

Patrizio, Ross Finn (2026) *Testimony, trust, and teleology: essays on the epistemology and ethics of interpersonal deference.* PhD thesis.

https://theses.gla.ac.uk/86356/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Testimony, trust, and teleology

Essays on the epistemology and ethics of interpersonal deference

Ross Finn Patrizio
MSc, M.A.

A thesis presented for the degree of
Doctor of Philosophy



School of Humanities
College of Arts
University of Glasgow
April 2026

Abstract

We are profoundly social creatures – deeply dependent on others across so many of the projects and plans that shape our lives. This thesis can be seen as a small contribution to the much larger task of making philosophical sense of the epistemological and ethical dimensions of this dependence. 'Testimony' in the title refers to the epistemological component, which concerns the nature of our dependence *on the word* of others. 'Trust' refers to the ethical component, which concerns the attitude whose adoption constitutes an indispensable yet perplexing manifestation of this dependence. Finally, 'teleology' refers to a crucial assumption I adopt throughout: that we are not only social creatures but *goal-directed* ones. I argue that we make better sense of the epistemological and ethical dimensions of interpersonal deference by first making sense of the ways that our aims and goals normatively govern the practices in which we participate. It is only with these aims and goals in clear view that we may begin to shed light on how – via healthy channels of testimony and trust – we help each other to realise them.

Table of contents

TESTIMONY, TRUST, AND TELEOLOGY 1
ABSTRACT 2
ACKNOWLEDGEMENTS 5
AUTHOR'S DECLARATION 8
STATEMENT OF LENGTH 9
PUBLISHED MATERIAL 10
INTRODUCTION 11
CHAPTER-BY-CHAPTER SUMMARY 15
CHAPTER I 25
TESTIMONIAL LIBERALISM AND THE BALANCE OF EPISTEMIC GOALS 25
§1 Introduction *25*
§2 Pinpointing the debate *28*
§3 Changing the metric *30*
§4 Liberalism and productivity *33*
§5 Liberalism and reliability *35*
§6 Conclusion *40*
CHAPTER II 41
ON THE ZETETIC VALUE OF TESTIMONY 41
§1 Introduction *41*
§2 Complex testimony *42*
§3 Norms on complex testimony *45*
§4 Testimony and inquiry *47*
§5 Zetetic epistemology of testimony *49*
§6 Zetetic dependence *52*
§7 Overinflation of the epistemic realm *55*
§8 Conclusion *57*
CHAPTER III 58
GRICEAN MAXIMS AS EPISTEMIC NORMS 58
§1 Introduction *58*
§2 Merely conversational failings *59*
§3 Gricean epistemic norms *62*
§4 Objections and replies *76*
§5 Conclusion *80*
CHAPTER IV 82
TRUST, REASONS, AND RATIONAL CONTROL: A DEFENCE OF THE BELIEF-LIKE VIEW 82
§1 Introduction *82*
§2 The significance of therapeutic trust *84*
§3 Theory-neutral constraints *87*
§4 Explaining the appeal of ALV *95*
§5 Conclusion *98*
CHAPTER V 99
TRUST'S MENO PROBLEM: CAN THE DOXASTIC VIEW ACCOUNT FOR THE VALUE OF TRUST? 99
§1 Introduction *99*
§2 The doxastic view and the trust Meno problem *101*
§3 Interpreting Trust's Meno Problem *104*
§4 The Final Value of Trust *109*
§5 Conclusion *113*

CHAPTER VI. 114
THE GENERALITY PROBLEM AND THE C-TYPING SOLUTION 114
§1 Introduction *114*
§2 The C-Typing Solution *116*
§3 Counterexamples to the C-Typing Thesis *118*
§4 Counterexamples from implicit bias *122*
§5 Conclusion *127*
REFERENCES 128

Acknowledgements

I couldn't have written this thesis without the help, support and encouragement of many people – undoubtedly more than I'll manage (or remember) to mention here. One person I definitely shouldn't forget is my primary supervisor, Chris Kelp. It's hard to overstate how supported I've felt throughout the past five years, and how much I've enjoyed talking about epistemology with Chris. Same goes for Adam Carter and Mona Simion, who read countless drafts of papers, discussed new ideas with me, and were otherwise always willing to lend an ear and offer much-needed advice along the way. It scares me to think that I might never have taken Adam's Virtue Epistemology class in my undergraduate, and might then never have taken Chris' or Mona's classes in my MSc.

Looking back on my (long) time at Glasgow, perhaps the single thing I am most grateful for is that it overlapped with Katie Moody's. Meeting Katie during lockdown changed my life in the best possible way and if there are any mistakes in anything that follows they can be reliably traced back to her. I hope that we'll continue to do everything together, including philosophy, for a very long time. Next is Rory Aird. It is difficult to express just how much more enjoyable the whole process has been for having a friend like Rory by my side at every step of the way. I'm not confident that I would've even enjoyed doing a PhD that didn't involve us reading every word of each other's writing, frantic back-and-forths in the office, or going to the pub afterwards (and during, and sometimes before). Last but not least, Gonca Yalcin. I'll always be grateful that our lockdown library job brought us together when it did, and will never forget the day that we were on shift together when we both found out that our scholarship applications had been successful. If we hadn't gone on to spend most of that money in Little Italy we probably would in fact be as rich as we thought we were about to be at the time.

There have been so many wonderful people at Glasgow throughout my time there. Thanks to the staff who offered generous support and guidance at so many crucial points throughout: Katharine Jenkins, Joe Slater, Glen Pettigrove, Jack Lyons, Ben Colburn, Neil McDonnell, Emma Gordon, Nico Cote, Robert Cowan, and Bryan Pickel. Special thanks to the COGITO postdocs, from whom I learnt so much and who I'm now proud to call friends: Matthew Vermaire, Chris Willard-Kyle, Giada Fratantonio, Spencer Paulson, Matt Jope, Tim Kearl, and Oscar Piedrahita. To the SEECRs team, Lilith Mace, Dario Mortini, and Angie O'Sullivan: thanks for having me as part of the group – SEECRs was such an important part of my PhD experience and I'm very proud of the small community we created. And beyond just SEECRs, I am forever indebted to Lilith, Dario, and Angie for supporting and inspiring me throughout such a formative period; it was in no small part through chatting and interacting with real philosophers like them that I began to think I might one day like to be one. Further thanks to my wonderful PhD colleagues at Glasgow over the years – definitely too many to name but they know who they are. Oakfield Avenue, and our basement offices with the barred windows, believe it or not, made for the best possible PhD environment.

I've been extremely lucky to have an incredibly strong and ever-present support network outside of philosophy. Firstly to Humour and its surrounding team: Andreas, Jack, Ruairidh, Lewis, Mark,

Rosie, Joe, James and Rod, who, whether they realise it or not, all played a crucial role in shaping and facilitating the completion of this thesis. They all made sacrifices at various crucial times, no matter what we had on or where in the world we were, that made it possible for me to balance two pursuits that mean so much to me. Sometimes, albeit only on the particularly long drives, they even asked me what I was writing about. It feels apt that I'm currently putting the finishing touches on the thesis from the back of the van – where so much of it was written – as we drive from Madrid to Lisbon and everyone else fights over a quiz.

One person who doesn't fit neatly in any of my carefully curated categories but whom I nonetheless want to mention is Emilios Christodoulidis. Emilios was an ongoing source of inspiration – throughout my life but particularly since starting university – he always showed a keen and genuine interest in my work, and I'll always cherish the memory of our conversations spent trying to establish mutual academic ground. I will miss him dearly.

Lastly and perhaps most importantly, I want to thank my amazing family: my parents for their unwavering and unquestioning love and support across not one but two highly (economically) questionable long-term decisions; and Matthew, for being the best brother anyone could wish for. As they each know and only remind me very occasionally, there is absolutely no chance that I could have done this without them. Thanks also to my two wonderful grandads, *grandad* Joe and *granda* Cecil, who each supported and encouraged me in their own unique ways. And, of course, to Evie and Oscar, for the furry and fragrant companionship along the way.

For Anne and Shirley.

Author's declaration

I declare that all the work in this dissertation is the result of my own efforts, except where properly indicated by means of quotations and references. This work has not been submitted for any other degree of professional qualification. This work has received generous financial support from the Scottish Graduate School for Arts and Humanities.

Statement of length

This dissertation, including footnotes and references, contains around 71,000 words. It meets the minimum word limit set by the University of Glasgow College of Arts (70,000) and does not exceed the maximum word limit (100,000).

Published material

At the time of submission, adapted versions of Chapters I and V have been published in the following journals:

- Patrizio, R. F. (2024a). Testimonial liberalism and the balance of epistemic goals. *Philosophical Studies* 181 (11):2997-3015.

- Patrizio, R. F. (2024b). Trust's Meno problem: Can the doxastic view account for the value of trust? *Philosophical Psychology* 37 (1):18-37.

Introduction

This thesis comprises six independent yet interrelated chapters pertaining to the epistemological and ethical dimensions of interpersonal deference. There is something normatively significant that occurs when we place our trust in others; either by trusting the accuracy and sincerity of their testimony, or by trusting them to carry out some action. The key contribution of this thesis is to nudge forward our understanding of the epistemological and ethical dimensions of this normatively special relationship between trustor and trustee.

Before laying out, chapter by chapter, how the thesis proceeds, here I offer some big-picture, stage-setting remarks regarding each of the three key words in its title. Some context with respect to the three core components of the thesis should help the reader to understand its structure and to appreciate how I see these components hanging together and interacting with one another.

The first three chapters are primarily concerned with the epistemology of testimony. In this part of the thesis my main concern is the normative, epistemic relationship between speaker and hearer in a testimonial exchange. In other words, I am specifically interested in the normative significance of placing trust *in the word* of others. My contribution in these chapters is, I think, best understood as a tripartite plea for *expansion* in the epistemology of testimony. I think that the epistemology of testimony has, in a couple of distinct but interrelated ways, been problematically oversimplified – too narrow in its focus – to make good sense of the speaker-hearer relationship. The simple model of testimony, assumed without argument across vast swathes of the extant literature, can be summed up, simply for now, as the idea that testimonial exchanges essentially revolve around speakers' assertions and hearers' resulting testimony-based beliefs. The ideal, or model testimonial exchange, around which our normative theories are built, involves a speaker S, who asserts that p, where p is some simple proposition (e.g., "there's milk in the fridge", or "Lisbon is the capital of Portugal"), and a competent hearer, H, who on that basis comes to believe – ideally also to *know* – that p.

My contributions to the epistemology of testimony in this thesis can all be understood as interacting in different ways with this simple model of testimony. However, there are a number of different ways in which one might brush up against this model. Some recent work (in particular, by Fraser (2021), Keller (2024), and Hazlett (2025)) delineates concrete instances of testimony that fall outside of its narrow scope. *Narrative* testimony, for example, and testimony *by presupposition*, are supposed to be ways in which we engage in the practice of testifying which do not obviously or straightforwardly reduce to the simple assertions and beliefs therein. Other recent work, across epistemology and the philosophy of language, instead pushes back by casting doubt over the idealistically smooth process – often simply assumed in the epistemology of testimony – from assertion with particular propositional contents to a resulting testimonial belief with matching contents.[1]

[1] See, e.g., Goldberg (2007); Buchanan (2010); Peet (2015a; 2015b; 2016; 2023); MacFarlane (2020a; 2020b); Pollock (2021; 2023; 2024); and Abreu Zavaleta (2021).

I follow neither of these strategies in what follows, though I engage with them on occasion and in passing (particularly in Chapter 2). Instead, my focus is straightforwardly on the epistemic normativity of testimony. My concern with the model, though I press it in different ways, is that the normative picture that emerges from it is mistaken. I argue that the normative resources it equips us with for epistemically evaluating testimony are unfit for purpose.

'Trust' in the title of this thesis accounts for the *ethical*, as opposed to *epistemological*, component that runs through what is to come. If, as seems plausible to me, the epistemology of testimony is to be understood as a study of one particular subspecies of the practice of trusting others – namely that of trusting *their word* – then from this vantage point I became interested in the more general practice or attitude that undergirds the epistemological one. The guiding question is a simple one: what is the nature of trust? In particular, how does it differ from nearby attitudes that seem in various ways less lofty, less highbrow, less impressive - such as, for example, mere reliance? The trust component of this thesis is thus an attempt to better understand the nature of the attitude itself, by engaging with the broad and dispersed array of literature in the philosophy of trust, which spans ethics, action theory, and epistemology. Interestingly, if the epistemological component of this thesis is understood as a plea for expansion, then the ethical component is quite the opposite: a plea for *quietism*.

To better understand the approach, consider a natural and attractively simple view of trust's nature: trust is essentially belief-like, or cognitive. A cannot trust B to φ unless they believe that B is trustworthy with respect to φ. It does not matter, on such a view, how much I really wish I could trust my sociopathic boss, if I do not in this moment judge him to be trustworthy, then I cannot overcome my doubt by sheer force of will and choose to trust. Trust takes time and effort. It does not and cannot simply jump into existence at the click of my fingers. Trust need not be *mere* belief, on such a picture – it may be special in virtue of whichever additional component that, along with the cognitive one, constitutes it – but it is an attractively simple picture nonetheless. Trust is not particularly metaphysically or normatively spooky or idiosyncratic. It is an attitude that by its nature seeks to pick out a valuable property in others, and it does its job when it accurately picks out that property.

Many people deny this simple picture. The main ambition in the ethical, trust-oriented component of this thesis can be understood as an attempt to restore it. Two, key putative frailties in the cognitivist's picture are addressed. First, perhaps the most pervasive and influential objection to the cognitivist view: alleged cases of *willed* or *therapeutic* trust – cases in which it feels as though one is capable of directly *choosing to trust*, in a way that would not be possible if trust were belief, given the assumption of doxastic involuntarism. The other putative frailty concerns the special and distinctive value of trust. How, if trust is a species of belief, can we account for its indispensable value, given that one can reap all of trust's rewards by simply *acting as if* one has the relevant trusting belief, and one need not *actually have it*.

My plea for *quietism* in the ethical component of this thesis is thus at its heart an essentially negative thesis: neither of these considerations give us reason to abandon the simple, cognitive theory of the nature of trust. There are, of course, other objections to the cognitivist's view – nothing in what follows comes with the promise or pretence of comprehensiveness – but these two

objections seemed to me to be particularly important. The first, on therapeutic trust, has been extremely influential. As I mentioned, it is perhaps *the* single biggest thorn in the side of the cognitivist view. The second, on the value of trust, while more recent and less influential, is an incisive one. It seems, at least prima facie, to highlight or draw attention to a genuine limitation of the cognitivist's view. While I argue that this worry turns out to be misplaced, it nonetheless warrants careful, considered discussion. Another thesis would address other objections, perhaps set out my own version of the view, fleshed out in more detail, but that thesis is not this one; I leave plenty of work to be done in pursuit of a full, rigorous defence of the cognitivist view.

Lastly by way of big-picture preamble, a word on teleology. Teleology in epistemology refers to the massively influential idea that epistemic normativity is not independent of our epistemic goals. This tradition builds on and stems from teleology in ethics, which can be helpfully characterised, in slogan form, as the view according to which "…the good is prior to the right" (Berker 2013). As Berker (2013) notes, teleology has been dominant in epistemology; we overwhelmingly tend to place the *epistemic* good prior to the *epistemic* right. What this means, more specifically, is that facts about how we epistemically ought and ought not to behave are *grounded in* facts about the value of being in certain cognitive states, i.e., of having true beliefs, knowledge, and understanding. The precise epistemic norms that we have, that we abide by, and that we hold ourselves and others accountable for, are determined at least in part by the conduciveness to and realisation of epistemically valuable states in creatures like us. There is a myriad of different manifestations of this broad approach in epistemology, many of which constitute some of the most popular and prominent views in the current literature, e.g., virtue epistemology, reliabilism, functionalism, zetetic epistemology, among others. My own preferred framework is in zetetic epistemology, which places our practice of *inquiring* at the front and centre of epistemological theorising. Epistemic norms, on this view, are norms of inquiry: they are the rules that we follow in order to bring about the proper aims of inquiry. However, I will make clear and explicit the sections in which I draw specifically on *zetetic* – as opposed to broad, teleological – epistemology.

Teleology is best understood as constituting the *backdrop* as opposed to the *subject matter* of this thesis. The thesis is not *about* epistemic normativity per se, so I will not defend the teleological tradition from its rivals directly – though some of Berker's (2013) own challenges are addressed, particularly when they are found to have bite against my particular deployment of teleological epistemology in the domain of testimony and trust. The influence of the framework is more or less ever-present in everything that follows. Rather than defending it, then, for the most part I simply assume it and use it to demonstrate current theorising surrounding interpersonal dependence do not offer fully apt resources for explaining how it is that creatures like us realise value via our practices of testimony and trust.

