossian (2008b: 9-51)

metaphysical claims about epistemic norms or what it means to follow an epistemic norm. For that reason, I shall keep my engagement with the rule-following literature to a minimum. Rather less ambitiously, what I shall suggest is that the best explanation of how the *ordinary* agent follows a particular norm will make reference to that agent's social context and upbringing. As we shall see later, this is directly relevant to discussion of deep disagreement.

The first point to emphasise is one that is widely noted in the rule-following literature. Simply put, to follow a rule is more than to conform to a rule. My answers to the questions on an algebra test might, for instance, correspond precisely to the answers I would have got if I had successfully followed the rules of algebra when doing the test – even though I guessed all of those answers. If my answers are the product of guesswork, though, my answers are not the product of my following the relevant rules. Since it is possible for an agent's actions to conform to a rule without following that rule, rule-following cannot consist of mere conformance. The natural suggestion, then, is that following a rule requires that one in some way regulates one's actions in accordance with what one takes to be the prescriptions of that rule. Since we can understand epistemic norms in terms of rule-like propositions or imperatives, the same would apply to epistemic norms.

Taking up this suggestion, then, the question is what kind of regulation is necessary for rule-following, or, more specifically, for following epistemic norms.

One answer we might consider is that following rules consists of one's acting on an explicit intention to follow that rule. I.e. If I follow the rule 'If C, then  $\Phi$ !', when I do  $\Phi$ , I do so because I intend to follow the rule 'If C, then,  $\Phi$ !' and because I have judged the conditions specified in C to obtain. Following Boghossian (2008a), we can call this model the 'Intention View'. As Boghossian points out, rule-following on this view – and so the following of epistemic norms – takes the form of inference. As he explains:

To act on this intention, it would seem, I am going to have to think, even if very fleetingly and not very consciously that its antecedent is satisfied. The rule itself, after all, has a conditional content [...] And it is very hard to see how such a conditional intention could guide my action without my coming to have the beliefs that its antecedent is satisfied (Boghossian 2008a: 127).

There are a number of problems with the Intention View so understood.

First, as Boghossian points out, inference – including a practical inference of the kind Boghossian ascribes to the Intention View – is a paradigmatically rule-following practice. Thus, if all instances of rule-following involve an inference as to how to follow the rules in question, all instances of inference must involve an inference as to how to follow the respective rules of inference. But,

Boghossian argues, that way regress beckons. For instance, on the Intention View, even prior to discovering her disagreement with Prudence, Fax's following her norms in coming to believe 'the public don't trust either of us' would necessarily involve an inference from her intention to follow those norms, her beliefs about the content of those norms, and her beliefs about the circumstances she finds herself in. Since inference involves rule-following, however, *that* inference would necessarily involve Fax's making a further inference from her intentions and beliefs in respect to the rules of inference and the circumstances she finds herself in. But, again, that inference would necessarily involve a further inference. And so on. As Boghossian puts it, 'the Intention View would look to be hopeless: under its terms following any rule requires embarking upon a vicious infinite regress in which we succeed in following no rule.' (Boghossian 2008a: 128)

Any defence of the Intention View must grapple with Boghossian's regress argument. Putting aside the question of whether there is a good response to that argument, however, there are also issues with how well the Intention View applied specifically to epistemic norms is a realistic account of how most of us go about our epistemic business. The issues I have in mind should be familiar, and echo points made in the discussion of how best to characterize deep disagreement in Chapter 1 and the discussion of epistemic practices in the Chapter 5.

In a nutshell, most of us do not have conscious beliefs about our norms, nor would it seem that most of our beliefs (or other epistemic actions) are the product of reasoning from such beliefs. But this would seem to be just what the Intention View requires for it to be so that a belief is the product of an agent's following epistemic norms. On the Intention View, then, very few of the beliefs that people hold will be the product of their following epistemic norms. On a similar line of thought, as I noted in Chapter 1, identifying precisely what one's epistemic norms are is a difficult (if not impossible) theoretical task. The Intention View of rule-following, however, would appear to require just that kind of specificity. I.e. I could hardly make a practical inference as to how to act in any given circumstance from the belief that 'I follow and ought to follow some norm or other to do with evidence'. To ground the kinds of inference in question, my beliefs about that norm would have to offer some specification of the conditions under which I ought to, ought not to, or am permitted to take it that I have sufficient evidence to justify belief. Few, if any of us, however, can be expected to have such well-specified beliefs about our epistemic norms. Again, then, on the Intention View, few of us would have beliefs that are the product of our following epistemic norms. Presuming that many of our beliefs and other epistemic actions are formed or performed in ways that can be characterised in terms of our following epistemic norms, then, the Intention View would appear to be thoroughly off-track as an explanation of our ordinary practices

of following epistemic norms. Thus, irrespective of whether there is a good response to Boghossian's regress argument, some other explanation of how we follow norms is required.

What I would like to suggest is that, for most of us at least, the regulation of beliefs in accordance with epistemic norms can be explained by our having been inculcated into normatively loaded ways of apprehending features of the world via our exposure to the epistemic practices of the communities, and members of the communities, of which we are a part. We might call this an 'inculturation' model of norm-following. Precedent for this can be found in a wide range of literature. This includes the aforementioned social practice based responses to Wittgenstein's remarks on rule-following, John McDowell's (1978, 1979) related work on the moral virtues, Miranda Fricker's (2007) work on testimony and epistemic injustice, and Sripada and Stich's (2012) work on the psychology of social and moral norms. Given the breadth of this literature, I shall not endeavour to lay out a systematic account of this understanding of norms and norm acquisition. Rather, I draw on some of the aforementioned work to explore the features of norms and norm-following that I take to have significance when it comes to questions of self-trust in contexts of deep disagreement.

As I conceive it an 'inculturation' in norms typically consists of an agent's exposure to how others conduct their epistemic practices. As the agent observes and interacts with others in the community to which they belong, they witness how those others attribute authority, who they trust and don't trust, under which conditions people (appear to) form beliefs, make assertions, offer testimony, and so on. Immersed in this stream of experience, and later becoming an active participant in it, the agent comes to see these practices as the way things are done and, so, develop a sensibility about and sensitivity to the epistemically salient features of different situations. Fricker offers a useful description of what I have in mind here in her account of how we develop a (virtuous) testimonial sensibility:

[W]e should think of the virtuous hearer's sensibility as formed by way of participation in, and observation of, practices of testimonial exchange. There is, in the first instance, a passive social inheritance, and then a sometimes-passive-sometimes-active individual input from the hearer's own experience. Together the individual and collective streams of input are what explain how our normal unreflective reception of what people tell us is conditioned by a great range of collateral experience—our informal background 'theory' of socially situated trustworthiness (Fricker 2007: 83).