This is helpful to have in mind throughout – particularly in the testimony-oriented chapters – and it also explains the pertinence of the final chapter of the thesis, "The Generality Problem and the C-Typing Solution". The Generality Problem is a well-known, much-discussed issue. It is commonly levelled specifically at *reliabilist*[2] theories of justification, though its pertinence spans far wider, and perhaps even to *all* teleological views. Roughly, the worry is that if what ultimately

[2] See Goldman (1979) for the locus classicus.

matters (epistemically, normatively speaking) are systems, processes, or practices that bring about epistemic value, then the teleologist faces a problem: at what level of specificity/generality are those processes to be defined? This worry looks particularly pressing given that, depending on how they are to be specified, it looks like you get different results regarding their epistemic, normative status (i.e., whether they do or do not conduce to the epistemic good). My interest in this problem stems from my investment in and belonging within the broad teleological tradition.

With this preamble complete, let me now summarise what is to come, chapter-by-chapter.

Chapter-by-chapter summary

Chapter I. Testimonial liberalism and the balance of epistemic goals

The first chapter concerns by far the most widely discussed question in the epistemology of testimony, namely that of testimonial justification. Under what conditions is some hearer, H, epistemically permitted to believe that p on the basis of some speaker, S's, say-so? Two broad camps, liberals and conservatives, emerged early (with Hume and Reid, respectively) and have dominated discussion ever since. Testimonial Liberalism, as I call it, is the view that testimonial justification simply requires an absence of defeaters. In other words, mere say-so is enough, provided one lacks any reason to think that something is epistemically awry – hence 'liberalism'. Testimonial Conservatism, on the other hand, is the view that mere say-so is not enough; testimonial justification requires independent, non-testimony-based reasons for thinking that one's interlocutor is right, or at least sufficiently likely to be telling the truth.

My novel contribution in this chapter is twofold: I introduce a new metric for settling the debate between liberals and conservatives about testimonial justification; then demonstrate that, on this new metric, the liberal view is stronger than has been appreciated. To understand the argument, consider the most prominent and natural objection to the liberal view: the view is too permissive, one lacks defeaters for all sorts of people that one nevertheless should not trust. A mere absence of defeaters is in this way too weak a condition to track the relevant normative, epistemic property of justification.

However, extant ways of levelling this criticism at the liberal view smuggle in a sceptical consideration that all parties to the debate would, and should, reject. In particular, as I show, they smuggle in the idea that one epistemically ought to do all that one can to reduce the risk of false belief. But, on pain of scepticism, this is false. There are limits. One should not overreact to the risk of error such that one precludes the attainment of true beliefs, for example by locking oneself away or closing one's eyes and ears. What needs to be shown, then, is not that the liberal view is more permissive than the conservative view (we already knew that it was). Rather, what needs to be shown is that the conservative view does a better job of *balancing our epistemic goals*, that is, of attaining truth and avoiding error. On this way of looking at things, however, the conservative view does no such thing. What matters is not avoidance of error but rather balance, and balance favours the liberal. Or so I argue in this opening chapter.

In the introduction I mentioned that the first three chapters of this thesis, all on the epistemology of testimony, are helpfully understood as interacting with the simple model of testimony. Recall that this model assumes that testimony has essentially to do with *assertion* and *belief*, that in an ideal testimonial exchange a speaker asserts truly *that p*, and, all being well, the hearer can come to know *that p*. Now, in this first chapter I granted this simple model. I allow that the target process is a relatively straightforward one, involving for the most part simple assertions and simple beliefs. Still, even granting this, our metric has been too simple. We have evaluated competing theories of

testimonial justification using a metric that tracks only one part of what epistemically matters. In the next chapter, I abandon the assumption of the simple model and instead attack it head-on.

Chapter II. On the zetetic value of testimony

We epistemically depend on one another in a range of different ways. I epistemically rely on you to shape my thinking, to shift my perspective, to focus my attention, to bring me to understand, and to influence my beliefs. Traditional theories of testimony, however, proceed as though epistemic dependence takes only the latter form. As I have already alluded to, they in one way or another take as our methodological starting point the idea that testimony is, at its heart, a *vehicle for propositional knowledge*. As a result, most of the ink has been spilt on our practices of *assertion* – on the speaker's side of the testimonial exchange – and *testimonial belief* – on the hearer's. But if, as seems eminently plausible, we do more in testimonial exchanges than simply assert that p and thereby bring our hearer to believe that p, then this model will not do. If, in communication, we can also epistemically influence hearers' *understanding*, their patterns of *attention*, their *perspective* on some topic, then there is no reason, at least in principle and from the outset, that our theorising about the epistemological significance of testimony should proceed whilst excluding such elements of our testimonial practices.

In this chapter, I proceed in the following way: let us suppose that the traditional model of testimony is indeed overly simple – that in testifying we do more that is of epistemic significance than merely transmit propositions to one another. How should we proceed? What is the upshot for normative theorising about testimony? In the growing literature that raises worries about the simple model of testimony, it has been suggested that by assuming it we are bound to get the epistemic normativity wrong. This is most explicit in Fraser (2021), who notes that "…a normative epistemology informed by such a [simple] picture will be, at best, incomplete and at worst, distorted." (2021: 4026) But how so? Why exactly will extant norms – of the kind that have dominated our discussions – fail to 'scale up', as it were, to more complex exchanges than those they were built around? Noting that we have engaged in oversimplification does not yet make a normative problem precise.

This chapter thus offers two novel contributions: a careful diagnosis of the normative problem and a novel, inquiry-based solution. By way of diagnosis, I argue that a concrete normative problem arising from oversimplification has to do with *value*. The problem with the traditional model is that it affords us an impoverished picture of the epistemic value of testimony. On at least one (albeit hugely influential) theory of epistemic normativity, that is, the teleological one I outlined in the introduction, it is all but guaranteed that from this starting point normative problems will follow. On this view, there is an intimate link between the norms that are supposed to govern our conduct, on the one hand, and the value, good, or goal that sets the standard according to which that conduct is evaluated, on the other. In taking the simple model as our starting point, we overlook a range of ways in which testimony, qua socio-epistemic practice, can realise epistemic value. In this way, then, the norms that are built from this starting point are indeed, as Fraser (2021) warned, at best, incomplete. The ever-present and near-unanimous focus on our practices of *assertion* and

belief assume, even if only implicitly, that *these are the only epistemically significant practices that speakers and hearers engage in via testimonial exchanges.*

Then, having diagnosed and made precise the normative problem that arises once we notice that testimony takes many forms beyond its most simple, I offer the outlines of a solution. My solution deploys zetetic epistemology, the approach to normative epistemology that places our practice of *inquiry* front-and-centre. Whereas the bulk of traditional epistemology focuses on *belief* as the sole locus of epistemic evaluation, we should, on a zetetic approach, instead focus on our practice of *settling questions*, where the latter involves a range of different inquiring activities before (and perhaps also after) a belief is formed. We gather evidence, we reformulate questions, we direct our attention toward particular features of our object, and more - all in the process of inquiring. Epistemology should no more concern itself with the culmination point of inquiry than everything else that the process involves.

I argue that zetetic epistemology offers uniquely sharp resources for explaining testimony's epistemic value, in particular when compared to the doxastic paradigm it seeks to supplant. If we think of testimony not as merely a *source of belief,* but rather a *resource in inquiry*, then we do a better job of accounting for its epistemic role and significance. If we begin to think of speakers not merely as *asserters*, but more like *teachers* or *mentors* in inquiry, then we do better justice to the range of epistemically significant contributions they can and often do in fact make via testimony. And if we begin to think of hearers not merely as *believers* but rather as *inquirers*, then we better account for the full range of ways in which hearers epistemically depend on testimony for their picture of the world – beyond merely their beliefs about it.

Chapter III. Gricean maxims as epistemic norms

Chapter II advocated expansion in our conception of testimony's epistemic value, and subsequently in the range of epistemic norms governing speakers and hearers in a testimonial exchange. Chapter III explores in more detail one particular manifestation of that idea, specifically as it pertains to the speaker's side of the testimonial exchange. The literature on the epistemic normativity of the *offering* of testimony is dominated by the *epistemic normativity of assertion*. But assertion is an action, and as a result it is governed by a range of different norms (epistemic and non-epistemic alike). My target in this chapter is the received view on the intersection between epistemic and non-epistemic norms governing assertion. I understand and articulate the view in terms of Grice's (1975; 1989) influential theory of conversation, and dub it *Quality-exceptionalism* (QE). Grice's maxims are as follows:

> ***Quantity*** Make one's contribution no more or less informative than is required for the current purposes of the talk exchange.
> ***Quality*** Do not say either what you believe to be false or lack sufficient evidence for.
> ***Relation*** Be relevant.
> ***Manner*** Be perspicuous; avoid unnecessary obscurity or ambiguity.

According to QE, there is an *epistemic* norm corresponding to the maxim of Quality (irrespective of whether we choose to put it in Grice's terms exactly), but not Quantity, Relation, nor Manner. Those standards may govern assertion, but they do not do so *from the epistemic point of view*. Most commonly, they are chalked up to the domain of pragmatics – a somewhat mysterious normative domain, usually thought of as a kind of conversational subdomain within *practical* rather than *epistemic* normativity. In any case, QE, the received view, is clear: the norm governing assertion in accordance with standards such as *truth, justification, knowledge,* or *evidence*, is an epistemic one; the norms governing it in accordance with standards such as *relevance, quantity, clarity,* or *manner* are not.

The main argument in this chapter is that the received view is mistaken. On a range of well-established and independently motivated ways of thinking both about (i) epistemic normativity in general; and (ii) the ways in which it interacts with and governs our practice of assertion in particular, QE is unmotivated. On the contrary, there is no such neat and principled line to be drawn between the properly or distinctively epistemic norms governing assertion, on the one hand, and the merely pragmatic or conversational norms, on the other. Epistemically good testimony is, all else equal, not only true and well-evidenced, but also *relevant, clear,* and *perspicuous*. I call my view Gricean Epistemic Norms (GENs, for short).

My strategy in this chapter is to argue for this perhaps initially surprising view from relatively mundane starting assumptions. The main such assumption is that of Boundedness, that is, the idea that our limitations – both psychological and environmental – are normatively relevant; they make a difference to what and how creatures like us ought to think and believe. In short, if you like Boundedness, as many do, you ought also to like GENs. With Boundedness in place, I advance six interrelated arguments for GENs. The first has to do with epistemic accountability practices. Such practices are commonly thought to be indicative of underlying epistemic norms (Kauppinen 2018; Simion & Kelp 2021; Flores & Woodard 2023). We have such practices, I argue, for all Gricean maxims, not merely that of Quality. However, it is quite possible that we have such practices for illegitimate or non-epistemic reasons. Next, I thus demonstrate that not only do we have such practices, but that they can be *justified* or *rationalised* by association with epistemic, as opposed to merely practical or social, goals. This is to be expected only if they really are epistemic norms in the full-fledged sense.

I then develop and refine this argument from epistemic goals in a range of ways. I firstly draw on the Jamesian notion of twin epistemic goals (from Chapter I) to demonstrate that GENs help us to achieve a better *balance* between reliability and productivity – and are thereby conducive to epistemic goals on a more sophisticated conception of those goals. Then I shift from thinking about epistemic norms in general to the more specific matter of how they are supposed to shape and structure our practice of assertion in particular. Assertion, as is commonly noted both in epistemology and in philosophy of language, is epistemically evaluable by virtue of its *epistemic function*. It is because assertion functions to generate or transmit knowledge to hearers that we have epistemic norms on it in the first place. However, with this function, and thus the rationale undergirding its place in the realm of epistemic evaluation both in clear view, I argue that a practice of assertion regulated only in accordance with Quality is not yet fulfilling its epistemic function. GENs are required to ensure that not only are epistemic goods asserted, but subsequently successfully taken up.

Lastly, I consider two more radical arguments for GENs. All prior arguments adopted an uncomplicated conception of the epistemic goals that our practice of assertion is normatively governed by; all the arguments can be cashed out in terms of the epistemic goals of true belief, or propositional knowledge. But of course, these are not all that is of epistemic value – indeed, as was emphasised in Chapter II. I thus consider two ways of developing the operative conception of the relevant epistemic value(s) in developing epistemic norms on assertion, and suggest that they each strengthen rather than hinder the case for GENs. The first has to do with the epistemic value of *understanding*. On all accounts, understanding has final, epistemic value; and on most accounts, that value does not straightforwardly reduce to that of knowledge. Plausibly, then, our practice of assertion epistemically ought to conduce to the realisation not only of knowledge but also of understanding. Others can help us not only to know *what is the case*, but to *understand why* it is the case; they can provide us not only with the relevant facts, but also an explanation or account of how those facts hang together and interrelate. GENs, I argue, have a substantive epistemic role to play if what we seek is not only knowledge but understanding. Finally, I make the case on zetetic grounds, as is in-keeping with the rest of the thesis. If there is something behind the recent, so-called 'zetetic turn' in epistemology, then it is good news for the proponent of GENs. If we think that the locus of epistemic evaluation is not merely our practice of believing, but rather that of *inquiring*, then we will need a range of further norms, governing relevance, quantity of information, clarity and intelligibility in presentation, that all speak recognisably to standards of *good inquiry* – even if they do not straightforwardly influence *belief*. One may resist either or both of these two ways of expanding our conception of epistemic goals, in which case one must contend with the first four arguments. But if one is tempted by either of these increasingly popular ways of thinking about epistemic value, then the case for GENs is strengthened further still.

On multiple ways of thinking both about epistemic normativity in general, then, as well as the way it is supposed to shape and structure our practice of assertion, Quality-exceptionalism is undermined. At least for creatures like us, limited as we are, GENs help to squeeze our practice of assertion for all its epistemic worth.

Chapter IV. Trust, reasons, and rational control

Chapter IV represents the beginning of a foray into the *ethical*, as opposed to *epistemological* dimensions of interpersonal deference. In this chapter and the next, I am concerned with the nature and normativity of the trusting attitude itself – irrespective of its particular application to epistemology. As noted in the introduction, I take a cognitive, or doxastic view of trust as my starting point, and critically address some key reasons for abandoning it in favour of one of its opponents.

Perhaps the single most prominent objection levelled at doxastic views of trust is that it mischaracterises trust's relationship with reasons and cognitive control. Trust, at least in certain circumstances, can be directly willed for practical or moral reasons. The same, as we all know, is not so for belief. Trouble then abounds for any view that insists upon a trusting belief. The most

widely discussed example purporting to demonstrate trust's non-doxastic relationship with reasons and control is from Richard Holton (1994: 63):

> **SHOPKEEPER** Maureen runs a small shop. She discovers that one of her employees, Matthew, has been convicted of petty theft. She decides, nevertheless, to trust him with the till. Doing so, she hopes, is the best way to draw him back into the moral community and promote trustworthiness in him.

The thought is that if such instances exemplify real, bona fide trust, of the kind that we are trying to build our theories around, then trust cannot be like belief. At the very least, its rational profile is deeply divergent from that of belief, and our theories of trust's nature should reflect this point. For the key point straight from the objector's mouth, here's Faulkner on the main lesson of such cases:

> "She can choose to accept her new employee's statement that he will not steal without assessing his testimony as evidence for its truth, and ignoring the evidence she possesses for its falsity... The [trustor] does not need evidence to support this presumption - she need not have assessed his character, for instance - *she simply chooses to trust.*" (Faulkner 2007a: 885-886, emphasis added)

Given doxastic involuntarism, one cannot simply choose to believe. And yet, if such cases – which have come to be known as 'therapeutic trust' – are taken at face value, one can, in the right circumstances, simply choose to trust.

This literature has now reached an impasse. Objectors such as Holton (1994); Faulkner (2007a; 2011; 2025); McGeer (2008); Frost-Arnold (2014); and Carter (2024b) are convinced that this is real, full-fledged trust, of the ilk that our theories ought to capture. Proponents of doxastic, or doxastic-adjacent, views (e.g., Hieronymi (2006); McMyler (2011); Keren (2014; 2020)) do not share the objectors' intuition, and tend to be more inclined to categorise such instances as something less than a full-fledged instance of trust *proper*. Indeed, as Faulkner himself puts it, "… the phenomenology is not decisive." (2025: 87). And what is worse, the debate risks becoming entrenched: objectors are highly motivated to treat such cases as central and illuminating; defenders as anything but. But without recruiting heavyweight theoretical commitments, or retreating to and reaffirming one's sharply contested intuitions, a route beyond this impasse is not obviously forthcoming.