A crucial element of this developmental process is that the agent witnesses, not only the patterns by which individuals conduct their own epistemic affairs, but also the responses to those activities by others – and especially the affective dimensions of those responses. Just as significant, for instance, as seeing who and when individuals trust, is seeing the association of reactive attitudes



such as resentment and betrayal with failures to make good on that trust, and gratitude when it is acted on.<sup>73</sup> Similarly for other epistemic actions. If an individual consistently fails to form beliefs or make assertions that reflect the available evidence (or whatever else is considered epistemically salient in the community in question) they are likely, not only to be corrected, but to be subject to criticism.<sup>74</sup> The further their actions stray from the norms, or the greater the practical consequences of their doing so, the harsher and more censorious this criticism is likely to become. These experiences, I suggest, provide the catalyst for the development of the agent's sense of the way things are into a genuinely normative and genuinely epistemic sensibility. Thus, as we come to inhabit these communal practices of praise and blame, so too we come to see compliance with the associated patterns of conduct, not just as something that people do, but something that people, including the agent, *ought* to do. Sripada and Stich describe this in the moral context in terms of the internalization of norms – where internalization is a term of art from anthropology and sociology to describe ‘a characteristic style of motivation in which the individual values compliance with moral rules even when there is no possibility of sanction from an external source’ (Sripada and Stich 2012: 292). As Sripada and Stich point out, the internalization hypothesis explains why, once indoctrinated into a rule, ‘people exhibit a lifelong pattern of highly reliable compliance with the rule’ (Sripada and Stich 2012: 292).

Importantly, the internalization of epistemic norms need not require, nor rule out, a role for more deliberate forms of instruction. Instruction, both generally and in specific domains, is undoubtedly an important part of both how we learn to reason, inquire and so on, and of how epistemic actions come to be coordinated across a community. Paralleling a suggestion from McDowell's account of the moral virtues, however, we might note that much, if not all, of our instruction in the whys and wherefores of how to think comes, not in the form of explicit explanation of the relevant norms, but a more subtle combination of exhortation, suggestion, and ostension. So, McDowell writes:

[O]ne exploits contrivances similar to those one exploits in other areas where the task is to back up the injunction "See it like this": helpful juxtapositions of cases, descriptions with carefully chosen terms and carefully placed emphasis, and the like (McDowell 1978).

Related to the point about instruction, that we are inculcated into normative epistemic practices does not preclude the role of explicit beliefs about how others conduct their epistemic practices

---

<sup>73</sup> See Baier (1986), Jones (1996), Holton (1994), and Faulkner (2011) on the reactive attitudes and interpersonal trust.

<sup>74</sup> Sripada and Stich suggest that the evidence from anthropology and experimental economics indicates, not only a reliable connection between norm violations and punitive reactions, but that motivations to censure norm violators are often ‘truly intrinsic’ (Sripada and Stich 2012: 294)

and how they are received in a wider social context – nor indeed explicit consideration of what conditions under which one ought to form beliefs, make assertions, ask questions, and so on. Indeed, space for reflection, and especially critical reflection, upon the practices we inherit from the communities we are embedded within would appear to be essential to the possibility that those communities might develop and evolve. Likewise, it is essential to the idea that our practices correspond to epistemic *norms* and not mere patterns of indoctrination. As Fricker makes the point in reference to testimonial sensibility, ‘The claim that testimonial sensibility is a capacity of reason crucially depends on its capacity to adapt in this way, for otherwise it would be little more than a dead-weight social conditioning that looked more like a threat to the justification of a hearer’s responses than a source of that justification’ (Fricker 2007: 84).

Finally, the model I have suggested of how we come to internalize epistemic norms extends also to how we implement those norms when acquired. Once again echoing aspects of McDowell’s account of the moral virtues and Fricker’s account of testimony, we might describe this in terms of the agent having acquired a normatively loaded way of *apprehending*<sup>75</sup> situations. So, McDowell writes:

In moral upbringing what one learns is not to behave in conformity with rules of conduct, but to see situations in a special light, as constituting reasons for acting; this perceptual capacity, once acquired can be exercised in complex novel circumstances, not necessarily capable of being foreseen and legislated for by a codifier of the conduct required by virtue (McDowell 1998: 85).

And Fricker,

The virtuous hearer does not arrive at her credibility judgement by applying pre-set principles of any kind, for there are none precise or comprehensive enough to do the job. She ‘just sees’ her interlocutor in a certain light, and responds to his word accordingly (Fricker 2007: 73).

On this understanding applied to the topic of epistemic norms, just as we might come to subscribe to certain norms without explicitly reasoning upon their content, beliefs formed via norm-following do not necessarily depend upon inference from intentions and beliefs about one’s norms and the situations one finds oneself in. The perceptual analogy helps us to see how this might be.<sup>76</sup> For, typically at least, the beliefs we form via perception do not involve reasoning from beliefs about the reliability of one’s perception faculties and one’s perceptual experiences. Likewise, on

---

<sup>75</sup> I prefer apprehending to feeling, seeing, or sensing since norm-following is clearly not perceiving in a literal sense.

<sup>76</sup> Arguably, my reliance on my perceptual faculties to form beliefs is also structured by my internalization of relevant epistemic norms. I don’t take a stance on this matter here – the comparison between perceptual beliefs and norm-acquired beliefs in the text should be seen as illustrative, rather than contrastive.

this analogy, the beliefs we form via following our epistemic norms, typically at least, come about through, to use Fricker's terms, 'spontaneous and unreflective' responses to consideration of the circumstances in which one finds oneself. So, for instance, considering the question 'what is in front of me?', I turn my attention to the object in front of me. I have the visual experience of a grey table, in so doing, I come to believe 'there is a grey table in front of me'. Considering the question 'how many books are on that table?', I see that there are two books on the right side, two on the left, and none in the middle, in so doing, I come to believe 'there are four books on the table in front of me'. In the second instance, my belief is formed, in part, because I have internalized epistemic norms (and rules of mathematics) – but I need not consciously consider those norms and rules when I form the belief. Just as I don't consciously reflect upon the reliability of my perceptions when I form the belief in the first instance. Instead, we might say that in considering the question 'how many books are on the table' I have primed myself to pay attention to those features of the situation that may correspond to the application conditions of the relevant norms (and mathematical rules). When I see those conditions to apply, I respond as those norms prescribe. This way of apprehending things is so ingrained into my way of thinking, relying on myself in respect to it is automatic.

There are four benefits to this account.

First, it explains how the majority of people come to follow particular norms. Second, it explains how individual followers of norms regulate their actions in accordance with what they take to be the prescriptions of that norm. Third, it does so whilst explaining the normativity of that practice. Fourth and most significantly, it achieves this without requiring that agent's realise what norms they follow (or even that they do follow norms), or explicitly reason from that realisation when determining how they ought to act, epistemically speaking, in different circumstances.

At this point it is important to reiterate that what this account does not offer are necessary and sufficient conditions of what it is for something to be an epistemic norm, or of what is required to follow an epistemic norm. This is important, both generally and in respect to the discussion of deep disagreement. If it *were* necessary and sufficient that norms correspond to social practices, there would be no facts outside of those social practices by which different norms could be said to be defective as ways of acquiring justification. For a disagreement to be deep, though, it must be so that those party to the disagreement have followed their respective norms competently. (I.e. If the justificatory structure of one party's beliefs is explained by some other flaw in their reasoning, the disagreement cannot be said to be explained by the difference in norms.) But, if there are no facts outside of the relevant social practices by which norms could be said to be defective, and

those party to a deep disagreement have followed those norms competently, then, there will be no defects in the justificatory structure of any of the parties' beliefs. Thus, on an account of norms such that they just are social practices, deep disagreements would be a kind of epistemically faultless disagreement. This I take it runs against intuitions about both the paradigmatic cases of disagreement such as POLITICIANS and, perhaps more so, domain-specific cases like DOCTORS-2. For, if as was the suggestion, the steadfast response to disagreement is not always the appropriate response – and in some cases some parties may have to abstain from their respective norms – then, it cannot be so that these are faultless disagreements. Correspondingly, were the account given to provide necessary and sufficient conditions for norm-following, there would be no normative problem of deep disagreement;