This chapter attempts to chart a path beyond. In order to do so I take as my starting point a datapoint over which debating parties can agree: trust is deeply normative; we trust for reasons, we revise trust in light of evidence, we hold one another accountable when trust is misplaced or illegitimately withheld. So far, so uncontroversial. But what has gone largely unnoticed is that such platitudes about the normativity of trust create space for a range of constraints on reasons-responsive and norm-governed attitudes. If, as we all agree, trust is one such attitude, then a range of theory-neutral constraints come to the fore. These constraints, I argue, can be deployed as impartial adjudicative resources with which to assess each of our competing views.

Using such resources, I aim in this chapter to make two main contributions: reinforce the defence of the belief-like view from one of its main objections; and, from there, develop our positive characterisation of the view. On the defensive side, I firstly deploy the aforementioned theory-neutral constraints on reasons-responsive and norm-governed attitudes to argue that, ultimately, the view that omits therapeutic trust from its analysis offers the more attractive account of the rationality of trust. Therapeutic and non-therapeutic instances of trust vary in ways that I argue would be surprising if they were, as the action-like view maintains, instances of the same type of attitude.

Then, from this vantage point, I seek to develop the belief-like view in two main ways. Belief-like views have been thought unable to account for trust's characteristic agency – the sense that we are not merely passive in whom we trust. However, I demonstrate that belief-like views can perfectly well accommodate the thought that trust is under our rational control; but the relevant conception of control, I argue, is *attitudinal*, not *volitional*. We can deliberate over trust, we can revise trust in light of reasons, reasons we take to bear on the attitude, we can trust differently over time. All of these aspects demonstrate our attitudinal control, yet none of it involves direct volition. The mistake in the discussion surrounding therapeutic trust was that it encouraged a false dichotomy: either trust is action-like and voluntary, and thus deeply agential; or it is belief-like, involuntary, and thus largely passive. But the literature on doxastic (in)voluntarism and attitudinal control shows us that voluntariness and rational control are not equivalent. There is a massive agentive space lying between rational passivity and voluntary control. It is in this space, I argue, that our conception of trust's agentive character should lie.

Lastly, I seek to demonstrate that the belief-like view can offer something of an error theory: whilst denying the verdict that the objector draws from Holton's case, the belief-like view can nevertheless explain why such cases have exerted such pressure on the view. The key in offering such an explanation, centres the long tradition of analysing and individuating attitudes according to their *teleological* or *functional* structure. In short, therapeutic instances *aim at*, or *function to produce*, non-therapeutic variants. That is their normative structure. Therapeutic trust *qua* attitudinal category is derived from, normatively governed by, and, when all goes well, can entirely give way to non-therapeutic trust. In this way, I aim to use resources deployed in response to the objection to enhance the belief-like view's positive picture of trust's relationship to reasons and rational control. Trust and therapeutic trust are meaningfully distinct, it is thus no objection that our view rules out or relegates such cases to a distinct attitudinal category. But that is not the end of the story: as we see, the belief-like view can rule out such cases whilst offering (i) a comprehensive account of trust's agency; and (ii) an explanation of why the counterexample felt compelling. In this way, the belief-like view can capture what was enticing about its opponent's view without paying its theoretical costs.

Chapter V. Can the doxastic view account for the value of trust?

Having addressed, in Chapter IV, the question of trust's relationship to reasons and control, and the dialectical stalemate that emerged over the tricky case of therapeutic trust, I turn my attention

to a more recent and lesser-known objection to the doxastic view. To bring it into light, note that the philosophical study of trust can be divided into two distinct questions: a *metaphysical* or *ontological* one that concerns its nature; and an *axiological* one that concerns its value. Most of the ink that has been spilt in this domain is over the first question (or set of questions); theorists have, for the most part, been united in their silence over questions concerning trust's axiology.

But in a recent article, Keren (2020) draws to our attention the importance of the axiological question. Trust is commonly thought of not merely as some handy commodity – like a smartwatch or a hot water bottle – but rather as an attitude of *distinct* and *indispensable* value. As Locke put it, trust is "… the bond of society" (1663; echoed in Hollis 1998: 4). Bok similarly writes: "Whatever matters to human beings, trust is the atmosphere in which it thrives." (Bok 1978: 31) But if this is right, then philosophical theories of trust can be compared and evaluated not only by reference to questions of their nature – such as whether they adequately distinguish trust from nearby attitudes on metaphysical grounds – but also by reference to questions of their value – such as whether they do so on axiological grounds. If trust is distinctly and indispensably valuable, then our accounts can be evaluated according to the conception that they provide of this value. This challenge, of accounting for trust's indispensable value, Keren labels *Trust's Meno Problem* (TMP, for short), owing to its obvious connection with Plato's Meno problem regarding the value of knowledge.

In this chapter I take up TMP on behalf of the doxastic view of trust. No sooner does Keren raise the problem than does he suggest that it will have particular application for proponents of the doxastic view of trust.[3] The reason, in short, is that if trust is just a species of belief, then it is hard to see why it could be of distinct and indispensable value. In particular, given that it seems at least on the face of it as though one could simply extract all its value by simply *acting as if* one has a trusting belief, i.e., no added value seems to be added by one's *actually having* the belief in question.

In this chapter I make three main contributions: (i) I firstly contextualise and elucidate TMP, to which Keren gestures but does not address in detail; (ii) then I draw on the epistemological literature on the aforementioned Meno problem for knowledge to disambiguate among multiple possible interpretations of TMP; (iii) finally, having found the most philosophically challenging interpretation of TMP, I demonstrate that the doxastic view has the resources to address it. In particular, I draw on the highly plausible claim, found in Katherine Hawley's (2012; 2019) influential work on the topic: that trust *pays a compliment* to the trustee. By taking this idea seriously and folding it into one's doxastic view, we can adequately capture the difference between cases in which one has the relevant trusting belief, and those in which one merely acts as if this is so. The payoffs of exploring the doxastic view in the context of Trust's Meno Problem are twofold: we better understand the nature of the problem itself, and we see that the doxastic view can give a satisfying answer.

Chapter VI. The generality problem and the c-typing solution

[3] Of which, I should note, he is one.

In the last chapter I address a hugely influential objection to teleological views in epistemology, owing originally to Conee & Feldman (1998), *the generality problem* (TGP, for short). For sake of clarity and concision, I discuss it in relation to reliabilist views in particular, though its application to other teleological views is easy to see. The problem, roughly, is that reliability is the crucial normative property on reliabilism, but it applies only to *types* of processes, not *tokens*. The onus is thus on the reliabilist to give principled and non-ad hoc specifications of the relevant *types* of process in determining whether any token output is epistemically justified. And herein the problem lies. For any given token process T, if one specification renders it justified and another does not, then what determines which of these specifications is the correct one?

This problem, in various versions, has been a thorn in the side of reliabilists for decades. In this chapter I consider and, ultimately, reject a recent attempt to solve it. Titus[4] (2017) proposes what she calls the 'C-Typing Solution'. On Titus' proposal, *credences* play the crucial role in determining and specifying process types. A myriad of factors – environmental, psychological and otherwise – contribute to the formation of a belief. But for any belief, we can always ask which features the subject herself was sensitive to in assigning whatever credence they have in the relevant proposition. To this question there will always be a fact of the matter: for any belief, there will always be a particular set of features that were causally relevant in the formation or retention of the subject's credence.

According to Titus, C-Types suggest a workable solution to TGP. Here is the proposal:

> **C-Type:** X is the C-Type for the token way Y that a subject S forms or retains a belief b(p) in p iff X is the type that is individuated by exactly the features of Y causally relevant to the formation or retention of S's credence c(p)
>
> **C-Typing Thesis:** For any belief b(p) held by any epistemic agent A, the way Y in which b(p) is formed or maintained is a token of a unique C-Type X, as defined above. X is the type of belief-forming process that determines Y's reliability relevant to epistemic justification. (Titus 2017: 112)

Key to the success of Titus' proposal, I argue, is what I call the 'repeatability claim' – the idea that, as Titus herself puts it, "… although we often form degrees of confidence in claims about particulars on the basis of token interactions with those particulars, what credence we form is determined by repeatable features of belief-forming processes that are instantiated on that occasion." (Titus 2017: 112) She continues: "Even where particular interactions are crucial to credence-formation, it is highly plausible that we systematically form credences on the basis of general features of such interactions, features we take to indicate how likely a belief is to be true." (Titus 2017: 112) Credences hold the key to a solution to TGP, then, insofar as they pick out, for every token belief, a non-arbitrary process types, which can in turn be assessed for reliability.

In this chapter I target Titus' repeatability claim. Credences, I argue, are often formed in decidedly *un*systematic ways, in response to unique features of each particular situation and interaction. I

[4] Lisa Titus went by 'Lisa Miracchi' at the time of publication of the (2017) paper that forms the target of Chapter VI. I have updated all references to reflect her name change.

firstly present a challenge from testimony. Sometimes, in testimonial exchanges, we are simply overwhelmed with a gut feeling that something is awry; so much so that, as a matter of fact, the credence we place in the proffered proposition plummets. Irrespective of whether this is a good or a bad thing, epistemically, the claim is simply that – as a matter of fact – it is something that we often do. But such gut feelings, I submit, do not arise in response to general, repeatable features of our interactions; to the contrary, such feelings seem to arise quite at random. If this is right, and our credences are oftentimes fixed by gut feelings, which, in turn, are oftentimes fixed at random, then the repeatability claim looks to be in trouble. I then develop the challenge by considering cases involving not gut feelings but rather implicit bias. Ultimately, the C-Typer faces a dilemma: specify C-Types to *include* factors such as gut feelings and implicit bias or specify them instead to *exclude* them. Opting for the former, I argue, risks rendering the thesis insufficiently *reliabilist* to count as a solution, while opting for the latter risks rendering the solution ad hoc. Either way, Titus' C-Typing thesis comes up short.

Chapter I.

Testimonial liberalism and the balance of epistemic goals[5]

Abstract: There are two broad views in the epistemology of testimony, conservatism and liberalism. The two views disagree over a particular necessary condition on testimonial justification: the positive reasons requirement (PRR). Perhaps the most prominent objection levelled at liberalism from the conservative camp stems from *gullibility*; without PRR, the thought goes, an objectionable form of gullibility looms large for liberals. In this paper I aim to make two main contributions: to introduce a new metric for adjudicating this debate; and to argue that, from the perspective of this new metric, the liberal view is stronger than has been appreciated. Drawing on work from James (1896), Goldman (1986), and Kelp et al. (2020), I firstly countenance the distinction between *positive* and *negative* epistemic measures. Positive measures concern, roughly, the acquisition of truths, whereas negative measures concern the avoidance of falsehoods. Both, it is argued, are relevant to epistemic justification, but this debate has proceeded in such a way as to overemphasise the importance of the latter over the former. Once this distinction is made, new conceptual terrain opens for the liberal. Rather than being resigned to a predominantly defensive role - of mitigating worries about negative measures - the liberal can go on the offensive, and explore the independent epistemic strengths of their position. The upshot is that liberals have a new way to dispel their most prominent objection.

§1 Introduction

Consider a foolhardy epistemic agent, Fred, who believes everything he's told, no matter the content, source, or circumstances. You testify to something, he'll believe it. A very natural critique of Fred's testimonial belief-forming policy is that it is simply too risky to be justification-conferring. In other words, the fact that his chosen strategy involves no safeguards whatsoever against the non-negligible risk of false testimony means that he does not count as justified in believing when he does. Fred's is a clear-cut case of gullibility.

This thought is a straightforward and intuitive one, yet two details are worth noting. The first is that the style of epistemic critique at play here is *policy*-based.[6] Fred's beliefs aren't just unjustified,

[5] This chapter is an adapted version of Patrizio (2024a).

[6] This 'policy' language enters the epistemology of testimony with Elizabeth Fricker's seminal work on the topic (1994; 1995). I return to Fricker's work in detail in §2.

they're unjustified *because produced by a bad policy*. The critique thus instantiates a very plausible idea, namely that the justificatory status of one's belief bears some overlap with the epistemic quality of the policy that produced it.[7] Moreover, the precise worry with Fred's doxastic policy is not just that it is bad, it is bad *because it is objectionably gullible*. The policy he adopts in response to testimony exemplifies gullibility, and gullibility is no route to justified belief.

I home in on this policy-based critique because it has played a prominent role in the epistemology of testimony - in particular, in the well-known debate between liberals and conservatives about testimonial justification. Here's the liberal view:

> **Testimonial Liberalism** A hearer has testimonial justification for believing that p on the basis of some speaker's say-so just so long as they lack any undefeated defeaters for the proffered testimony.[8]

According to liberalism, a 'no-defeater' clause is all that's required: it is perfectly permissible to go ahead and believe just so long as one lacks defeaters. Conservatives disagree. They object that without independent support - i.e. non-testimonial, positive reasons - for thinking that one's interlocutor is trustworthy, liberalism is vulnerable to the style of critique just outlined. While perhaps not as egregious as Fred's, the liberal's policy is similarly criticisable: ultimately, it fails to confer justification because it simply lacks the necessary safeguards against the risk of false testimony. Such a policy is, as Fricker puts it, "…an epistemic charter for the gullible and undiscriminating." (1994, 126)[9]

Conservatives advance a competing policy:

> **Testimonial Conservatism** A hearer has testimonial justification for believing that p on the basis of some speaker's say-so just so long as they are in possession of sufficient independent reasons for thinking that the proffered testimony is likely to be true, where such reasons are not themselves defeated by any undefeated defeaters.[10]

An absence of defeaters won't suffice on conservatism; one must also be in possession of positive reasons for justified testimonial belief. This *positive reasons requirement* (PRR) thus divides the views and, at least according to conservatives, allows them to avoid the gullibility worries that plague more lenient approaches.

Fred's foolhardy policy was certainly too gullible to be a source of justification, but are conservatives right in thinking that the same goes for liberalism? The issue can be understood in terms of reliability: the conservative worries that without PRR liberalism is too unreliable to sanction non-gullible belief. In this article I defend liberalism from this charge with two broad

[7] For the same idea put to work elsewhere in epistemology, see Douven (2008) and Kelp (2014).
[8] The view originates with Reid (1764) and is endorsed by Burge (1993); Coady (1994); Goldberg & Henderson (2006); Graham (2006, 2010); Simion and Kelp (2020). The views have operated under various labels; I borrow the 'liberal' and 'conservative' labels from Jope (2021)
[9] See also Fricker (1995: 404) and Lackey (2008: 170, fn. 32).
[10] The view originates with Hume (1740) and is endorsed by Fricker (1994, 1995, 2006a, 2006b); Adler (1994); Malmgren (2006); Kenyon (2013).

theses: the first concerns the *metric*, or standard of evaluation, at play in the debate; the second concerns liberalism's reliability-maintaining resources. I firstly argue that the metric employed in the debate up until now has, by anyone's standards, overemphasised the importance of reliability. What matters is not only reliability - its ratio of true-to-false beliefs - but also *productivity* - the total number of truths produced. I thus defend a metric which evaluates policies by reference to their overall balance between reliability and productivity.

Secondly, I argue that, from the perspective of this metric, the gullibility worry can be dispelled: even if somewhat less reliable, the liberal policy strikes a good overall balance. Firstly, and unsurprisingly, I show that liberalism scores well on productivity. The main contribution of this second section, however, is to argue that liberalism's productivity advantage is not outweighed by a reliability deficit. In particular, if we understand the no-defeater clause as incorporating the notion of *normative defeat*, we see that liberalism is in fact more reliable than has thus far been appreciated. With my two theses taken together, we see not only that reliability matters less than we thought it did, but also that liberals can say more about it than we thought they could. Therefore, if you accept my new metric, and you accept the notion of normative defeat, the gullibility problem for liberalism is dispelled.

Here, in more detail, is how I proceed. In §2 I briefly canvass the way in which liberals and conservatives have approached this debate until now. I question whether liberals should accept the current terms of the debate - whether by doing so they are not playing into the hands of the conservative. In particular, I draw the distinction between positive epistemic measures, which aim at the acquisition of truths, and negative epistemic measures, which aim at the limiting of falsehoods. I argue that whilst both are relevant to epistemic justification, the debate has played out in such a way as to overemphasise the importance of the latter over the former.

In §3 I advance and defend an alternative framework, the *balance metric,* which evaluates doxastic policies by reference to the overall balance they strike between competing epistemic goals. In a nutshell, whether or not some doxastic policy is a prima facie source of justification is not only a matter of its *reliability* (truth-to-false ratio), but also its *productivity* (total number of truths produced). This metric, though friendly to the liberal, is based on some well-established, theory-neutral insights from outside of the current debate.