In response to this line of thought, I return to the point with which I began this discussion. I.e. that the explanation given here is only intended to explain how the *ordinary* human being, despite their epistemic limitations, can acquire a sense of how they ought to conduct their epistemic affairs.<sup>77</sup> As discussed earlier, however, I take it that there are moments in which members of a community – even one that is isolated from others – are able, through their own reflections and reasoning, to apprehend supposed shortcomings in the practices of their wider community and having done so, to come to new and purportedly better ways of doing things. And, for that to be the case, we have to deny that norms are always and entirely socially dependent. The consequence of this is that this account does not, and indeed cannot, ground any claims about the metaphysical nature of norms/norm-following. Once we lower our metaphysical ambitions in this way, however, we are able to retain the intuition that there are significant normative issues associated with the possibility of deep disagreement. For whilst most of us will depend upon our social experiences when it comes to learning how we ought to conduct our epistemic affairs, that is not to say that norms just are social practices, or that individuals cannot in exceptional circumstances come upon better ways of doing things. Similarly, the possibility of taking a critical stance on one's epistemic norms, despite the unconscious way in which those norms are acquired and implemented, allows us to retain the intuition that at times those party to deep disagreement may be held to account for their failure to respond to the possibility that their own norms are in some

---

<sup>77</sup> Compare to Kripke's (2002) discussion of rule-following, in which the metaphysical claims are advanced via consideration of how an *ideal* epistemic agent might follow a rule.

way defective.<sup>78</sup> With that said, let's finally return to consider the question of how we ought to respond to deep disagreement.

## 6.4 Responding to deep disagreement: The Higher-Order Trust approach extended

Now that we have a general account of what it means to follow an epistemic norm, we can return to 'The Normative Problem of Deep Disagreement'. As noted previously, a distinctive feature of deep disagreement is that disputants may be effective peers just because they follow different epistemic norms, whilst at the same time not having the kind of evidence that would enable them to recognise this. The suggestion at the beginning of this thesis, and around which much of this discussion has been framed, was that the answer to the question 'How ought one to respond to the realisation of disagreement?' cannot require the agent to place significance upon facts to which the agent does not have access. If, then, one is not evidentially in a position to recognise that one's interlocutors are one's effective peers, or that one's interlocutor's are one's effective peers *because* they do not share one's epistemic norms, one cannot and ought not to place significance upon those facts in responding to the realisation of disagreement. In other words, in such cases, one ought to respond to the realisation of disagreement *as if it were* an ordinary disagreement. This being so, the Higher-Order Trust approach to ordinary disagreement can be straightforwardly extended to at least some cases of deep disagreement.

Let's say then, that:

If  $S_1$  disagrees with  $S_2$  about  $Q$  and that disagreement is explained by  $S_1$  following norm  $N_1$  and  $S_2$  following norm  $N_2$ , so long as  $S_1$  does not have evidence that the disagreement is deep,  $S$  ought to continue following  $N_1$ . Thus, on the presumption  $N_1$  is good:

---

<sup>78</sup> Compare this point to Fricker's discussion of how we might ground a feeling of 'disappointment in resentment' towards the character Herbert Greenleaf, in Anthony Mingella's 'The Talented Mr Ripley, in his failure to rise above the biases and prejudices of the historical moment in reception of testimony from his daughter-in-law Marge: 'Any such disappointment is grounded in the conviction that it was historically possible, in our newly extended sense, for him to have made a better credibility judgement: even Greenleaf could have judged exceptionally. [...] By judging merely routinely in these heightened circumstances, he does something disappointing that it seems appropriate to hold him responsible for in some way; but the sense in which his moral and epistemic character is besmirched by his failure to rise above the routine is not enough to warrant anything as strong as blame. One cannot be blamed for making a routine moral judgement. But one can none the less be held responsible for making a merely routine judgement in a context in which a more exceptional alternative is, as a matter of historical possibility, just around the corner' (Fricker 2007: 105).

- (1) If  $S_2$  would not be  $S_1$ 's effective peer were the disagreement ordinary,  $S_1$  ought to favour whichever party  $S_1$  ought to believe has the relative advantage according to  $N_1$ .
- (2) If  $S_2$  would be  $S_1$ 's effective peer were the disagreement ordinary,  $S_1$  will be justified in believing  $Q$  *iff*  $S_1$  has well-placed affective self-trust. If not,  $S_1$  ought to revise her belief  $Q$  accordingly.

With this much said, we can cash out the suggestion from 6.2 that in some cases of deep disagreement, the 'mixed response' to deep disagreement will sometimes be appropriate. For, as per (2), if  $S_1$  does not have well-placed affective self-trust,  $S_1$  will lose whatever justification she had for believing  $Q$ . Yet, at the same time, since she has no evidence that the disagreement is deep, she will retain any entitlement to follow  $N_1$ . To put it in the terms I suggested earlier, then, in this kind of case, the realisation of disagreement places the disputants in a position that requires trust in their epistemic practices in general but does not require trust in their norms. In such a position, the mixed-response to deep disagreement may be the appropriate response.

I suggested earlier that the majority of deep disagreements will be cases in which the disputants do not realise that their disagreement is deep and do nothing epistemically wrong in failing to realise that. Thus, in the majority of deep disagreements the appropriate response to the disagreement at the level of belief will be the same as if that disagreement were an ordinary disagreement. At the same time, the appropriate response at the level of norms will typically be to stick by one's norms. In other words, for the majority of cases, either the steadfast or mixed-response to deep disagreement will be appropriate.

Whilst we might suppose that the majority of deep disagreements will be ones in which those party to the disagreement are not aware of its nature, however, this will not always be the case.

To illustrate, suppose that in the world described in DOCTORS-2 it is a matter of live controversy within the medical domain as to what role – if any – clinical judgement ought to play in the deliberations of medical practitioners. As a highly trained doctor, we might expect Quinn to be aware of this controversy. Suppose also that Quinn and Zaius continue to confer about the source of their disagreement and discover the details about their different training and the different ways in which each responded to the (shared) feeling that something was off with the evidence. Given this additional evidence – combined with the fact that prior to discovering her disagreement with Zaius, Quinn was justified in believing Zaius to be competent – in this situation it will no longer be reasonable for Quinn to infer that the disagreement is explained by some contextual or personal

failing on her or Zaius's part. Instead, it would seem more reasonable that Quinn suppose that the disagreement is explained by her and Zaius's differing stances on the wider controversy about the relevance of clinical judgement. Thus, we might say that in this situation, unlike DOCTORS-2 and POLITICIANS, Quinn would have reason to believe that her disagreement with Zaius is explained by the kind of systematic difference in practices that we might characterise in terms of epistemic norms. Call this case, DOCTORS-3. In DOCTORS-3, in effect, Quinn ought to recognise that her disagreement with Zaius is a deep disagreement.