Having motivated a move to a new metric, I then assess the results for the liberal. In §4 I discuss liberalism's performance on productivity, demonstrating that on this score it fares significantly better than its opponent. In §5 I discuss its performance on reliability, showing here that it fares better than has hitherto been appreciated. In particular, I draw on the notion of normative defeat to strengthen an already powerful reliability-maintaining resource: the no-defeater clause. If we accept this development of the clause, we see that liberalism's impressive score on productivity need not come at an excessive cost to reliability. In §6 I conclude.

The main payoffs of employing the balance metric in this context are as follows: we have a new and independently plausible metric for this debate and at the same time we see that, once it is employed, liberals have a new way to dispel their most prominent objection.

§2 Pinpointing the debate

Consider the following case to demonstrate the gullibility problem for liberalism:

> **ANONYMOUS BLOGGER** One day, at the beginning of the March 2020 global coronavirus pandemic, Tracy is mindlessly surfing the web when she comes across a post from an anonymous blogger. The post claims that the global coronavirus pandemic is in fact a cleverly-designed hoax to enable governments to install millions of 5G stations during lockdowns. Tracy adopts the corresponding belief on the basis of the anonymous blogger. Crucially, as of this moment, Tracy has no other evidence one way or the other as regards the proper cause and explanation of the pandemic.

Tracy behaves gullibly in believing on the basis of an anonymous blogger, yet she lacks defeaters for said belief. The epistemic problem in such a case is thus not explained by defeaters, for there are none; the problem rather has to do with a lack of independent support.[11] If this is right, then an absence of defeaters is compatible with the agent's nevertheless believing gullibly, and liberalism looks to be in trouble.[12] This is the gullibility problem (GP).[13]

Now, evaluating competing belief-forming policies is a complex business. And there are two distinct questions that are relevant here. On the one hand, there is the question of whether liberalism really does license a justification-undermining form of gullibility.[14] But there is a more fundamental question, often-overlooked, which needs to be settled before we can answer the first, namely the question as to the *metric*. We need to agree upon the standards of evaluation here before we quarrel over the verdict delivered by these standards.

§2.1 Clarifying the metric

In what sense is Tracy gullible? And, importantly, in what sense would she avoid this charge if she adopted conservatism? Answering these questions will help us identify the metric at play in the debate.

The work of Elizabeth Fricker, the most prominent contemporary proponent of the GP, will be instructive here. Fricker (2006a) characterises the objection as follows: a hearer behaves gullibly when she "adopts a policy for doxastic response to testimony which fails to screen out false testimony". This is the case, moreover, for any agent who forms a belief in "…any environment in which false testimony is likely enough for that risk to be epistemically relevant." (Fricker 2006a, 620)

Here's the thought: given the method (testimony), and the environment (one in which testimony is often false), the liberal policy (which in this environment licenses testimonial belief even in the

[11] In what follows I grant the conservative that, at least *prima facie*, there are no defeaters in such a case, though I return to this case in §5.2. For a liberal defence that makes no such initial concession, see Perrine (2014).
[12] Fricker (1994; 1995; 2006a) most notably presses this worry, though Lackey (2008) also presses a version of it in defending her hybrid alternative to both conservatism and liberalism. I set aside hybrid views such as Lackey's (2008), Faulkner's (2011) and Greco's (2020) in what follows.
[13] The gullibility problem has elsewhere been referred to as 'the subjective source problem' (Simion and Kelp 2020).
[14] See Fricker (1994; 2006a) and Goldberg and Henderson (2006) for debate over this question.

absence of positive reasons) is a sanction for an epistemically objectionable (and thus justification-undermining) form of gullibility.[15]

The above is instructive when it comes to clarifying the metric. Here's what we've learned. Just like our critique of Fred at the outset, the method of evaluation at play is *policy*-based. The idea, roughly, is that prima facie testimonial justification corresponds to the doxastic policy the agent adopts in forming their belief. One's belief may be unjustified by virtue of its being the result of a bad policy. The second point worth noting is that the metric is *reliability*-based. The objection to the policy is not just that it licenses gullibility in some unspecified sense. It is gullible because (at least in that environment) it is unreliable.[16] This much is clear from the Fricker quotes above. The liberal policy would not do a good enough job of screening out false testimony to be justification-conferring. One should instead adopt the conservative's policy, which presumably would.

The metric in place for evaluating competing policies is thus what I'll call the *reliability metric* (RM). On this metric, the prima facie justification of a belief is a function of the reliability of the policy that produced it. In what follows I take no issue with the metric's being *policy*-based, and instead focus my attention on its being *reliability*-based. I'll grant the overlap between doxastic policy and epistemic justification, in order to properly target the idea that reliability is the appropriate standard of evaluation.

To be sure, it is very plausible that whether or not a policy is a source of prima facie justification is (in part) a function of its reliability. A policy with no negative measures - e.g. Fred's foolhardy policy of 'believe everything you're told' - is not a plausible candidate source of prima facie justification. And, indeed, it is plausible that the explanation for this has to do with its being a license for gullibility. The RM is thus, at least at first glance, a sensible way to settle the debate.

It is worth noting that in their responses to GP, many liberals show that they, too, endorse something like RM, though they of course disagree on the verdict it delivers. Take Goldberg and Henderson (2006), Graham (2010; 2012; 2018a), and Simion and Kelp (2020), for example. All three cases exemplify a common strategy: the liberal defence is mounted by appeal to various theoretical mechanisms which serve to boost liberalism's reliability score - its negative measures. More specifically, Goldberg and Henderson (2006) appeal to the existence of an innate *monitoring,* capacity in hearers, which sifts through all consumed testimony and alerts the agent whenever there are clear signs of deception, incoherence, and so on. Graham (2010; 2012; 2018a), across a series of papers, defends liberalism by appeal both to a *filtering* capacity in hearers and the existence of *social norms* influencing speakers. Filtering, like monitoring, increases the reliability of testimony by flagging to hearers when there are signs of deception. Moreover, it also incentivises truth-telling in the first place, for lies will often get caught. (2010: 173) Alongside filtering, Graham also appeals to the existence of social norms in making the case for default acceptance. Even if is in their best interests to lie, internalised social norms - including the norm on truth-telling informatively -

[15] In this paper I set aside Fricker's empirical premise that false testimony is sufficiently prevalent in the real world to undermine liberalism in this way. For critical discussion on this premise, see Michaelian (2010), Shieber (2012), and Graham (forthcoming).

[16] I am thinking of reliability here in Goldman's (1986) sense, as denoting the ratio of true-to-false beliefs.

explain why speakers often nevertheless tell the truth.[17] Simion & Kelp (2020), for their part, similarly appeal to the existence of such social norms in explaining the reliability of testimony, and thus in grounding the view that default acceptance, even absent positive reasons, is a prima facie source of justification. Thus, while they differ in the particular features of testimonial exchanges they emphasise, all aforementioned liberals appeal to theoretical mechanisms which - granting that they exist and are effective - would serve to boost the reliability of believing on the basis of mere say-so in the absence of positive reasons.

It is beyond the scope of the current paper to evaluate whether liberals are successful in this endeavour. Instead, I want to pose the question as to whether liberals, simply by accepting the terms of the game, are not playing into the conservative's hand.

§3 Changing the metric

§3.1 Sceptical policy

Consider a third option as a testimonial belief-forming policy:

> **Sceptical Policy (SP)** A hearer is entitled to believe on the basis of some speaker's say-so just so long as (i) they lack any undefeated defeaters, (ii) they are in possession of positive reasons, *and* (iii) they have established that there is no chance whatsoever that their interlocutor is lying or mistaken.

As a conservative, why not go a step further and endorse SP? Here's why: while it is of course true that one's policy must encode certain precautions against the risk of falsity, there is a limit. Even supposing we granted that the liberal policy amounts to gullibility, it is doubtful that many would deem SP an adequate replacement. If liberalism can be accused of an epistemically objectionable form of gullibility, SP would certainly be accused of an equally epistemically objectionable form of cynicism. The upshot of this is that a *pure* RM, which tests only for ratio-maintenance and delivers the straightforward verdict that liberalism is objectionably gullible because less reliable, is inadequate, even by conservative standards.

§3.2 James, Goldman, and Kelp et al.

This thought will be recognisable to most as a manifestation of the familiar Jamesian (1896) insight that the epistemic goal of acquiring truths (the truth goal (TG)) is in fact ambiguous between two distinct subgoals: the avoidance of falsehoods (AF goal); and the obtaining of truths[18] (OT goal) (Alston 1985; Foley 1987; David 2001; Riggs 2003; Carter et al. 2016).

For our purposes it is worth emphasising that these two goals are *competing*. In other words, they can pull in opposite directions; they are not two ways of stating an identical commandment (James 1896, 203-4). As Carter et al put it:

[17] The thought that there are social norms increasing testimony's reliability is popular one. Examples, which come from each side of the divide, include: Fricker (2006b, 2017); Graham (2010; 2012; 2018a; 2019); Faulkner (2011).
[18] James sometimes spoke in terms of truth, and sometimes in terms of knowledge. For the sake of simplicity, I will continue with truth, though nothing substantive hinges on this decision for my argument.

> Weighting the [first] aim more would lead one to be more cautious in order to avoid possible misrepresentation. Weighting the [second] aim more would lead one to be bolder in order to possess more truths. (2016, 2336)

Goldman makes a similar point in the following passage:

> Complete reliability can be achieved by extreme caution, or conservatism: producing beliefs only in the 'safest' circumstances, where it is virtually impossible to go wrong … But such radical caution would probably be purchased at the price of (extensive) ignorance. If hardly any beliefs are produced, hardly any true beliefs are produced. This spells (extensive) ignorance. (1986, 26)

If overly concerned with the OT goal, one may be pushed toward carelessness; but equally, if overly concerned with AF goal, one would be pushed toward cynicism. Neither represents a good policy.[19] To reflect this insight, Goldman highlights the importance not only of reliability but also of what he calls 'power'.[20] He continues:

> If the antidote to error is reliability, the antidote to ignorance is (intellectual) *power*. Power is the capacity of a process, method, system, or what have you to produce a large number of true beliefs; or, slightly differently, the capacity to produce true beliefs in answer to a high ratio of questions one wants to answer or problems one wants to solve. Although power, like reliability, incorporates the ingredient of true belief, it is clearly a distinct standard. A method or system can be very reliable without being very powerful; and a method or system can be pretty powerful but not terribly reliable. (1986, 27)

We can find similar insights reflected in Kelp et al. (2020), who distinguish between different standards of appraisal we might employ when assessing epistemic performances. Adopting a Sosa-style telic normativity framework, Kelp et al. consider an idealised model of performing (epistemic) agents, and ask: what makes one epistemic performance better than another? They note that, alongside reliability, performing agents can also be more or less *productive*. (2020, 5196) In other words, what matters is not merely the proportion of successful-to-unsuccessful performances, but also the total number of successful performances produced.[21] As Kelp et al. put it, "All else equal (including the degree of reliability), agents are better at producing performances of a certain type if they are more productive." (2020, 5196)

---

[19] This thought is closely related to the debate between White (2005) and Kelly (2013) over doxastic 'uniqueness' and 'permissivism'. The former is the thesis that if an agent whose total evidence is E is fully rational in taking doxastic attitude D to P, then any subject with E who adopts an attitude other than D to P is less than fully rational. The latter, on the contrary, is the thesis that different agents with the same evidence might well be fully rational in adopting different doxastic attitudes to P. The overlap here is complicated, but fortunately I do not need to take a stand on that debate one way or another. One may read my argument as conditional: if uniqueness is true, then my argument is for the conclusion that liberalism is the superior policy; if permissivism is true, then my argument is for the conclusion that it is no worse than conservatism on account of being less cautious. Either way, the force of the GP is dispelled. Thanks to Chris Willard-Kyle for helpful discussion on this point.

[20] Alongside reliability and power, Goldman also discusses 'speed', though I set it aside for present purposes.

[21] Productivity, at least when the performance at issue is an epistemic one, is thus very similar - if not identical - to what Goldman called 'power'.

Consider the following example to make these points salient.

> **RELIABLE RANDY AND PRODUCTIVE PETE** In forming his beliefs Randy employs belief-forming policy A, which produces *only* true beliefs but does so very slowly. Pete, on the other hand, employs policy B, which is a little more fast and loose; it produces only *mostly* true beliefs, but produces beliefs quicker than A. As a result, Randy has 1 belief in total, and it is true. Pete, on the other hand, has 100 beliefs, 99 of which are true, 1 of which is false.

Here's the thought: the correct verdict is that Pete's policy is the better one, epistemically speaking. But this verdict is not delivered by RM, which tests only for ratio-maintenance. Strictly speaking, Randy's is more reliable. But, of course, it is epistemically worse; an epistemic agent ought to adopt Pete's policy rather than Randy's in spite of the reliability deficit.

Bringing this together, we can draw a distinction between *positive* epistemic measures, which concern the acquisition of truth, and *negative* epistemic measures, which concern the avoidance of falsehood. In what follows, I'll simply speak of a policy's 'productivity' when discussing its *positive* evaluation (assessment of its truth-production), and 'reliability' when discussing its *negative* evaluation (assessment of its truth-to-falsity ratio). All of the accounts just discussed point to a compelling idea: insofar as we are in the business of epistemically evaluating competing doxastic policies, we had better look not only to reliability, nor only to productivity, but rather to some balance of the two.

This, I think, carves out some untapped territory for the liberal. They may grant that there is important overlap between epistemic justification and one's doxastic policy, and yet deny that RM is how policies are to be evaluated. The conservative line in response to a sceptic ought to be that while their policy is indeed less reliable than the sceptic's, it is not *gullibly* so. But, the liberal may point out, likewise for liberalism in comparison to conservatism. Heightened reliability alone does not deliver heightened epistemic justification.

Let's specify a metric that puts these insights to use:

> **Balance Metric (BM)** Some doxastic policy X is epistemically superior to doxastic policy Y if and only if X enjoys a better overall balance with respect to reliability and productivity, where reliability denotes a policy's true-to-false belief ratio, and productivity denotes the number of true beliefs a policy produces.[22]

BM simply makes explicit the intuitive distinction between reliability and productivity. All-else-equal, (including degree of productivity) the more *reliable* the better; similarly, all-else-equal, (including degree of reliability) the more *productive* the better (Kelp et al 2020, 5196). Significant increases in either property might outweigh slight deficits in the other; some policy might properly

[22] Strictly speaking, for reasons pertaining to the well-known problem of trivial truths, it cannot be brute number of true beliefs. I take it, however, that for present purposes this formulation gives us a sufficiently tight grip on a metric that would better capture what we're looking for. For discussion of trivial truths, see Zagzebski (2003), and Grimm (2008).

be considered epistemically superior to another, even if less reliable, so long as it is significantly more productive, and vice versa.[23]

Perhaps none of this would apply if we were limitless epistemic agents - free from our cognitive and temporal constraints. In such circumstances, plausibly, reliability really would be our sole concern; we should do what we can - no matter the cost to productivity - to form reliable testimonial beliefs. But we are not limitless in this way: our minds are limited and so is our time. Alone, there is only so much we can perceive, there are only so many phenomena into which we can inquire, and only so many conclusions we can infer or deduce. My proposal is that we therefore need a metric which responds to our epistemic situation by placing explicit emphasis on the balance between reliability and productivity (hereafter, just *balance*).[24] BM does just this.

§4 Liberalism and productivity

I've argued for a new metric, BM, for settling the debate between liberals and conservatives over reliability. The metric we had been employing until now overemphasised the importance of reliability, and in so doing played into the hands of the conservative. Liberalism's reliability worries are downplayed if we accept BM; reliability is no longer the sole issue determining whether a policy is justification-conferring, because productivity also matters. The next stage of my argument is to assess liberalism's performance on BM. I assess its productivity in this section (§4), and its reliability in the next (§5).

Recall that satisfaction of PRR requires independent reasons on the part of the hearer which somehow speak in favour of the credibility of their interlocutor.[25] The problem, however, is that we lack precisely these kinds of reasons in many ordinary, everyday cases. This, I will argue, leads to a significant productivity advantage of omitting PRR.

Consider the following examples.

> **BAR** Sophie is running late to meet a friend at a bar. When she arrives, she asks the first person she sees whether someone fitting her friend's description has arrived at the bar. They say they have not, and so Sophie forms the (true) belief that her friend has not yet arrived.

[23] It is worth noting that, in emphasising the importance of balancing both reliability and productivity, one is not thereby obliged to provide a neat formula for doing so. As Goldman points out, we should take comfort here from ethics, where noting that there are competing values to be balanced (individual rights with social justice, for example) does not mean that there is a neat or precise formula for doing so. (Goldman 1986, 125)

[24] Balance here is not intended to denote *neutrality, or evenness,* with respect to these competing epistemic measures. A more neutral distribution between the two measures does not represent a better balance. When assessing two policies by utilising the current proposal, a reliability advantage might contribute to an overall advantage, provided the productivity deficit - if indeed there is one - is not so severe as to outweigh this advantage. And vice versa. What is ultimately crucial is that both features matter, and interact with one another.

[25] Note that I proceed with a *local*, rather than *global,* interpretation of the conservative's position. The reason for this, in short, is that the shortcomings of the former have been well-rehearsed in the literature, including, notably, by Fricker herself (1994). For further discussion, see Lackey (2007, 146-149); Weiner (2003, 258); Coady (1992, 82).

CHILD A child, Ben, asks his caregiver whether his parents will be home in time for tea, his caregiver replies that they will, and so Ben forms the (true) belief that his parents will indeed be home in time for tea.

Given that these are perfectly ordinary, unremarkable cases, the like of which take place on a daily basis, one would be forgiven for thinking that there's therefore nothing particularly epistemically interesting going on here. Indeed, such cases are just part and parcel of the everyday exchanges we commonly find ourselves engaged in. But their banality is what makes them pertinent. Such cases represent the kinds of instances in which some of liberalism's independent epistemic strengths, *qua* doxastic policy, are on show.

In **BAR** and **CHILD**, true beliefs are available, but positive reasons are not. Sophie and Ben simply lack independent reasons that speak in favour of their interlocutor. Sophie lacks any background information whatsoever, and Ben - being a child - lacks the ability to cognise independent, positive reasons speaking in favour of his interlocutor.[26] Though for slightly different reasons, then, Sophie and Ben are in the same epistemic boat insofar as all they have to go on is the proffered testimony, combined with the absence of any particular reason to think that it is wrong or misguided.

Such cases are worth drawing attention to because they highlight a significant cost of adopting the conservatives' more stringent policy. If the heightened stringency of conservatism meant that they got things exactly right - that is, it ruled out belief whenever one's belief would be false and allowed belief whenever one's belief would be true - then it would be the perfect policy. However, both policies come with costs, and it is worth reflecting on them in order to ascertain which has the overall advantage. The heightened reliability of conservatism doesn't come for free. Cases such as **BAR** and **CHILD** are the problem cases for conservatism. They are, I contend, the quite ordinary everyday contexts in which we lack independent support for true testimony. Given that conservatism disallows any testimonial belief in the absence of positive reasons, it thus follows that there is a class of would-be true beliefs which are placed out of reach by their policy. Insofar as what we care about is, roughly, obtaining truth and avoiding falsity, there is a significant epistemic cost associated with adopting the more stringent policy.[27]

Of course, there is nothing particularly novel or radical about pointing out that positive reasons are oftentimes in scarce supply. Indeed, one of the most common objections to conservatism is

[26] See Graham (2018b) for helpful clarification and extended discussion of the challenge conservatives face from infant/child testimony.

[27] Ahlstrom-Vij (2015; 2016) brings similar considerations to bear on the GP. He distinguishes between multiple possible interpretations of the problem, and argues that all fall short of establishing the claim that there is anything epistemically remiss about deferring to testimony in the absence of positive reasons. He concludes that a liberal policy leaves us no worse off, epistemically speaking - and might well leave us better off - than a conservative policy. Though similar in spirit, our approaches have important differences. In particular, there is a way of understanding the GP that is overlooked in Ahlstrom-Vij's disambiguation project, namely that an agent who adopts a liberal policy is gullible precisely because said policy is too unreliable. In other words, we might interpret the conservative as complaining that satisfying PRR, and nothing less, is what ensures a sufficient reliability level to avoid the gullibility charge. Perhaps this reading can be subsumed under one of Ahlstrom-Vij's, but it is not obvious that it can, and it is this reading that my solution speaks to.

precisely that it makes testimonial justification too hard to come by.[28] What liberals have not yet done, however, is attempt to thereby mount a defence of their view by appeal to testimony's productivity.[29] If positive reasons are hard to come by, then believing only in their presence will have a negative impact on one's productivity.

§5 Liberalism and reliability

BM tests for both reliability and productivity. Unsurprisingly, liberalism compares favourably on productivity. But how significant is the reliability deficit thereby incurred? The BM puts the reliability worries into perspective; if we accept this metric, reliability is no longer the sole issue driving policy evaluation. However, any policy needs some reliability-maintaining measures. The final piece of the puzzle is to assess the liberal's.

To better understand this issue, I'd like to introduce the distinction between *prima facie* and *ultima facie* justification. Conservatism and liberalism are views of *prima facie* justification: they recommend competing policies for how one ought, all-else-equal, to form testimonial beliefs. We are all in agreement that some countermeasures are needed to address the risk of false testimony, and the conservative proposal is to implement them at the *prima facie* level, by adopting a more stringent testimonial policy across the board. But another route, available to the liberal, is to implement measures at the *ultima facie* level - to explore what resources we have for placing principled restrictions on the prima facie justification licensed by our policy. The focus here, perhaps unsurprisingly, will be on the liberal's no-defeater clause (NDC), which is their main limiting resource. However, in this section I'll argue that if liberals countenance the notion of normative defeat - as they have good reason to do - then it is more powerful than has been fully appreciated.

§5.1 No-defeater clause

The NDC blocks a hearer's justification to believe whenever the hearer is in possession of any particular reason *not* to believe their speaker. But defeaters come in different shapes and sizes; indeed, there is a well-known distinction in the literature between what are known as *rebutting* and *undercutting* defeaters (Pollock 1986). One has a rebutting defeater when one acquires some reason to think that their belief is in fact false, whereas one has an *undercutting* defeater when one acquires some reason to doubt that, in the circumstances in which it is formed, their source of belief is adequate (Pollock 1986: 484; Kelp 2023: 7). The NDC is a powerful liberal resource insofar as a

[28] See Coady (1973: 151-3); Schmitt (1987: 49); Plantinga (1993: 79) for early formulations of the problem as it pertains to conservatism. See Greco (2020: 30) for more recent discussion of the same idea.

[29] A notable exception here is Zollman (2015), who uses a computer-aided thought experiment to demonstrate that a liberal community of epistemic agents would fare better than a conservative one insofar as they would cultivate larger and more heterogeneous pools of interlocutors. In other words, they would score higher on productivity insofar as there would be more - and more diverse - groups of people from which they could solicit testimony. The conservative groups, on the other hand, would remain smaller and more homogeneous, such that though there would be tight-knit groups of highly reliable individuals, there would equally be tight-knit groups of highly unreliable individuals, with little crossover between them. While we are engaged in different projects, then, it is good news for the view I develop here if Zollman turns out to be right. My argument concerns the conditions of testimonial justification, his work addresses what we should do when constrained as to how much testimony we can solicit. As he notes himself, my question is more abstract than the one he discusses, and its connection to action more remote. But if the two answers line up, then all the better for the liberal; such a result would simply be taken as further indication of the underlying plausibility of the view.

great deal of would-be unjustified testimonial beliefs can be handled by the presence either of a rebutting or an undercutting defeater.

Greco's well-known cases will serve us well here as a case-in-point:

> **Case 1.** A suspect in a murder mystery tells the detective that he was not at the scene of the crime.
> **Case 2.** A used car salesman tells a customer that the car has never had any faults.
> **Case 3.** A teacher tells his pupil that France is in Europe.
> **Case 4.** A mother tells her child that there is milk in the fridge. (Greco 2020, 4)

Greco argues that even if positive reasons are not required in cases like 3 and 4, they surely are in cases like 1 and 2. Even if liberalism is broadly plausible, then, cases such as 1 and 2 seem to suggest that the NDC is not always enough - PRR is at least *sometimes* required to avoid gullibility.

However, once we look a little closer at the supposed problem cases, it becomes clear that they can be adequately handled by a properly understood NDC. In short, here's why: cases 1 and 2 very plausibly involve undercutting defeaters. The hearers in such exchanges (the detective and the customer) have significant reason to doubt what they're told (Simion and Kelp 2020, 2862). In particular, the fact that the speaker is a suspect is a strong reason not to simply go ahead and believe what they tell you. Similarly, the fact that the speaker in case 2 sells used cars for a living is a strong reason not to simply go ahead and believe that the car is in mint condition. Of course, one might prefer to characterise these factors as contextual cues that signal a need for positive reasons. But, at the very least, it is equally natural to characterise them as defeaters. They signal a need for defeater-defeaters, and this much can be accommodated by the liberal's NDC, which forbids belief in the presence of undefeated defeaters. Some, e.g. Simion and Kelp (2020), have even argued that a liberal treatment of such cases is in fact *preferable* to a conservative treatment.[30] But for present purposes we need not endorse this stronger conclusion. Suffice it to note that the NDC, properly understood, is all the liberal needs to handle a wide range of cases, including the seemingly problematic ones (e.g., **Case 1.** and **Case 2.** above).

This much has been noted elsewhere, so I will not belabour the point here. Instead, I want to countenance the notion of *normative* defeat, the power of which has yet to be fully demonstrated by liberals in this debate. If this notion exists - and, though not uncontroversial, it is highly plausible that it does[31] - then this would further enhance the power of NDC, and thus further enhance liberalism's credentials with respect to negative epistemic measures.

---

[30] To see why, note that on the most plausible construal of PRR, what's needed is some independent reason for thinking that this particular subject is a generally reliable testifier, perhaps with respect to the particular topic at hand. However, note that in these cases this kind of reason will not suffice. The murder suspect may in fact be an esteemed FBI agent themselves, and be a generally highly reliable testifier (in particular on the topic of murder mysteries). Of course, such a reason won't help in this case, when they themselves are the suspect. What's needed in such a cases is a more particular kind of reasons: one that defeats the undercutting defeater *that they are a suspect in the murder mystery*. But this is just the result that the liberal - and not the conservative - will get.

[31] See Bergmann 2006; Kelp 2023; Simion 2023; 2024 for the case that it does. See Nottelmann 2021 for the case that it does not.

Consider the following case:

> **SYLLABUS** Regina takes a metaphysics class on Wednesdays. When she wakes and begins planning her day, she infers that, as it is Wednesday, she has a metaphysics class. She has the belief 'I have metaphysics today'. However, she has not checked the course syllabus, and were she to check it, she would see that this week is a reading week and so there's no class scheduled today.

It seems very natural to most to say that Regina ought not to have believed that her class was on, but note that the explanation for why this is so is going to be importantly different than ordinary instances of undercutting or rebutting defeat. In this case, she has a rebutting defeater that renders her belief unjustified (that the course syllabus says otherwise). What's different about this kind of rebutting defeater, however, is that Regina has not psychologically registered it. In order to accommodate the intuition that failing to psychologically register a defeater does not get one off the normative hook, we need to countenance the existence of defeaters *external* to the agent's psychology.[32] Enter the notion of *normative defeat*. One has a normative defeater when there is new information which undermines one's justification for their belief *irrespective of whether or not they register this new information*. Many have argued that cases like SYLLABUS support the existence of such defeaters (Harman 1973; Goldberg 2015a; 2017; Kelp 2023; Simion 2023, 2024).

Further support for the phenomenon of normative defeat can be found in cases of testimonial injustice (Fricker 2007). Testimonial injustice occurs when a speaker is assigned less credibility than she is owed through systematic identity prejudice (against her sex, gender, or ethnicity, for example). To see why such cases can be used to support the existence of normative defeat, consider the following case, developed from Lackey (2018) and discussed in Kelp (2022):

> **SCIENTIST** A sexist male scientist, Pete, owing to his prejudice, simply fails to register the contributions of his female colleagues. The thought that a female scientist could make any serious contribution to science is so far from his worldview that he will zone out when they speak, ignore their emails, and so on. Now suppose that Pete comes to believe some scientific hypothesis SH on the basis of he and his male colleagues' research. However, as is pointed out to him by his female colleague, Sara, SH is demonstrably false. Unsurprisingly, Pete fails to even register Sara's decisive counterevidence, and remains undeterred in his belief that SH is true.[33]

Clearly, Pete's belief that SH is true is unjustified despite the fact that he is not in possession of any reasons to doubt it. If that's right, then we need the notion of normative defeat to explain the situations in which one's justification is undermined by rebutting or undercutting defeaters one has not psychologically registered. (Kelp 2022)

---

[32] For this reason, some have opted for the label 'external defeaters' (e.g. Kelp 2022). There are reasons to prefer this label, however for the sake of simplicity, and consistency with the majority of the literature on the topic up until now, I will proceed with the label 'normative defeaters' in this paper.

[33] See Kornblith (2007, 290-3) for an earlier discussion of a case in which an expert's belief is similarly rendered unjustified by virtue of their having ignored the contributions of their peers.

§5.2 Incorporating normative defeat

I lack the space to further bolster the case for normative defeat. Instead, in what follows, I will simply take it for granted that examples like **SYLLABUS** and **SCIENTIST** suffice to demonstrate that there is strong, independent motivation for countenancing such a notion.[34]

Recall that what is at issue is liberalism's negative epistemic credentials - what resources it has at its disposal for restricting the justification it licenses. NDC is the obvious place to start, and the cases just outlined speak in favour of a refinement of this liberal resource. Understood in this way, the NDC is capable of handling not only instances in which the agent psychologically registers a defeater, but also certain instances in which their justification is undermined despite having failed to do so.

To appreciate the value of incorporating normative defeat, let us return to the **ANONYMOUS BLOGGER** case with our sharpened NDC in hand. The worry was that Tracy lacks a defeater for such a belief but is nevertheless gullible. NDC alone is thus said to be insufficient for avoiding the gullibility charge, and the case for conservatism is strengthened. However, now, having developed our NDC, we see that there is a plausible avenue of response. In particular, it looks as though the case is underspecified, and that, depending on how the missing details are spelled out, the liberal has a solid response.

We firstly need to know more about the agent's testimonial environment. After all, in a world in which blog posts are, as a matter of fact, the most reliable source of news - perhaps because they go through a rigorous peer-review process - and that they are, in fact, always anonymous so as to maintain the integrity of this process, then Tracy would not be gullible in believing on its basis. If, however, much like the real world, the environment was such that forming beliefs about complex, global, socio-political phenomena on the basis of an anonymous blog post on a random website is a wildly unreliable guide to the truth, then the liberal has the resources to explain this away as an instance of defeat. If Tracy has psychologically registered any of this information about the quality of testimony she is likely to receive from an anonymous blog post, then she has reason to doubt her source. The liberal's NDC handles this as an instance of non-normative (psychologically registered) *undercutting* defeat. If instead she has psychologically registered any of the abundance of counterevidence that exists for such conspiracy theories, then she has reason to think her belief is false. The liberal's NDC would thus handle this as an instance of non-normative *rebutting* defeat. Finally, she might have failed to register either but the world she inhabits, much like the real world, is one in which both kinds of reasons are readily available to her at minimal effort. This final interpretation of the case would be handled by the liberal's NDC as an instance of normative defeat. She failed to register the defeating information, but clearly ought to have done so. As we have seen from **SYLLABUS** and **SCIENTIST**, there is good, independent reason to think that negligently failing to acknowledge evidence does not get one off the epistemic hook. Insofar as

[34] These motivations are independent in the sense that they are not theory-laden. Though I will argue that developing NDC to incorporate normative defeat does important work for the liberal, notice that the motivations for doing so more broadly do not come from the liberal camp. It is friendly to the liberal, but this is not the reason the notion was originally deemed necessary. Indeed, Jennifer Lackey, e.g., has defended the notion of normative defeat whilst, in other work, rejecting liberalism precisely because of the GP.

the phenomenon of normative defeat exists, then, I contend that this would constitute an uncontroversial instance.

There is, in fact, one final option. Suppose the details of the case were spelled out such that Tracy lacks any defeater, normative or non-normative; that is, she has not registered a defeater and it is not even the case that she ought to have. Such a world, presumably, would be one in which it is not abundantly clear that anonymous blog posts are a wildly unreliable guide to the truth - perhaps, again, they're actually quite reliable - and there is not, in fact, an abundance of counterevidence for the conspiracy theory proffered. Belief could not be disallowed in such a case because, even on the refined NDC, there is no defeater. However, I submit, this is the right result. Indeed, the subject in such a case is *not* gullible in believing. If the world was such that anonymous blog posts were in fact a decent source of information and there was not, in fact, a whole host of readily available counterevidence to the conspiracy theory, then it is no longer clear that it would be gullible to believe.[35] The case is thus crucially underspecified and, once the relevant details are filled in, the liberal has the resources to get the right result. Either (i) Tracy is indeed gullible but the liberal does not permit belief, or; (ii) the liberal permits belief but Tracy is not gullible.