In light of cases like this, then, let us say that disputants in this kind of situation are *recognized peers* in respect to deep disagreement. More specifically, disputants will be *recognized peers* when the following conditions are met:

- I. They are engaged in deep disagreement
- II. They are effective peers in respect to that disagreement.
- III. They recognise that it is a deep disagreement

As per the points about DOCTORS-3, we might expect these conditions to be met more often in cases where the relevant norms are domain-specific, or relatively close in content. However, this is not to say that it cannot be the case that those engaged in more general and wider deep disagreements such as POLITICIANS are recognized peers. Indeed, much of the motivation for discussing deep disagreements would appear to stem from the fact that theoreticians at least do consider disagreement between scientific and religious communities, conventional and alternative medicine, left- and right-wing political groups, and so on, to be well explained by differences in epistemic principles, commitments, and norms. If the theoretician is in a position to realise this, however, there is no reason in principle that those engaged in such disagreements might not also realise this. Indeed, since the theoreticians in question are rarely neutral on such issues, we might take it as a given that at least some of those party to such disagreements have already realised this! As a consequence of this possibility, and despite the points about the majority of deep disagreement, then, we still need an explanation of how recognized peers ought to respond to the realisation of disagreements between recognized peers.

The first thing to consider is the significance that the status of recognized peerhood has upon the kind of defeater one might acquire in case of deep disagreement.

In this light, note first that, as per the discussion of higher-order evidence in Chapter 4, any defeating higher-order evidence disputants might acquire where they are not in a position to recognise that their disagreement is deep (e.g. in DOCTORS-2) will typically be in the form of a

contextual undercutting defeater – i.e. the same kind of defeater typically acquired in case of ordinary disagreement with an effective peer. This is because, without evidence that the disagreement is deep, it will typically be the case that the best available explanation of the disagreement will be some localised performance error, or other feature specific to the context. In contrast, when disputants are recognized peers, the higher-order evidence they acquire on realising the disagreement will be in the form of a general undercutting defeater. This is because evidence that gives one reason to doubt the norms one followed in coming to believe Q, is evidence that one may have come to believe Q because of a *general and non-context specific* flaw in one's epistemic practices. Since the evidence that recognized peers have that their disagreement is deep effectively rules out other explanations of the disagreement, then, it will constitute a general undercutting defeater for whatever justification they had in respect to the disputed beliefs. Moreover, given the nature of general undercutting defeaters, that higher-order evidence will be a defeater, not only for the justification they have for the beliefs under dispute in the current context, it will be a defeater for the justification they have for any other beliefs formed via the contested norms. In other words, in lieu of a defeater-defeater, not only does the realisation of deep disagreement with a recognized peer mandate the conciliatory response to deep disagreement – but, it will also have the sceptical consequences we saw to be associated with the principle of Total Revision. As we saw, these consequences are likely to be (at best) extensive if the norms in question are domain specific, near-global if the norms in question are fully general.

If recognized peers are not to face the sceptical consequences of Total Revision, then, they must have some kind of defeater-defeater for the higher-order evidence they acquire on realising the disagreement. And here it is crucial that one's epistemic norms count amongst one's epistemic practices. For that being so, reflecting the discussion in 6.3, we are able to extend the Higher-Order Trust approach to cases of deep disagreement between recognized peers, just as we were able to extend it to other cases of deep disagreement between. Thus, we can say:

If  $S_1$  disagrees with  $S_2$  about Q, that disagreement is explained by  $S_1$  following norm  $N_1$  and  $S_2$  following norm  $N_2$ , and  $S_2$  is  $S_1$ 's recognized peer,  $S_1$  will be justified in believing Q *iff*  $S_1$  has well-founded affective self-trust. If not,  $S_1$  ought to abstain from  $N_1$  and so, in lieu of some other source of justification, ought to revise her belief Q accordingly, as well any other beliefs she has formed via following  $N_1$ .

Why only if  $S_1$  has affective self-trust? Because, recognized peerhood entails effective peerhood. Thus,  $S_1$  will have no independent reason to believe that she has the relative epistemic advantage over  $S_2$  (and be in a position to recognise this) As such,  $S_1$  will have no first-order defeater-defeaters



for the higher-order evidence – including any non-circular justification for believing that  $N_1$  is true. Thus, as per the discussion in Chapter 5,  $S_1$  will only have a defeater-defeater if she has well-founded affective self-trust. More specifically, since the doubts induced by the higher-order evidence concern  $N_1$ ,  $S_1$ 's affective self-trust must pertain to her continued reliance on  $N_1$ .

Whether or not, then, the discovery of deep disagreement has the sceptical consequences associated with Total Revision will depend on whether the agent in question has well-placed affective self-trust. In the previous chapter, we saw that there are two components to this. First, if an agent has affective self-trust toward some practice of theirs, that self-trust will be well-placed only if that practice is a source of justification.<sup>79</sup> Correspondingly, where affective self-trust is toward their following some epistemic norm, it will be well-placed only if following that norm leads to justified beliefs. Second, and more fundamentally, affective self-trust must be *available* to the agent – if, not, then whatever attitude they have toward the relevant practices, it will not be one of affective self-*trust*. It is in respect to the question of availability that the discussion of the nature of epistemic norms in 6.3 becomes relevant. We saw that there is not a principled account we can give as to when affective self-trust is available. However, we were able to pick out its key determiners. Thus, I suggested that the availability of affective self-trust in cases of ordinary peer disagreement will be dependent upon the circumstances of the disagreement and the character of the individual epistemic agent. What I would suggest here is that that the social character of one's norm-following practices introduces a third element to the mix. Correspondingly, a further determiner of the availability of affective self-trust in case of deep disagreement between recognized peers will (at least for the ordinary agent) be the wider social contexts within which the disputants were enculturated into their respective norms – as well as each agent's relationship to those social contexts. What I want to suggest in the concluding paragraphs of this chapter, is that the social nature of our norm-following practices can make all the difference to how we ought to respond to deep disagreement with a recognized peer. As we with the availability of affective self-trust in cases of ordinary peer disagreement, I shall endeavour to illustrate this notion by reference to examples. POLITICIANS and DINERS-3 will do nicely.

Take POLITICIANS. In this case, as we saw earlier, the difference between Fax's and Prudence's norms is both fundamental and wide. For that reason, I argued, we would not expect Fax and Prudence to be effective peers even in the ordinary sense. Given this divide, though, I would suggest that it would be exceptionally difficult for Fax to see *even the idea* of Prudence's way of

---

<sup>79</sup> To refer back to the terms in the previous chapter, we might understand this as a part of what it is for self-trust to be well calibrated. Self-trust in practices that are not a source of justification is poorly calibrated self-trust.

reasoning as a genuine alternative to her own way of reasoning. To do so, Fax would have to overlook the evidence she would otherwise have to view Prudence as incompetent *and* somehow inhabit a radically different way of apprehending the world than her own. Such an exceptional imaginative leap, it strikes me, could not be achieved merely by reflecting on any evidence she might have that the disagreement is deep. Rather, much as I have suggested we typically come to internalize our own norms through exposure to the practices of others, this kind of leap would typically require an agent to have a similar stock of experiences with followers of the rival norm on which to draw. Given the divide between the pairs' norms, this seems extremely unlikely. For the kinds of experience required are one's that involve the recognition of epistemic authority, relationships of dependence and trust, cooperation and collaboration. In other words, just the kind of relationships that depend upon the presumption of competence in others that we would not expect to obtain between followers of Fax's and Prudence's norms.