One may worry just how liberal this new version of liberalism is. In particular, are normative defeaters not simply defeaters we *do not* have but *should have* had - i.e., ones we would have caught had we been more vigilant? And is this heightened vigilance not precisely what the conservatives have been advocating for all along?[36]

In response to this worry, it is worth emphasising that the view we have ended up with is liberal just by virtue of omitting PRR - a speaker need not have positive reasons to have justification. Once committed to the liberal camp in granting *prima facie* justification in the absence of positive reasons, the question that follows concerns what resources the liberal has for restricting *ultima facie* justification. To this end, I've argued that a normative defeat refinement of NDC can do important work for the liberal - it places further restrictions on justification in a way that is consistent with the omission of PRR, thus strengthening the case that liberalism, while it may be less reliable than conservatism, is not thereby a charter for gullibility. If the NDC, properly understood, allows liberals to handle the seemingly problematic cases, then so much the better for the view that omits PRR, and thereby secures the productivity benefits that come with so doing.

There is a trade-off between competing epistemic pressures, and liberalism strikes a different balance than does conservatism, but normative defeat helps us to see that the productivity advantage need not come at an excessive reliability deficit. If the conservative can help themselves to similar such reliability-maintaining resources, then so be it. The point remains: this is a powerful resource that is within the liberal's toolbox.

[35] For an extended treatment of the effect of one's environment on the epistemic permissibility of belief, see Levy (2021).
[36] My thanks to an anonymous reviewer at *Philosophical Studies* for pressing me on this point.

§6 Conclusion

Liberalism and conservatism recommend different doxastic policies when it comes to testimony. These competing policies are divided over the positive reasons requirement, and the conservative insists that its omission sanctions gullibility. In this article I've sought to address this worry for liberalism in two main stages. The first has been to argue that we should accept a *balance metric* in adjudicating this debate, which evaluates policies by reference to the balance between reliability and productivity. This metric should be acceptable to all: it vindicates the conservative tenet that some policies are too unreliable to be justification-conferring, whilst maintaining that others are too sceptical to be any better. Secondly, I argued that liberalism strikes a good overall balance. Unsurprisingly, it enjoys a productivity advantage over its opponent. But more notably, if we understand the no-defeater clause as incorporating *normative* defeat, then this productivity advantage is not outweighed by a reliability deficit. Therefore, if we accept the balance metric, and incorporate the notion of normative defeat, liberalism's most prominent objection is dispelled.[37]

[37] For comments on earlier drafts of this paper, and extremely helpful discussion along the way, I am indebted to Chris Kelp, Christopher Willard-Kyle, Mona Simion, Adam Carter, Katie Moody, Rory Aird, Ísak Ólafsson, Dario Mortini, Matt Vermaire, and two anonymous reviewers at *Philosophical Studies*. Versions of this chapter were presented at Bled Epistemology Conference, the European Epistemology Network, the European Congress for Analytic Philosophy, the Social Epistemology Network Event, Scottish Epistemology Early Career Researchers (SEECRs), University of Glasgow Graduate Seminar, and the COGITO research seminar; I am grateful to audiences on each occasion for questions and discussions that significantly improved the paper.

Chapter II.

On the zetetic value of testimony

Abstract: We epistemically depend on one another in a range of ways beyond the mere exchange of propositional knowledge. It is argued that traditional frameworks in the epistemology of testimony fail to account for this thought, and that a novel, inquiry-centric alternative does better. If we think of testimony not as a source of belief but rather as a resource in inquiry, we begin to do justice to the rich plurality of roles that it plays in our epistemic lives.

§1 Introduction

Consider a simple case of testimony in which a knowledgeable speaker, S, asserts that *p* on the basis of excellent evidence to an appropriately diligent hearer, H. It is uncontroversial that, on the basis of such an exchange, H can come to know that *p*; indeed, all extant views on the epistemic normativity of testimony will secure this result. But many exchanges look importantly different: instead of a new belief, the speaker offers a way of thinking - an analogy, explanation, or conceptual tool that reshapes how the hearer frames the matter at hand.

There is now a burgeoning body of work that draws attention to these more subtle facets of our communication,[38] and yet it is arguable that the epistemology of testimony, owing to its preoccupation with the simple cases, has not fully caught up. For this reason, Hazlett unsympathetically dubs the traditional model of testimony "overly simple." (2025: 13) Fraser, in a similar vein, notes that "…a normative epistemology informed by such a [simple] picture will be, at best, incomplete and at worst, distorted." (2021: 4026) Speakers often do more in testimonial exchanges, the thought goes, than merely induce hearer belief. Normative theories of testimony - theories of its epistemic value and the norms that govern its practice - should be able to account for this fact.

In this article I aim to advance the discussion in two main ways: to clarify the problem and then offer a solution. Recent work gestures toward a problematic oversight in our theorising about testimony but does not make clear what exactly it amounts to. First step, then, is to clarify why exactly extant normative theories of testimony will fail to scale up to more complex exchanges than those they have been built around. Then, with the problem in clearer view, I offer a solution: a new, inquiry-based theory of testimony that better handles the data. If we think of testimony not as a source of belief but rather as a resource in inquiry, we begin to do justice to the rich

[38] See, e.g., Camp (2017a; 2017b); Fraser (2021); Munton (2021); Whiteley (2024); Hazlett (2025); Sliwa (2025); El Shazly (forthcoming).

plurality of roles that it plays in our epistemic lives. The upshot is twofold: we better understand a growing worry in the literature; and see that zetetic epistemology has resources to address it.

§2 Complex testimony

The bulk of the literature in the epistemology of testimony coalesces over the model, or ideal, testimonial exchange, around which theories are built. Schematically: S asserts that *p* (where *p* is a relatively simple proposition[39]) to H, who, on that basis, comes to know that *p*. Accordingly, the normative discussion has centred almost exclusively on the speaker's *assertion* and the hearer's testimonial *belief*. We agree over the model, then take as our key task that of establishing which norms best realise it. Hence, we ask: what is the epistemic norm of assertion? Must a speaker *know*, or merely *reasonably believe*, that which they assert?[40] And on the hearer's side: must a hearer have non-testimonial, positive reasons before believing their interlocutor's assertion? Or, on the other hand, is mere say-so enough?[41] On this simple *assertion-and-belief* model, normative theories of testimony are modelled around its core epistemic capacity: transmitting knowledge of what is said.[42] The problem with this model, however, is that it fails to account for a range of everyday testimonial exchanges. In this section I'll consider three such examples from the recent literature, gathering them under the banner of 'complex testimony.'[43]

§2.1 Explanation

Suppose I ask whether your daughter is an introvert. You reply: *"Well, she steals the show at dinner parties, but those same parties exhaust her."* (Green 2019: 54-5; discussed in Hazlett 2025: 13). What I gain from such an exchange is not merely a pair of true beliefs. Rather, what I gain is a way of framing, or understanding, your daughter's character and personality. Your testimony communicates more than mere facts; it offers an interpretation that guides how I should think about them. To appreciate the point, contrast with a flat assertion: "*Yes, she is an introvert.*" That answer might well yield knowledge regarding whether she is an introvert, but it lacks the richness

[39] Consider the familiar set of stock examples, e.g., "Paris is the capital of France", "There's milk in the fridge", etc. (Audi 2006; Zagzebski 2012; Hills 2016).

[40] For the knowledge norm view, see Williamson (2000) and Kelp & Simion (2021); for truth norms, see Weiner (2005) and Whiting (2013); for justification norms, see Lackey (2007) and Douven (2008); and for context-sensitive variations, see McKinnon (2015), Goldberg (2015b), and Gerken (2015). This list is far from exhaustive, see Pagin & Marsili (2025) for an overview.

[41] For the reductionist view, which requires positive reasons, see Fricker (1994, 1995), Van Cleve (2006). For the anti-reductionist view, which eschews the positive reasons requirement, see Burge (1993), Graham (2006; 2010), Simion (2021), Simion and Kelp (2020). This list, again, is far from exhaustive, see Leonard (2023) for an overview.

[42] I will for the most part bracket the thorny issue of specifying what, exactly, is said by a speaker in a given exchange, whilst noting that this is likely to be another respect in which the literature is guilty of oversimplification. As others have noted (e.g. Goldberg (2007: 54) and Peet (2016: 1)), the literature has tended to assumed a relatively unproblematic recovery process: roughly, S asserts that p, and H, in turn, recovers precisely that same proposition. I use the terminology of 'what is said' only to home in on an aspect of the literature that will later be relevant: namely, that further knowledge, beyond knowledge of the particular proposition uttered by a speaker, is not typically deemed relevant to the epistemic evaluation of a testimonial exchange. The paradigmatic epistemically good exchange, of the kind we've built our theories around, involves knowing *that p* on say-so *that p*. For work that addresses these questions head on, see Peet (2015a; 2015b; 2016) on context-sensitive utterances and the recovery problem, as well as Fricker (2012), Hawthorne (2012), and Haziza (2022) on implicatures.

[43] I will here restrict myself to three such examples for ease of exposition, but note that in the recent literature further examples of complex testimony abound. See, in particular, Fraser's (2021) *narrative testimony* and Keller's (2024) *testimony by presupposition*.

that allows me to better grasp the nature of her character. In such exchanges, testimony functions less as a vehicle for propositional content and more as a means of transmitting understanding.[44]

§2.2 Advice

Consider another case, this time from Sliwa (2025). Sara, a stay-at-home mother, feels frustrated and resentful toward her husband, who expects her to take an interest in his workday after she has spent her own day on household and childcare labour. A friend advises: *"It sounds like you need a vacation - or at least a weekend. You work all day too. You provide household and childcare labour, and then John asks for emotional labour on top of that. No wonder you're exhausted and irritated!"* (Sliwa 2025: 3). As Sliwa points out, such a communicative exchange does not merely leave Sara with some new belief about her situation - i.e., 'the concept of *work* applies to my situation', or 'my resentment is justified'. Rather, her friend suggests a new way of organising her thoughts about and perceptions of her situation. (Sliwa 2025: 4) As a result, she comes to see different features of her situation differently; she has new conceptual tools with which to think about and communicate her situation to others. Once again, the epistemic effect of such an exchange does not lie simply in the transmission of true propositions, but rather in Sara's new conceptual tool, or framing device.[45]

§2.3 Analogy

Consider one final case, this time borrowed and adapted from Hills (2020). A father wants his son to understand why stealing his sister's toy was wrong. Rather than assert as much, he offers an analogy: *"Do you know the story of the mother bear and her two cubs? A mother bear hunts for food for her cubs, then trusts the cubs to distribute it amongst themselves evenly. But, as soon as she falls asleep, the older cub hoards everything and the younger sibling goes hungry."* The son, by grasping the analogy, comes to see his own action differently: as akin to the selfish cub depriving his sibling. What he acquires is not merely a belief, i.e., that 'my actions were wrong,' but rather a particular way of appreciating why they were. Contrast the analogy with a case of straightforward assertion: "*What you did was wrong because it breaches the categorical imperative*." Even granting that such an assertion could similarly yield knowledge about the wrongness of his actions, the analogy plays an importantly different epistemic role: it makes certain aspects of his conduct salient to him (namely, those that bear the relevant similarities to that of the selfish cub), it suggests a particular way of framing and interpreting his situation that the straightforward assertion does not. Here again, testimony achieves something the standard model - which, recall, focuses primarily on assertion and belief - cannot explain. The epistemic value of the exchange lies not in straightforward proposition

[44] For the purposes of this paper I set aside the debate regarding the precise communicative role testimony plays in facilitating understanding. One may be a pessimist, and think that testimony can at best 'plant the seed' of understanding, rather than transmitting or generating it. Still, planting the seed of understanding is an importantly different epistemic role than transmitting propositional knowledge of what is said. For relevant discussion, see, e.g., Hills (2020); Grimm (2020); Malfatti (2021; 2025a); Hazlett (2025; forthcoming); El Shazly (forthcoming).

[45] Of course, this is not to say that new conceptual tools or framing devices will not facilitate, or make available, further propositional knowledge. Indeed, it seems very likely that they will. Sliwa's point is rather that the epistemic significance of such an exchange involves the transmission of a cognitive tool, or framing device, which does not itself reduce to merely believing some new proposition.

transmission, but in guiding the hearer toward a new way of conceptualising, organising, structuring their thoughts about the object of testimony.[46]

§2.4 Simple vs. complex testimony

Recall the traditional *assertion-and-belief* model of testimony, which has for the most part dominated discussion. This paradigm takes simple propositions as its guiding data, and interrogates the epistemic normativity of speakers asserting and hearers coming to believe such propositions. The three cases discussed above, however, do not fit this model. In *Explanation*, *Advice*, and *Analogy,* we find testimony doing something more (or perhaps just something different) than merely inducing hearer belief. If this is right, then theories of testimony built solely around the assertion-and-belief model cannot account for the full range of epistemically significant roles that testimony can play. To make better sense of testimony's epistemic value, we require a more expansive and multidimensional framework than one based solely on the exchange of propositional knowledge.

But one may push back: it may be argued that in our cases testimony bears epistemic value only (and precisely) to the extent that it yields knowledge. Complex testimony, on this view, does not show that the transmission of propositional knowledge is the wrong model, only that the path is sometimes less direct than in simple cases.

The problem with this response, however, is that it underplays the epistemic significance of communicative acts such as reframing, guiding attention, or providing conceptual tools. The hearer who has their attention shifted by a speaker, or is given a new conceptual tool, is subject to a significant epistemic interjection from their interlocutor. To be sure, for all I'll say here, epistemic evaluation of that interjection may well have to do with the propositional knowledge it produces, transmits, or otherwise makes available (I return to this question in the next section). But the point would remain: testimony's epistemic role in such cases is not merely reducible to the exchange of propositional knowledge of what is said. The problem lies not with propositional knowledge itself, then, but with the overly simple model of how testimony contributes to it.

Another worry is that *Explanation*, *Advice*, and *Analogy* can all be analysed simply as sequences of ordinary assertions, each assessable by existing theories. On this view, there is nothing epistemically distinctive about complex testimony - we already have all the resources we need to account for it.

[46] I will remain neutral over precisely how to cash out the distinctive epistemic effect of complex testimony. Perhaps, as some have suggested (e.g., Fraser 2021; Sliwa 2025) the speaker gives the hearer a new *perspective*, in Camp's (2017a; 2019) sense - a non-propositional, open-ended suite of dispositions to notice, explain, and respond to situations in the world. Perhaps, instead, the effects are best explained in terms of testimony-induced *understanding* (e.g., Hills 2020; Grimm 2020; Hazlett 2025, forthcoming). It may even be that such exchanges merely bring about more propositional knowledge than their simple counterparts. I will say more about the link between complex testimony and epistemic value in the next section, but suffice it to note for now that all such options - even the one drawing on further propositional knowledge - represent a departure from the traditional model, which, recall, sets out to explain testimony's epistemic value purely in terms of knowledge of what is said.

But the cases we've discussed resist such reduction; one has not necessarily grasped the analogy just by virtue of believing its constituent propositions.[47] I may believe that $p_1 \ldots p_n$ without thereby seeing, or grasping, how these propositions are supposed to hang together in explanation, or without thereby having my attention drawn to the fact that my conduct - in the relevant ways - resembles that of the selfish cub. Of course, it is an interesting question what exactly is involved in this further step that complex - and not simple - testimony involves, but we need not answer it to see that there is something distinctive going on here.[48]

§3 Norms on complex testimony

So far, I've drawn on recent literature to lay out the case for thinking that testimony plays a wider range of epistemically significant roles than the traditional model suggests. In this section I aim to shed light on why norms engineered in the face of simple testimony will fail to 'scale up' to more complex exchanges than those it was built to handle.

§3.1 Pluralism

To better understand the issue, let us frame it in terms of a broadly *teleological*, or *goal-directed* conception of epistemic normativity.[49] On this way of thinking about things, there is an intimate link between the norms that are supposed to govern our conduct, on the one hand, and the value, good, or goal that sets the standard according to which that conduct is evaluated, on the other.

On this line, epistemic norms are directed toward the realisation of epistemic value. Epistemic norms that pertain to our practice of testimony should be directed toward the ways in which epistemic value is realised via this practice. However, extant views in the vicinity have assumed that there is but one way in which testimony realises epistemic value: by functioning as a vehicle for propositional knowledge. Sometimes this view is given explicitly in our definitions of 'testimony'.[50] But even when it is not explicit, this account of the main (perhaps exclusive) way in which testimony realises epistemic value is nevertheless on clear display in the literature via the aforementioned focus on assertion and testimonial belief. We have only norms of assertion and norms of testimonial belief because those are the kinds of norms we need for ensuring that testimony does a good job in its core, epistemic role: of transmitting knowledge of what is said.