In a case like *POLITICIANS*, then, my suggestion is that the social dimensions of deep disagreement can have a kind of inverted significance in respect to the availability of affective self-trust. For, where the difference in norms is wide and deep enough, the disputants are unlikely to have the stock of social engagements and interactions with followers of the other norms that would spur the exceptional leap of imagination needed to see the rival norms as genuinely epistemic alternatives to their own. And, if one cannot see the rival norms as a genuine alternative to one's own, it seems that recognition that the disagreement is deep will fail to exert any influence upon one's inclination to trust one's own norms. Though Fax may be in a context of realisation, we might say, she will not be pushed into a context of doubt. Thus, if the agent would be inclined to trust their norms in other contexts, so we would expect that trust also to be available in a deep disagreement of this kind. And that being so, steadfastness as a response to the realisation of disagreement will be a rational and psychological possibility.

As noted toward the beginning of the chapter, the kinds of deep disagreement which threaten global scepticism tend to be of a kind with *POLITICIANS*. Thus, whilst we might not be able to lay down any hard and fast principles as to when trust is appropriate and inappropriate in the face of deep disagreement, the thought that affective self-trust will tend to be available (or as available as it would be in other circumstances) in these cases allows us to stave off the worries about global scepticism that the possibility of these cases might lead to.

The flip side to this suggestion, of course, is that the possibility that recognized peers might be justified in staying steadfast in a narrower, domain-specific case of deep disagreement such as in *DOCTORS-3* will be comparatively lower. After all, as we noted in discussion of that case earlier,

the content of the rival norms is such that we would not expect followers of each to doubt each other's competence as a matter of course. And, that being so, there is no reason to think that they would not have the rich stock of social experiences that might *either* spur the imaginative leap to understanding the other's epistemic sensibilities *or* even justify believing that the other's norms do indeed represent a genuine epistemic alternative. Once those norms are seen to represent a genuine alternative, however, the context of realisation will become a context of doubt. One in which the possibility that either disputant might bracket out those doubts – with the kind of rational sensitivity that is characteristic of a trusting as opposed to a dogmatic response – will be remote to the point of being impossible. And, with affective self-trust unavailable steadfastness will no longer be a possibility – neither rationally, nor psychologically. As such, the conciliatory response to deep disagreement will be the appropriate one. As noted earlier, such a response will still have extensive sceptical consequences – though not the near global one's associated with a case like POLITICIANS. Again, though, this seems to match up well with the intuitions that led us to consider such cases to start with and to consider whether the Higher-Order Trust approach – as a non-uniform approach – might extend to disagreement.

## Conclusion

In this chapter, we have seen how the Higher-Order Trust approach to 'The Problem of Ordinary Disagreement' can be extended to cover also 'The Normative Problem of Deep Disagreement'. In the brief conclusion to the thesis that follows, I shall bring together the main findings of the past few chapters to provide a fuller statement of the Higher-Order Trust approach to disagreement.

# Conclusion

In this thesis I have considered two related questions concerning the normative significance of disagreement:

How should I respond to ordinary disagreement with an effective peer? And, how should I respond to deep disagreement with an effective peer?

I labelled these 'The Problem of Ordinary Disagreement' and 'The Normative Problem of Deep Disagreement'. In response to which I forwarded a normative theory of disagreement that I call the Higher-Order Trust approach to disagreement. In a nutshell, this says that – whether disagreement is ordinary or deep – the matter of how one should respond to the realisation of disagreement comes down to whether one has available a form of self-trust that I refer to as *affective self-trust*.

Given the nature of the determinants of the availability of self-trust in general as well as in the different kinds of disagreement, I suggested that it is not possible to give a principled account of the conditions under which these responses are available, nor, correspondingly under which these responses are justified. Nonetheless, it is possible to lay out some of the core claims of the Higher-Order Trust approach in more schematic fashion. I shall do so in this brief conclusion.

Effective peerhood, the reader will recall, was defined such that:

If,  $S_1$  and  $S_2$  disagree and  $S_1$  has no undefeated reason *independent of the substance of the disagreement* to believe specifically that  $S_1$  or  $S_2$  holds the relative epistemic advantage vis-à-vis the disputed propositions, then,  $S_2$  is  $S_1$ 's *effective peer*.

In case of disagreement with an effective peer, one acquires higher-order evidence that calls into question the practices by which one came to hold the disputed beliefs. In calling those practices into question, such higher-order evidence functions as a defeater for whatever justification one might have had for those beliefs prior to realising the disagreement. What form this defeater takes will depend on the kind of disagreement one is engaged in.

In cases of ordinary peer disagreement where the disputants have based their beliefs on different evidence and are unable to share that evidence, the higher-order evidence is a rebutting defeater. In such cases, where  $S$  believes  $Q$ ,  $S$  acquires higher-order *evidence of evidence* that justifies her

believing  $\sim Q$  and in light of which, in lieu of a defeater-defeater, S cannot rely upon her appreciation of *the evidence she has* to provide a reliable guide to the truth about Q.

In cases of ordinary peer disagreement where disputants share the same evidence, the higher-order evidence is a non-contextual undercutting defeater. In such cases S's realisation that her effective peer has come to believe  $\sim Q$  on the basis of the same evidence as she came to believe Q puts her in a position where, in lieu of a defeater-defeater, she cannot rely upon the practices by which she came to that conclusion. Since, typically, the best explanation of such cases will be that one party or other was subject to some kind of performance error specific to that context these doubts do not call into question S's justification for her un-disputed beliefs.

In cases of deep disagreement with an effective peer, things are slightly more complicated. In the majority of cases disputants will not be in a position to realise that their disagreement is deep. When this is so, the fact that a disagreement is deep will not play a role in determining the appropriate response to the disagreement. Instead, in such a situation S ought to respond to that disagreement as if it is an ordinary disagreement. However, we saw that it is possible that those party to deep disagreement could realise both that the disagreement is deep *and* that they are effective peers because of that. In such cases, the higher-order evidence will come in the form of a general undercutting defeater. In this situation, S's realisation of the disagreement and its nature puts her in a position where, in lieu of a defeater-defeater, she cannot rely on the norms that she followed in coming to believe Q. In so far as norms correspond to general features of our epistemic practices, S acquires a defeater – not only for her justification for believing Q, but, other things being equal, any and all other beliefs she formed via the relevant norm.

Where the forms of defeater in the different cases differ, what is common to all three is that defeating higher-order evidence works by placing an agent in a position where, in lieu of a defeater-defeater, they are not permitted to rely upon the practices by which they came to hold the disputed beliefs. We can call these 'contexts of doubt'. Following the discussion in Chapter 5, we can schematically lay out the nature of defeat from higher-order evidence as follows:

- (1) If S comes to believe Q on the basis of P via K, S will be justified in believing Q, *iff* P is a reason for S to believe Q and S is rationally permitted to rely on K.
- (2) If S acquires higher-order evidence impugning K, in lieu of a first-order defeater-defeater for R, S will be placed in a context of doubt vis-à-vis K.

As we saw in Chapter 6, to say that S follows epistemic norm N is to say that following N is amongst S's epistemic practices. That being so, this schema of defeat applies also to realised deep disagreement.

First-order defeat-defeaters are reasons that are not themselves impugned by the putative defeater. In the case of disagreement this amounts to the requirement that they be independent of the substance of the disagreement. Thus, by definition those party to disagreement with an effective peer will not have available any kind of first-order defeater. It follows, then, that either the appropriate response to disagreement will always be the conciliatory response; or there must be some other form of *higher-order defeater-defeater* that one might have available in such cases and that can underwrite one's continued reliance on the practices called into question by the higher-order evidence.