The problem, then, is that we have been theorising about the epistemic normativity of testimony from an incomplete picture of the ways in which it realises epistemic value. If testimony's epistemic role started and ended with the exchange of propositional knowledge, then it doesn't matter what form it takes, our normative theories could remain largely unchanged. Our cases from the last section, however, suggest that such a picture of testimony's epistemic value is

[47] Fraser (2021) makes a similar point with respect to the irreducibility of narrative discourses to their constituent propositions.

[48] Perhaps there is some reason to retain the 'epistemology of testimony' label solely for the study of our testimony-based beliefs, where this includes only beliefs whose content matches the propositional content of what is said. To avoid merely terminological dispute, if one is inclined by such a view on what counts as 'testimony', one should read everything I say in what follows as pertaining more broadly to the epistemology of communication instead.

[49] The reader is free to re-interpret the challenge in whichever way they prefer; nothing substantive should hang on my way of framing the challenge, the purpose is simply to get at least one concrete version of the problem in clear view. The teleological view allows me to do this in a neat way.

[50] See, e.g., Audi (2006), Zagzebski (2008; 2012), Hawley (2010) and Hills (2016).

incomplete. There are many ways, beyond the exchange of propositional knowledge, in which testimony seems to contribute to the realisation of epistemic value.

What we need, then, is a *pluralist* theory of testimony. Testimony plays not one but many epistemically significant roles, and our theories should not abstract away from this fact. As a result, there are a range of further norms, beyond those of assertion and belief, governing speakers and hearers from the epistemic point of view in testimonial exchanges. A speaker's having satisfied one's favourite norm of assertion, at least in the kinds of cases we've been interested in, won't explain their role in an epistemically fruitful testimonial exchange. A hearer's having satisfied one's favourite norm of testimonial belief, in those same cases, similarly won't explain what it is that they did epistemically well. A comprehensive and satisfying theory of the ways in which testimony helps us to realise our epistemic goals will not presuppose from the outset that it is propositional knowledge - and only propositional knowledge - that sets the standard according to which our norms are to be specified.

§3.2 Varieties of pluralism

Before setting out some desiderata for a theory such that it accommodates the pluralism we seek, let us firstly get clear on exactly what this pluralism amounts to. One might think that the challenge here amounts to epistemic value pluralism - the view that there are multiple, final epistemic values. But nothing I say here turns on that debate. The pluralism we seek concerns not epistemic value in general but rather *testimony's role in its realisation*. Even if knowledge is all that is of final, epistemic value, what the cases show is that this value is often realised via testimony in importantly different ways than the simple model suggests. As noted in §2.4, the problem is not that complex testimony cannot be explained in terms of knowledge, but rather that our model assumes it does so in only one way. Pluralism, then, of the kind we'll pursue in what follows, amounts to the view that testimony does not merely transmit propositional knowledge, it reframes questions, redistributes attention, and brings about understanding. These roles are epistemically significant in their own right; they are not reducible to the exchange of propositional knowledge.

§3.3 Desiderata for a normative theory of testimony

Let us now articulate some concrete desiderata for a new normative theory of testimony - one that can incorporate the pluralism we need.

Firstly, we need a wider range of norms on speakers' and hearers' conduct. Norms of assertion and norms of belief will not suffice. On the traditional view it was this conduct, and only this conduct, that falls within the remit of our epistemic (testimonial) norms because only this conduct bears directly on whether the relevant epistemic goal (namely, the exchange of propositional knowledge) is realised. Our epistemic norms do not need to evaluate the telling of jokes, for instance, because this is not an aspect of our conduct relevant to the exchange of propositional knowledge - unlike assertion and belief.

But if I'm right, we'll need a wider set of epistemic norms - governing different aspects of speakers' and hearers' conduct - to reflect the plurality of ways in which testimonial exchanges can realise epistemic value. On the speaker's side, epistemic norms on explanation, advice-giving,

and so on, will all be relevant. It is a speaker's epistemic success with respect to these roles - and not merely the role of asserting - that explains why some testimonial exchanges went well. And on the hearer's side, we'll need norms on mental conduct relating to the uptake of conceptual resources, explanations, perspectives, patterns of attention, and so on.[51] It is the hearer's epistemic success with respect to these aspects of our mental conduct - and not merely that of believing and knowing - that explains why some testimonial exchanges went well.

Secondly, accommodating pluralism will require not only a wider set of norms to reflect the wider set of relevant epistemic values, but also some way of determining which, of potentially many, is operative in any given exchange. The reason for this is as follows. Suppose you agree with everything said so far; we need to countenance a plurality of dimensions along which to evaluate testimonial exchanges to reflect the plurality of ways in which testimonial exchanges can go epistemically well. Still, whichever way you go in so doing, you will need a principled recipe for adjudicating which evaluative standard is operative in a given testimonial exchange. To illustrate, consider the following exchange.

> **Child** Mum, is there milk in the fridge?
> **Mother** Yes, there's milk in the fridge.

It is uncontroversial that, in this way, the child can come to know that there's milk in the fridge. Moreover, that propositional knowledge is successfully exchanged looks like a perfectly good explanation of the precise way in which testimony realises epistemic value in such exchanges. Even if other exchanges, like those involving explanations or analogies, conduce to epistemic value in importantly different ways, exchanges like this one do not. Such evaluative standards, whatever they may turn out to be, are simply not operative in exchanges like this one. But if this is right, then the pluralist must provide a supplementary story regarding which of potentially numerous evaluative standards are operative in any given exchange.

§4 Testimony and inquiry

In recent times there has been a surge of interest in *zetetic*, or *inquiry-centric*, epistemology.[52] I'll argue that a zetetic approach to normative epistemology has the resources to provide a promising, novel theory of testimony's epistemic value - one that can avoid the trouble that befell extant frameworks whilst meeting our new desiderata. In particular, I'll draw out two distinctive features of the zetetic approach and try to show how they can help us to account for the full range of testimony's epistemically significant roles.

§4.1 Normative breadth

Suppose I want to figure out how many windows the Chrysler Building in Manhattan has.[53] Doing so will require setting up camp outside Grand Central Station and dedicating roughly an hour to

[51] There is plenty of precedent for such epistemic norms in the literature upon which we may wish to draw, e.g., Watzl (2022); Munton (2022); Whitcomb (2023); Carruthers (2024); Willard-Kyle et al. (2025).
[52] Among increasingly many others, see: Hookway (2006); Friedman (2019; 2020); (Kelp 2018b; 2021a); Willard-Kyle (2023; forthcoming); Flores and Woodard (2023).
[53] This case is from Friedman (2020).

focussed counting. In that time, and in those circumstances, I will need to forego an abundance of epistemic gains - the opportunities I will have to either form new beliefs or revise existing ones will be close to unending. To get the job done, I need to forego these opportunities. And note that this is so even if the beliefs I would thereby form would be perfectly justified - perhaps even knowledgeable. Success in inquiry thus sometimes involves foregoing certain epistemic gains. Ignoring evidence, for example, at least for certain stretches, might be *zetetically* even if not *doxastically* justified.

As I'm thinking of it here, the zetetic approach counts the judgements and evaluations oriented toward success in inquiry as every bit as epistemic as those oriented toward success in belief. Contrast this with the bulk of traditional, analytic epistemology, wherein the sole locus of epistemic evaluation is an agent's *beliefs* and *belief-like attitudes* - i.e., whether they are justified, well-evidenced, coherent, and so on.[54] The broadening of epistemic evaluation beyond belief, then, is characteristic of the zetetic turn in epistemology. We should think of the epistemic agent less as a *belief-haver*, and more as a *question-settler* (Hookway 2006). For ease of exposition, let's summarise this fundamental, zetetic principle and give it a name:

> **Normative breadth** Epistemic norms pertain to an agent's conduct across the whole process of inquiry, not just their beliefs.

The traditional approach to epistemology focussed on one particular aspect of our conduct: the formation, revision and maintenance of beliefs and belief-like attitudes. **Normative breadth**, however, brings into play epistemic norms on non-doxastic conduct. An agent's curiosity, the questions they take up, and their relative attentional distribution, for instance, are - at least in principle - within the remit of the epistemic on this approach. In particular, they are so because they may contribute to success in inquiry, even if not success in belief.

§4.2 Normative flexibility

Alongside **Normative breadth**, a further zetetic principle - this time from Kelp (2018b) - will be helpful. Kelp draws the distinction between two kinds of inquiry we often engage in:

> **Inquiry-questions** Inquiry aimed at closing questions with *true belief/knowledge* whether p
>
> **Inquiry-phenomena** Inquiry aimed at closing questions regarding the nature of phenomena with *understanding* (Kelp 2018b: 1)

Here's the thought. We're not interested solely in belief, we're interested in inquiry; epistemic agents seek to settle questions, not merely form and revise beliefs. But good inquiry is not a uniform evaluative standard; different questions aim at different epistemic goods. The point may

[54] For fuller discussion on the relation between the zetetic and the traditional epistemic realm, see Hookway (2006), Friedman (2020), Thorstad (2021) and Fleisher (2022). For pushback on the idea that zetetic norms should replace or take precedent over epistemic norms, see Thorstad (2022) and Falbo (2023).

be spelled out in several different ways, but I take it that at its heart is a compelling idea: while inquiry aims at question-settling, what exactly this looks like will vary from question to question.[55]

Sometimes what I seek is an answer to a whether-question; I just need to know whether or not *p* is true. Along these lines, we can evaluate my conduct relative to its conduciveness to an epistemically appropriate culmination point for such an inquiry, namely true belief or knowledge. Other times my inquiry is different. It concerns not merely the answer to whether *p*, but rather the nature of some phenomenon. Such inquiries aim not at true belief, nor knowledge, but rather understanding. Coming to understand Newtonian mechanics, or the British sense of humour, will require something different of me - both cognitively and practically - compared to coming to know whether or not Jenny was at the party.

Drawing on Kelp (2018b), then, we see that while the relevant epistemic norms are structured according to the value of good inquiry, what exactly that looks like will depend on the nature of one's question.[56] We can thus make explicit a further zetetic principle that will soon come in handy:

> **Normative flexibility** Epistemic evaluation is inquiry-relative.

§5 Zetetic epistemology of testimony

Now, with the zetetic backdrop in place, we can shift our attention back to testimony. I'll first detail what a distinctively zetetic approach to testimony might look like, then I'll explain the way in which such an approach does well by our two desiderata.

§5.1 Deploying the zetetic approach

Zetetic epistemologists think that traditional epistemology has overemphasised the importance of belief, to the exclusion of other, equally important elements of inquiry. With this in mind, the zetetic epistemologist has a readymade diagnosis of what's gone wrong in our key cases from §2. All these cases are simply a communicative manifestation of what is a much broader issue: there are a range of epistemically significant aspects of our conduct beyond the straightforwardly doxastic. In the domain of testimony, just like elsewhere in epistemology, too much onus was placed on the culmination point of inquiry - the point at which agents form judgements. But good inquiry is multidimensional; some conduct may be epistemically relevant and epistemically evaluable without directly bearing on belief.

I thus submit that the zetetic gloss on the traditional approach to testimony is as follows. We have been supposing that testimony plays but one epistemically relevant role: *speakers close hearers' inquiries*. We have been preoccupied with the cases in which H, by way of S's testimony that *p*, thereby closes inquiry into whether *p*. We then ask whether S was right to recommend - by way of their testimony - that H close their inquiry.

---

[55] See also Smith (2020: 183)

[56] Following Dover (2024), we may also wish to countenance a further kind of inquiry, namely one that is not mediated by a question at all. I set this aside in what follows, whilst registering amenability to expanding further still.

But what our cases demonstrate is that closing inquiry is but one of many epistemically relevant roles that testimony can play. If we reconceive of the epistemic agent as an inquirer, and not merely a believer, then we create space for a range of epistemically significant roles that testimony might play in the life of an inquiring agent. Speakers in testimonial exchanges take epistemic responsibility for their hearer's inquiry, but they need not do so merely by recommending its closure. They might suggest a new way of framing the hearer's already-acquired evidence; they might offer a new conceptual tool with which to interpret one's situation; they might recommend that attention is redistributed - factor A is less important than you had realised, whereas factor B is more so. These roles need not change beliefs, but they nonetheless shape inquiry.

Zetetic epistemology's alternative characterisation of the epistemic agent (from *believer* to *inquirer*), then, offers a new way of conceptualising the epistemic significance of testimony. As a result, a range of further standards of evaluation come to the fore. With respect to the object or topic of testimony, we need to think not only about which propositions constitute the facts, but also about what a good allocation of attention, for such a topic, would look like.[57] We need to think about which facts are central, and which are merely peripheral, in order to understand whether my testimony draws undue salience to one feature over another.[58] We need to think about what the right, as opposed to the wrong, kinds of questions are to probe regarding some object or topic.

Of course, these are bound to be difficult questions, over which many will disagree. But that these are all significant moves in inquiry makes clear why they are nonetheless questions worth asking. Once we begin to think of the epistemic agent as an inquirer, and not merely a believer, the traditional normative epistemic roles carved out for speakers (asserters) and hearers (believers) begin to look unmotivated and implausibly restrictive.

§5.2 Desideratum 1: Normative breadth and flexibility

I've sketched an inquiry-centric framework within which to theorise about testimony. Let us now return to our desiderata, with **Normative breadth** and **Normative flexibility** in hand, to see how our new view fares.

Recall that the first desideratum concerned pluralism. We sought testimonial norms on a wider range of our communicative conduct to reflect the wider range of ways in which testimony conduces to epistemic value. To see why the zetetic approach fares well in this regard, firstly note that speakers often influence hearer's inquiries even when they do not straightforwardly change their beliefs. Now, while this is neither a novel nor revelatory insight about testimony, in **Normative breadth** the zetetic framework builds it into the normative apparatus of the view.[59] The problem for traditional views arose in the first place because normative differences between instances of complex testimony were not of the ilk that they can account for. As we saw, what

[57] For relevant discussion see Anderson (1995); Siegel (2022); Whiteley (2024).
[58] See Munton (2021; 2022) and Whiteley (2022; 2023) for (epistemically and ethically) normative accounts of salience patterns.
[59] Moran, for instance, noted that "There are more ways of changing someone's mind than changing his or her beliefs…" (1994: 100). Similarly, Hills pointed out that sometimes speakers do not merely transmit or generate knowledge, but instead "set the terms of hearers' inquiries" (2020: 404).

makes for epistemically good complex testimony is not merely a matter of epistemically good *assertion-and-belief*.

But **Normative breadth** shifts the focus away from belief. What it is that the speaker is on the hook for in a testimonial exchange is not solely the hearer's testimonial belief. Similarly, a hearer is epistemically evaluable not merely when they form a testimonial belief, but also when they take up their interlocutor's explanation, heed their advice, or grasp their analogy. Such modes of uptake shape and structure their inquiry even if the result is neither transmitted beliefs nor propositional knowledge. The zetetic framework, by broadening the remit of the epistemic beyond the doxastic, has the resources to evaluate the aspects of speakers' and hearer's conduct that fell beyond the scope of the traditional view.[60]

Normative breadth is thus helpful insofar as it expands epistemic evaluation beyond doxastic conduct. But what's more, it does so in a principled way: it is *because of* the range of ways in which testimony can realise epistemic value that epistemic evaluation is expanded in this way. To see this, recall that, by **Normative flexibility**, epistemic evaluation is inquiry-relative. What one ought to do, epistemically speaking, depends on the nature of their inquiry. Think back to our first case, in which a father describes his daughter's introverted personality. Now consider two contexts in which the exchange might've occurred. In one, the hearer is gathering data for a social psychology midterm, i.e., it is a simple whether question, and they're tracking the answers. In such a case, a variation on which the father simply asserts "*Yes, my daughter is an introvert*" will do perfectly well. The father contributes a simple assertion, for which he has excellent evidence, and the hearer for their part is well aware of this fact.

Now suppose that the exchange takes place in a different context. The speaker and hearer's daughters grew up together, but they haven't seen each other for years, and so the hearer is just curious to find out how her friend's daughter turned out. That is, the hearer is not merely pursuing what Kelp would call *inquiry-question*, but rather something more akin to *inquiry-phenomenon*. They want to know, and perhaps also to understand, what the speaker's daughter is like. In the context of this, rather than the former kind of inquiry, we can begin to see the epistemic value of the indirect explanation. In this context, it is very plausible that the indirect explanation can bring you to better understand her character - the simple 'yes/no' answer to the whether question lacks something that the indirect answer provides.