The Higher-Order Trust approach to disagreement says that one can have such a higher-order defeater-defeater – in the form of and by way of affective self-trust. Thus, we can complete the schema by adding that:

- (3) If S is in a context of doubt vis-à-vis K, S will be rationally permitted to rely on K *iff* S has some form of epistemic self-trust.

Affective self-trust as I conceived it can be characterised in terms of an agents having an attitude of optimism about relying on their own epistemic practices. To be optimistic in this way is to manifest a willingness to rely on one's practices *under the presumptions that* (i) if K is reliable I will be more likely to achieve my epistemic aims if I rely on K, and that (ii) by relying upon K I will accrue those benefits, just because (iii) K is reliable.

Importantly, these presumptions are not made in the form of belief. Rather they are the background commitments that constitute what it is to have affective self-trust. And in that, they provide the underlying rational structure of a normatively loaded way of perceiving situations of reliance as 'cognitively safe'. These presumptions together imbue affective self-trust with its own internal structure of reason, in light of which, affective self-trust is in itself sufficient explanation of why an agent would continue to rely. Thus, if self-trust is available to an agent that agent will thereby have reason to rely upon the practices in question. Whether that reason is good is merely a question of whether that trust is well-placed.

What this means, though, is that were S to have available affective self-trust vis-a-vis K, that would also give S access to a defeater-defeater for the higher-order evidence acquired on realising disagreement with an affective peer. This is because, in underwriting the rational permissibility of

her relying on the deliverances of K, S's affective self-trust would allow S to rely upon her original reasons for holding those beliefs to reason as follows:

- (i) In coming to believe Q on the basis of P, I judged that P establishes Q
- (ii) My interlocutor disagrees with me about Q
- (iii) But P establishes Q (via the presumptions of justified affective self-trust)**
- (iv) Thus, I did not err in judging that P establishes Q
- (v) Therefore, Q and 'my interlocutor must have erred in some way in coming to believe  $\sim Q$ '

Where the disagreement is deep, S could argue additionally:

- (vi) My interlocutor came to believe Q by competently following a norm other than mine
- (vii) Therefore, Q must have erred because her norm is somehow defective

In other words, affective self-trust is a response to a context of doubt. When it is available it opens the possibility of being rationally steadfast. But without it the only option is conciliation.

Thus, if we recognise the possibility of affective self-trust, we can recognise the following two principles of disagreement:

If  $S_1$  is justified in believing Q but comes to realise that  $S_2$  disagrees with her about Q, and  $S_2$  is  $S_1$ 's effective peer, then:

- I. If  $S_1$  does not trust, in the affective sense, the practices by which she came to believe Q,  $S_1$  will have no defeater-defeater for the higher-order evidence acquired on realising the disagreement, thus,  $S_1$  ought to revise her belief Q
- II. If  $S_1$  does trust, in the affective sense, the practices by which she came to believe Q,  $S_1$  will have a *higher-order defeater-defeater* for the higher-order evidence acquired on realising the disagreement, thus,  $S_1$  ought to stay steadfast in her belief Q

The Higher-Order Trust approach to disagreement is, as such, a non-uniform account of disagreement. What will determine which response to any particular disagreement is appropriate is largely a matter of whether affective self-trust is available. Here, though, we hit bedrock, so to speak. As discussed in Chapter 5 and 6, the determinants of self-trust include the character of the agent in question, the circumstances in which the need for self-trust arises, and in case of realised deep disagreement the social context and background of the disagreement. All three of these determining features, however, are varied enough and the ways that they interact in particular cases complex enough that we can expect any attempt to produce a principled account of the availability

(and so appropriateness) of affective self-trust to run aground. Thus, we might say, as far as the problems of disagreement go, the best we can do is to ask in any particular case: Should I trust myself?



Fig. 1: Varieties of defeater-Defeater for Higher-Order Evidence

	R	DD		
		<i>Rebutting defeater-defeater</i>	<i>Contextual undercutting-defeater-defeater</i>	<i>General undercutting defeater-defeater</i>
Rebutting defeater	(R)RB(~Q)	(DD)RB(Q)	(DD)RD(~R unless ~Q); S <i>can</i> justifiably believe (were it not for DD, (R)RB(~Q))	(DD)RD(~R unless ~Q); S <i>cannot</i> justifiably believe (were it not for DD, (R)RB(~Q))
Contextual Undercutting Defeater	(R)RD(~Q unless P); S <i>can</i> justifiably believe (were it not for R, (P)RB(Q))	(DD)RB(~Q unless P)	(DD)RD ((R)RD(~Q unless P)); S <i>can</i> justifiably believe (were it not for DD, ((R)RD(~Q unless P))	(DD)RB~((P)RD(~Q unless P)); S <i>cannot</i> justifiably believe (were it not for DD, ((R)RD(~Q unless P))
General Undercutting defeater	(R)RD(~Q unless P); S <i>cannot</i> justifiably believe (were it not for R, (P)RB(Q))	(DD)RB(~Q unless P)	(DD)RB~ ((P)RD(~Q unless P)); S <i>can</i> justifiably believe (were it not for DD, ((R)RD(~Q unless P))	(DD)RB~((P)RD(~Q unless P)); S <i>cannot</i> justifiably believe (were it not for DD, ((R)RD(~Q unless P))

RB: 'reason to believe'
RD: 'reason to doubt or deny'

# Bibliography:

- Alston, William P. 1986. 'Epistemic Circularity', *Philosophy and Phenomenological Research*, 47: 1-30.
- . 1988. 'The Deontological Conception of Epistemic Justification', *Philosophical Perspectives*, 2: 257-99.
- . 2005. *Beyond 'justification' : dimensions of epistemic evaluation* (Cornell University Press: Ithaca, N.Y.).
- Baier, Annette. 1986. 'Trust and Antitrust', *Ethics*, 96: 231-60.
- Ballantyne, Nathan, and E. J. Coffman. 2012. 'Conciliationism and Uniqueness', *Australasian Journal of Philosophy*, 90: 657-70.
- Barker, Simon, Charlie Crerar, and Trystan Goetze (eds.). 2018. *Harms and Wrongs in Epistemic Practice* (Cambridge University Press: London).
- Battaly, Heather. 2012. 'Sosa's "Reflective Knowledge": How damaging is epistemic circularity?', *Synthese*, 188: 289-308.
- . 2018a. 'Can Closed-mindedness be an Intellectual Virtue?' in S. R. Barker, C. Crerar and T. Goetze (eds.), *Harms and Wrongs in Epistemic Practice* (Cambridge University Press: London).
- . 2018b. 'Closed-Mindedness and Dogmatism', *Episteme: A Journal of Individual and Social Epistemology*, 15: 261-82.
- Bergmann, Michael. 2004. 'Epistemic circularity: Malignant and benign', *Philosophy and Phenomenological Research*, 69: 709-27.
- Blackburn, Simon. 1984. 'The Individual Strikes Back', *Synthese*, 58: 281-301.
- Boghossian, Paul. 2008a. 'Epistemic Rules', *The Journal of Philosophy*, 105: 472-500.
- Boghossian, Paul 2008b. *Content and justification: philosophical papers* (Clarendon Press: Oxford).
- Brandom, Robert. 1998. *Making it Explicit: Reasoning, Representing & Discursive Commitment* (Harvard University Press: Cambridge, MA).
- Burge, Tyler. 1993. 'Content Preservation', *The Philosophical Review*, 102: 457-88.
- . 2003. 'Perceptual Entitlement', *Philosophy and Phenomenological Research*, 67: 503-48.
- . 2013a. *Cognition through understanding: self-knowledge, interlocution, reasoning, reflection* (Oxford University Press: Oxford).
- . 2013b. 'Content Preservation.' in *Cognition through understanding: self-knowledge, interlocution, reasoning, reflection* (Oxford University Press: Oxford).
- Christensen, David. 2007. 'Epistemology of disagreement: the good news', *Philosophical Review*, 116: 187.
- . 2009. 'Disagreement as Evidence: The Epistemology of Controversy', *Philosophy Compass*, 4: 756-67.
- . 2010. 'Higher-Order Evidence', *Philosophy and Phenomenological Research*, 81: 185-215.
- . 2011. 'Disagreement, Question-Begging and Epistemic Self-Criticism', *Philosophers' Imprint*, 11.
- . 2013. 'Epistemic Modesty Defended.' in D. Christensen and J. Lackey (eds.), *The Epistemology of Disagreement: New Essays* (Oxford University Press: Oxford).
- . 2016a. 'Conciliation, Uniqueness and Rational Toxicity', *Noûs*, 50: 584-603.
- . 2016b. 'Disagreement, Drugs, Etc.: From Accuracy to Akrasia', *Episteme: A Journal of Individual and Social Epistemology*, 13: 397-422.
- Chuard, Philippe, and Nicholas Southwood. 2009. 'Epistemic Norms without Voluntary Control', *Noûs*, 43: 599-632.
- Clarke, Brendan, Donald Gillies, Phyllis Illari, Federica Russo, and Jon Williamson. 2014. 'Mechanisms and the Evidence Hierarchy', *Topoi*, 33: 339-60.
- Coates, Allen. 2012. 'Rational epistemic akrasia', *American Philosophical Quarterly*, 49: 113-24.

- Coliva, Annalisa. 2016. 'Which Hinge Epistemology?', *International Journal for the Study of Skepticism*, 6: 79-96.
- Dipaolo, Joshua. 2018. 'Higher - Order Defeat is Object - Independent', *Pacific Philosophical Quarterly*, 99: 248-69.
- Dotson, Kristie. 2011. 'Tracking Epistemic Violence, Tracking Practices of Silencing', *Hypatia*, 26: 236-57.
- . 2012. 'A Cautionary Tale: On Limiting Epistemic Oppression', *Frontiers: A Journal of Women Studies*, 33: 24-47.
- . 2014. 'Conceptualizing Epistemic Oppression', *Social Epistemology*, 28: 115-38.
- Elga, Adam. 2007. 'Reflection and Disagreement', *Noûs*, 41: 478-502.
- Enoch, David. 2010. 'Not Just a Truthometer: Taking Oneself Seriously (but not Too Seriously) in Cases of Peer Disagreement', *Mind*, 119: 953-97.
- Faulkner, Paul. 2011. *Knowledge on Trust* (Oxford University Press: Oxford).
- . 2016. 'Agency and Disagreement.' in P. Reider (ed.), *Social Epistemology and Epistemic Agency: Decentralizing Epistemic Agency* (Rowman and Littlefield International: London).
- Feldman, Richard. 1988a. 'Epistemic Obligations', *Philosophical Perspectives*, 2: 235-56.
- . 1988b. 'Subjective and Objective Justification in Ethics and Epistemology', *The Monist*, 71: 405-19.
- . 2001. 'Voluntary belief and epistemic evaluation.' in M. Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue* (Oxford University Press: Oxford).
- . 2006. 'Epistemological puzzles about disagreement.' in S. Hetherington (ed.), *Epistemology Futures* (Oxford University Press).
- . 2007. 'Reasonable religious disagreements.' in L. Antony (ed.), *Philosophers Without Gods: Meditations on Atheism and the Secular Life* (Oxford University Press).
- . 2009. 'Evidentialism, Higher-Order Evidence, and Disagreement', *Episteme*, 6: 294-312.
- . 2014. 'Evidence of Evidence is Evidence.' in J. Matheson and R. Vitz (eds.), *The Ethics of Belief* (Oxford University Press: Oxford).
- Fitelson, Branden. 2012. 'Evidence of evidence is not (necessarily) evidence', *Analysis*, 72: 85-88.
- Fitelson, Branden, and David Jehle. 2009. 'What is the "Equal Weight View"?' *Episteme: A Journal of Individual and Social Epistemology*, 6: 280-93.
- Fogelin, Robert. 1985. 'The Logic of Deep Disagreements', *Informal Logic*, 7: 3-11.
- Foley, Richard. 2001. *Intellectual Trust in Oneself and Others* (Cambridge University Press: London).
- Fricker, Elizabeth. 2016. 'Doing (Better) What Comes Naturally: Zagzebski on Rationality and Epistemic Self-trust', *Episteme: A Journal of Individual and Social Epistemology*, 13: 151-66.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford University Press: Oxford).
- Gerken, Mikkel. 2013. *Epistemic Reasoning and the Mental* (Palgrave Macmillan Basingstoke).
- Holton, Richard. 1994. 'Deciding to trust, coming to believe', *Australasian Journal of Philosophy*, 72: 63-76.
- Horowitz, Sophie. 2014. 'Epistemic Akrasia', *Noûs*, 48: 718-44.
- Hurley, Susan. 1989. *Natural reasons : personality and polity* (Oxford University Press: Oxford).
- Jones, Karen. 1996. 'Trust as an Affective Attitude', *Ethics*, 107: 4-25.
- . 2012a. 'The Politics of Intellectual Self-trust', *Social Epistemology*, 26: 237-51.
- . 2012b. "Response to Nancy Daukas." In, 1-7. Social Epistemology Review and Reply Collective.
- Kappel, Klemens. 2012. 'The Problem of Deep Disagreement', *Discipline Filosofiche*, 22: 7-25.
- . 2018. 'Higher Order Evidence and Deep Disagreement', *Topoi*: 1-12.
- Kelly, Thomas. 2005. 'The epistemic significance of disagreement.' in J. Hawthorne and T. Gendler (eds.), *Oxford Studies in Epistemology, Volume 1* (Oxford University Press: Oxford).

- . 2010. 'Peer Disagreement and Higher - Order Evidence.' in R. Feldman and T. A. Warfield (eds.), *Disagreement* (Oxford University Press: Oxford).
- . 2013. 'Disagreement and the Burdens of Judgment.' in D. Christensen and J. Lackey (eds.), *The Epistemology of Disagreement: New Essays* (Oxford University Press: Oxford).
- . 2014. 'Believers as Thermometers.' in J. Matheson and R. Vitz (eds.), *The Ethics of Belief* (Oxford University Press: Oxford).
- Kidd, Ian James, and José Medina (eds.). 2017. *The Routledge Handbook of Epistemic Injustice* (Routledge: London).
- King, Nathan L. 2012. 'Disagreement: What's the Problem? or A Good Peer is Hard to Find', *Philosophy and Phenomenological Research*, 85: 249-72.
- . 2016. 'Religious Skepticism and Higher- Order Evidence.' in J. Kvanvig (ed.), *Oxford Studies in Philosophy of Religion: Volume 7* (Oxford University Press: Oxford).
- Knight, Frank H. 1921. *Risk, Uncertainty and Profit* (Houghton Mifflin Company: Boston).
- Kripke, Saul A. 1982. *Wittgenstein on rules and private language: an elementary exposition* (Blackwell: Oxford).
- . 2011. 'On Two Paradoxes of Knowledge.' in, *Philosophical Troubles. Collected Papers Vol I* (Oxford University Press: Oxford).
- Kvanvig, Jonathan L. 1984. 'Subjective Justification', *Mind*, 93: 71-84.
- Lackey, Jennifer. 2010. 'A Justificationist View of Disagreement's Epistemic Significance.' in A. Haddock, A. Millar and D. Pritchard (eds.), *Social Epistemology* (Oxford University Press: Oxford).
- Lasonen - Aarnio, Maria. 2014. 'Higher - Order Evidence and the Limits of Defeat', *Philosophy and Phenomenological Research*, 88: 314-45.
- Littlejohn, Clayton, and John Turri. 2014. *Epistemic Norms: New Essays on Action, Belief, and Assertion* (Oxford University Press: Oxford).
- Lynch, Michael P. 2010. 'Epistemic Circularity and Epistemic Incommensurability.' in A. Haddock, A. Millar and D. Pritchard (eds.), *Social Epistemology* (Oxford University Press: Oxford).
- Lynch, Michael P., and Paul Silva. 2016. 'Why Worry about Epistemic Circularity?', *Journal of Philosophical Research*, 41: 33-52.
- Matheson, Jonathan. 2009. 'Conciliatory Views of Disagreement and Higher-order Evidence', *Episteme: A Journal of Individual and Social Epistemology*, 6: 269-79.
- . 2014. 'Disagreement: Idealized and Everyday.' in J. Matheson and V. Ritz (eds.), *The Ethics of Belief* (Oxford University Press: Oxford).
- . 2015. *The Epistemic Significance of Disagreement* (Palgrave Macmillan: New York).
- . 2018. 'Deep Disagreements and Rational Resolution', *Topoi*: 1-13.
- McCraw, Benjamin W. 2015. 'The Nature of Epistemic Trust', *Social Epistemology*, 29: 413-30.
- McDowell, John. 1978. 'Are Moral Requirements Hypothetical Imperatives?', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 52: 13-42.
- . 1979. 'Virtue and Reason', *The Monist*, 62: 331-50.
- . 1984. 'Wittgenstein on following a rule', *Synthese*, 58: 325-63.
- . 1993. 'Knowledge by hearsay.' in A. Chakrabarti and B. K. Matilal (eds.), *Knowing From Words* (Kluwer Academic Publishers).
- . 1998. *Mind, value, and reality* (Harvard University Press: Cambridge, Mass. ).
- McGinn, Marie. 2013. *The Routledge Guidebook to Wittgenstein's Philosophical investigations* (Routledge: New York).
- Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations* (New York : Oxford University Press, 2013: New York).
- Möllering, Guido. 2001. 'The Nature of Trust: From Georg Simmel to a Theory of Expectation, Interpretation and Suspension', *Sociology*, 35: 403-20.
- Moran, Richard. 2018. *The Exchange of Words: Speech, Testimony, and Intersubjectivity* (Oxford University Press: Oxford).

- Neta, Ram. 2014. 'The epistemic "ought".' in O. F. Abrol Fairweather (ed.), *Naturalizing Epistemic Virtue* (Cambridge University Press: Cambridge).
- Pasnau, Robert. 2015. 'Disagreement and the value of self-trust', *Philosophical Studies*, 172: 2315-39.
- Pollock, John L. 1986. *Contemporary theories of knowledge* (Rowman & Littlefield: London).
- . 1987. 'Defeasible Reasoning', *Cognitive Science*, 11: 481-518.
- Pritchard, Duncan. 2018. 'Wittgensteinian Hinge Epistemology and Deep Disagreement', *Topoi*: 1-9.
- Ranalli, Chris. 2018a. 'Deep disagreement and hinge epistemology', *Synthese*: 1-33.
- . 2018b. 'What is Deep Disagreement?', *Topoi*: 1-16.
- Rasmussen, Mattias Skipper, Asbjørn Steglich-Petersen Steglich-Petersen, and Jens Christian Bjerring. 2018. 'A Higher-Order Approach to Disagreement', *Episteme*, 15: 80-100.
- Salmon, Wesley. 1991. 'Hans Reichenbach's vindication of induction', *Erkenntnis*, 35: 99-122.
- Schafer, Karl. 2015. 'How Common is Peer Disagreement? On Self - Trust and Rational Symmetry', *Philosophy and Phenomenological Research*, 91: 25-46.
- Schoenfield, Miriam. 2018. 'An Accuracy Based Approach to Higher Order Evidence', *Philosophy and Phenomenological Research*, 96: 690-715.
- Silva, Paul. 2017. 'How Doxastic Justification Helps Us Solve the Puzzle of Misleading Higher - Order Evidence', *Pacific Philosophical Quarterly*, 98: 308-28.
- Sliwa, Paulina, and Sophie Horowitz. 2015. 'Respecting all the evidence', *Philosophical Studies*, 172: 2835-58.
- Sorensen, Roy A. 1988. 'Dogmatism, Junk Knowledge, and Conditionals', *The Philosophical Quarterly*, 38: 433-54.
- Sosa, Ernest. 2010. 'The Epistemology of Disagreement.' in A. Haddock, A. Millar and D. Pritchard (eds.), *Social Epistemology* (Oxford University Press: Oxford).
- Sripada, Chandra Sekhar, and Stephen Stich. 2012. 'A Framework for the Psychology of Norms.' in P. Carruthers, S. Laurence and S. Stich (eds.), *A Framework for the Psychology of Norms* (Oxford University Press: Oxford).
- Wedgwood, Ralph. 2010. 'The Moral Evil Demons.' in R. Feldman and T. A. Warfield (eds.), *Disagreement* (Oxford University Press).
- White, Roger. 2005. 'Epistemic Permissiveness', *Philosophical Perspectives*, 19: 445-59.
- . 2013. 'Evidence Cannot Be Permissive.' in M. Steup and J. Turri (eds.), *Contemporary Debates in Epistemology* (Blackwell).
- Wittgenstein, Ludwig. 1997. *Philosophical investigations* (Blackwell: Oxford).
- . 2013. *On Certainty* (Blackwell Publishing: Oxford).
- Wolterstorff, Nicholas. 2010. 'Ought to believe: two conceptions.' in T. Cuneo (ed.), *Practices of Belief*.
- Worsnip, Alex. 2018. 'The Conflict of Evidence and Coherence', *Philosophy and Phenomenological Research*, 96: 3-44.
- Wright, Crispin. 1980. *Wittgenstein on the foundations of mathematics* (Duckworth: London).
- . 2004. 'On Epistemic Entitlement', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 78: 167-245.
- . 2014. 'On Epistemic Entitlement (II): Welfare State Epistemology.' in D. Dodd and E. Zardini (eds.), *Scepticism and Perceptual Justification* (Oxford University Press: Oxford).
- Zagzebski, Linda Trinkaus. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford University Press: Oxford).