The first desideratum had two strands: we need a broader set of norms to reflect the broader set of ways in which testimony realises epistemic value. By **Normative breadth** and **Normative flexibility**, the zetetic approach captures both strands. Not only do we have a broader range of norms at our disposal, but we have these norms precisely because there is a broader range of ways in which testimony realises epistemic (or better: zetetic) value.

[60] Of course, it is an open question how best to specify the full range of zetetic norms on speakers and hearers. My purpose here is simply to point out that, by **Normative breadth**, the zetetic approach looks comparatively well-placed to account for complex testimony.

§5.3 Desideratum 2: Unification

So far so good, but one final hurdle remains: how, if at all, can we unify the standards of epistemic evaluation in testimonial exchanges? How and why is it the case that one testimonial exchange is held to an entirely different standard than the next? In particular, we need some principled, non-ad hoc way of specifying how this shifting of evaluative standards can occur.

And I think here, too, we see that the zetetic approach can offer a satisfying answer. To see why, note that the evaluative standard, across all epistemic evaluation, is *good inquiry*. This standard is itself flexible only to the extent that different questions admit of different settling conditions - it imposes a flexible standard only insofar as the objects of our inquiries themselves differ from one another in important ways. It is because good inquiry into *whether your daughter is an introvert* places different demands on an agent than does inquiry into *what your daughter is like* that an agent engaged in the former inquiry is evaluated differently than an agent engaged in the latter inquiry. This flexibility is thus neither ad hoc nor unprincipled; it is concretely tied to the nature of the inquiry. The demands we place on epistemic agents - the norms they must follow and the evaluative standards they must meet - do not float free of the epistemic value at issue in their inquiry. To the contrary, the value sets the standard. Progress toward the proper closure of one's question is thus always the unifying, gold epistemic standard, even if what proper closure looks like is dependent on the question itself.

We can now see, then, that our two zetetic principles, applied to the domain of testimony, help us to make sense of the phenomenon of complex testimony. There are different ways in which testimony can go epistemically well or badly, different ways in which speakers and hearers can behave epistemically well or badly, and this is no surprise once *good inquiry* is placed at the normative centre-point in epistemological theorising. Testimonial normativity pertains to more than just our practices of asserting and believing precisely because a wider range of our conduct is recruited in pursuit of good inquiry via communication.

§6 Zetetic dependence

I've argued that in **Normative breadth** and **Normative flexibility** the zetetic epistemologist has the resources to offer a uniquely appealing characterisation of the epistemic normativity of testimony - one that is both independently plausible and can accommodate the new, complex data that traditional views could not. In §7 I will consider some key objections to the inquiry-centric view here developed. But before that, I want to begin to fill in some of the details of the positive position I am developing. Though I will not be able to do full justice to it here, I want to at least gesture at an answer to the question of what *zetetic* as opposed to *epistemic* dependence looks like. In other words: if the normative, epistemic roles carved out for speakers and hearers are not, as the traditional view would have it, that of *asserter* and *believer* respectively, then what are they?

§6.1 Speakers assert, hearers believe

As noted at various points throughout, traditional normative theories of testimony maintain sharp focus on one particular evaluative standard: the exchange of propositional knowledge. Speakers, primarily, assert propositions; hearers, primarily, form testimonially-based beliefs. Epistemic norms on the speaker's side of a testimonial exchanges start and end with the epistemic norm(s)

of *assertion*. On the hearer's side, the key question is whether reductionism, anti-reductionism, or perhaps some hybrid view, best captures the epistemic normativity of testimonially-based *belief*.

Notice, moreover, that with the context of zetetic vs. doxastic epistemology now on the table, this approach to normative epistemology of testimony is somewhat unsurprising. If, as many have thought, beliefs and belief-like attitudes constitute the primary remit of epistemic evaluation, then of course the interesting questions surrounding the epistemic normativity of testimony will have to do with the ways in which testimony produces, generates, and transmits beliefs from speakers to hearers. Speakers communicate their beliefs by way of assertions, hearers take up speakers' assertions by way of testimonially-based beliefs.

Traditional epistemology of testimony thus adopts the following broad structure:

Speakers → *Asserters*
Hearers → *Believers*

The arrows denote the respective normative, epistemic roles occupied by speakers and hearers in testimonial exchanges, and, accordingly, the range of conduct for which each is thereby epistemically evaluable. On this, dominant line, speakers in testimonial exchanges perform better or worse, from the epistemic point of view, to the extent that their assertions are epistemically better or worse. Same goes for hearers and their testimonially-based beliefs. From this framework we get a narrowly circumscribed range of norms (i.e. norms of assertion and norms of testimonial belief.)

§6.2 Speakers teach, hearers inquire

But *zetetic* epistemologists, I contend, would and should be particularly unwilling to passively adopt this broad and hitherto pervasive conception of testimonial exchanges. Displacement of belief as the normative locus of epistemological theorising clearly displaces testimonial belief in the domain of testimony. And from there, the disproportionate focus on assertion as the primary, paradigmatic, or perhaps even exclusive vehicle for testimony, is unlikely to stand. It is of course not that belief and assertion are or would be sidelined on a zetetic approach; rather, what I am doubting is that they would simply passively assume centre stage in our normative, epistemic theorising in quite the same way.

Here's how I think a zetetic reimagination of speakers' and hearers' normative, epistemic roles might look:

Speakers → *Teachers*
Hearers → *Inquirers*

Just as there are a range of aspects of our conduct - beyond merely our doxastic conduct - that the zetetic approach draws into the remit of epistemic evaluation, so too, we might expect, for our *communicative* conduct. Our practice of testimony, and of communication more generally, falls into the remit of traditional epistemology to the extent that we exploit it to influence one another's

beliefs. But in zetetic epistemology there is significant scope for this practice to fall within the remit of epistemic evaluation irrespective of its effect on our beliefs. At least in principle, there are a range of different stages of inquiry, beyond the culmination point at which beliefs are formed, at which testimony might play a role. If so, the zetetic framework will epistemically evaluate its contribution to inquiry in much the same way as it would evaluate a false assertion's negative contribution to the hearer's inquiry.

Given this broad approach, I take it that the shift from *hearers as believers* to *hearers as inquirers* is somewhat self-explanatory. We're not epistemically interested in agents only to the extent that they form (/revise/abandon) beliefs, and so we're not epistemically interested in hearers only to the extent that they do such things on the basis of testimony. A zetetic approach to the hearer's side of a testimonial exchange thus would not think of such agents primarily as (testimonial) believers because they do not think of epistemic agents more generally primarily as believers. Inquiry involves many different things, epistemic evaluation looks to the whole process, not just its culmination point.

The shift from *speakers as asserters* to *speakers as teachers*, however, requires a little more unpacking. Whilst I am open to the possibility that other analogies might better capture a distinctively zetetic approach to evaluating speakers, there are a couple of reasons why I think this one is useful. First, note that it is sufficiently uncontroversial as to sound platitudinous that teachers frequently deploy creative and diverse communicative and pedagogical techniques in order to impart epistemic goods upon their students. Common and oft-cited techniques in the philosophy of education include narrative, storytelling, dialogue, metaphor, and analogy (Dewey (1916; 1933); Egan (1993; 1997); Noddings (1995); Elgin (1999a; 1999b; 2017); Freire (2000)). To the extent that such techniques clearly exploit non-assertoric (or at least not straightforwardly assertoric) modes of communication in pursuit of decidedly epistemic aims and goals, they support the idea that the teacher model is a useful one for the zetetic model of evaluating speakers - which as we have noted is apt to epistemically evaluate speech not only in terms of its effects on hearer's beliefs.

But it is also worth noting, in support of the teacher analogy, that there is already a substantive tradition in the epistemology of testimony that places teachers close to the epistemic gold standard when it comes to good speaking. Lackey's (2008: 48) well-known **CREATIONIST TEACHER** case is but one notable example; there are others.[61] These examples get their bite precisely because of the widely held intuition that teachers exemplify good communication in a distinctively and recognisably epistemic sense. It is not, after all, for their own amusement that teachers typically deploy such creative and diverse means of communication; rather, they do so with the aim of effectively and efficiently imparting epistemic goods such as knowledge and understanding upon their students. If the effective and efficient exchange of such epistemic goods is not the core locus of our analysis in the epistemology of testimony, then what is?

To the extent that teachers often (i) exploit creative, diverse, often non-assertoric, communicative means and (ii) do so in a way that is nevertheless decidedly epistemic, I submit that they constitute

[61] See, among others, Goldman (1991); Elgin (1999a; 1999b); Zagzebski (2008); Goldberg (2013); Carter & Nickel (2014); Malfatti (2019; 2025b); Gaszczyk & Grodniewicz (forthcoming).

a superior paradigm around which the normative, epistemic role of the speaker in a testimonial exchange ought to be theorised.

§7 Overinflation of the epistemic realm

I've argued that testimony's epistemic value does not lie in its capacity to transmit propositional knowledge but rather in its contribution to inquiry. A natural worry about such a view concerns overinflation: if any contribution to inquiry counts as epistemically evaluable, we risk including communicative phenomena that do not belong in the epistemic realm. Two cases will illustrate versions of the concern.

> **FALSE TESTIMONY** S's false testimony *that the sky is green* causes H to have a revelatory breakthrough in their PhD thesis on the metaphysics of colour.

Clearly, S's testimony is not epistemically good. It is false, unevidenced, and, we may stipulate, uttered purely for self-amusement. Still, it clearly has a profoundly positive effect on H's inquiry. If we want to model the epistemic normativity not on the transmission of epistemically good propositions but instead on testimony's effect on inquiry, are we not forced to epistemically commend such cases?

We are not. Even granting, as we have in this article, that epistemic normativity is teleologically structured, i.e., that epistemic norms are specified by reference to the realisation of epistemic goals, there are many plausible routes for blocking the idea that any conduct whatsoever that produces epistemic good is thereby epistemically required. One prima facie plausible option here is a kind of rule-consequentialism.[62] On this line, epistemic norms are epistemic *rules*: they are the rules that creatures like us should follow to systematically bring about epistemic value. On such a view, that some conduct produces epistemic value (even a great deal of it) in any given moment is insufficient for it to qualify as an epistemic rule. The results in terms of epistemic value were creatures like us to follow a rule like *speak falsely* would be catastrophic; and this is unaffected by particular, idiosyncratic instances in which things go exceptionally well.

We can rule out **FALSE TESTIMONY**, then, by virtue of the fact that *speak falsely* is not an epistemic rule. But notice that it is comparatively very plausible that the norms we'd find in the zetetic space I've carved out in this article are genuine epistemic rules. Norms forbidding speakers from foregrounding the peripheral information and backgrounding the central information, for instance, or drawing a misleading analogy, can clearly be defended on grounds that abiding by such norms systematically leads to epistemic value.

Consider a further, related case, which requires slightly different treatment.

> **CHARISMATIC TEACHER** Consider two teachers, Ann and Bo. Ann is further along as an educator - they are more experienced and employ innovative pedagogical strategies.

[62] There are of course others available. I leave it to other teleologists, with other commitments, to resist such cases in whichever way they choose.

> Bo, by contrast, is less advanced but has a charismatic speaking style that reliably keeps students engaged.

The earlier **FALSE TESTIMONY** case was easy to dismiss on rule-consequentialist grounds: systematically, false assertions are epistemically disastrous. But that strategy will not help here. Bo's charisma might stably promote epistemic value - it could systematically facilitate better inquiry. Yet this looks like the *wrong kind of factor*. Whether my testimony is epistemically good should not be determined by the tone of my voice.

One revisionary response would be to embrace the zetetic implications in full: if charisma systematically conduces to inquiry, then it is genuinely epistemic. Some may be open to such a move.[63] But more conservative strategies are available, ones that others may find easier to stomach. One option draws on Flores & Woodard's (2023: 11) responsibility-based method for delineating the epistemic.[64] Their view appeals to the idea that we hold each other epistemically responsible for conforming to epistemic norms. So, while it is true that many factors may help or hinder inquiry in an indirect way, our practices of epistemic appraisal are instructive when it comes to determining which of these contributing factors are underwritten by genuine epistemic norms. For example, a bad night's sleep might make our reasoning sloppy, but in such a case, it is the sloppy reasoning that is the subject of our epistemic criticism, not the bad night's sleep that set the stage for it (2023: 22-23). Following Flores & Woodard's lead, then, we can exploit a key difference between **CHARISMATIC TEACHER** and our guiding examples of complex testimony. We do not epistemically criticise testifiers directly on grounds of charisma. By contrast, we can and do criticise misleading explanations, distorting analogies, or ill-suited advice. The responsibility-based view thus offers one promising strategy for countenancing epistemic evaluations of complex testimony without thereby overinflating the epistemic domain.

A complementary strategy is to draw the boundary in terms of *content*. Very plausibly, epistemic assessments of testimony should be assessments of *what is said* (content) rather than *the way in which it is said* (mere style).[65] The challenge, then, is to broaden the conception of content enough to capture complex testimony while excluding merely stylistic features. Fortunately, this can be done in a principled way. On my view, content should be conceived broadly: conceptual tools, framings, and analogies conveyed by testimony all count as part of what is said, and so are genuine sites of epistemic evaluation.[66] By contrast, merely stylistic features such as charisma, tone, or volume do not, since they are not bound up with what is said. This broader conception also coheres with how recent literature presents the relevant cases from §2: they are introduced as genuine instances of testimony in their own right - not reducible to their constituent propositional content. Such cases are philosophically instructive precisely because they show that speakers often communicate more than brute propositional contents of their utterances.

---

[63] E.g. Friedman (2020), Singer & Aronowitz (2022); cf. Kelly (2003), Berker (2013).
[64] See also Kelp & Simion (2021: 24).
[65] This is not meant to suggest that epistemic evaluation of content reduces to bare truth or knowledge. On most views, such evaluation is partly sensitive to how content is produced (e.g. justification, reliability). The point is simply that the target of evaluation is what is said - broadly construed - rather than the stylistic manner of delivery.
[66] The literature on implicatures provides further precedent for a broad conception: many argue that epistemic norms govern not only explicitly asserted, but also pragmatically implicated contents (Hawthorne 2012; Peet 2015a; Haziza 2022; cf. Fricker 2012).

For those wary of full-blown zetetic revisionism, then, these responsibility- and content-based accounts offer principled ways of preserving zetetic pluralism about testimony's epistemic roles while denying that charisma or delivery style are themselves epistemically significant factors.

§8 Conclusion

The traditional view accounts for testimony's epistemic value primarily in terms of the exchange of propositional knowledge: speakers assert, hearers believe, and, all going well, knowledge is exchanged. I argued that this model cannot account for many ordinary testimonial exchanges that shape inquiry without simply transmitting belief. Such cases reveal a lacuna in extant normative theories of testimony. To fill that gap, I proposed a novel framework, drawing on recent work in zetetic epistemology, to highlight that testimony contributes to inquiry not only by transmitting knowledge but by supplying conceptual resources, framing questions, and directing attention. This inquiry-centric approach accommodates the plurality of epistemic roles testimony can play whilst avoiding problematic overinflation. If this is right, the epistemology of testimony should be integrated more fully into the zetetic turn: testimony is not merely a ubiquitous source of belief, but rather an invaluable resource in inquiry.

Chapter III.

Gricean maxims as epistemic norms

Abstract: Standard approaches to the normativity of assertion draw a sharp line between *epistemic* and *pragmatic* norms. The former govern our conduct in accordance with standards such as *truth, knowledge,* and *evidential support*, whereas the latter govern it in accordance with standards such as *relevance, clarity,* and *manner*. This article challenges that picture. Focusing on Grice's (1975; 1989) conversational maxims as a canonical articulation of a widely accepted epistemology-pragmatics divide, I argue that there are genuine epistemic norms captured by all Gricean maxims, not just Quality, as is standardly assumed. To reach this conclusion, I draw on familiar resources for theorising both about epistemic normativity in general and the epistemic normativity of assertion in particular. I then show how to embrace Gricean epistemic norms whilst avoiding some key objections. If I'm right, the upshot is significant: the boundary between epistemology and pragmatics must be reconsidered, as the two domains are far more tightly intertwined than has hitherto been assumed.

§1 Introduction

There are many ways to be a good speaker. One might be knowledgeable, polite, lucid, relevant, eloquent, insightful, to name but a few. But which of these numerous standards of evaluation, we may ask, denotes a genuinely epistemic standard? Some are obviously epistemic (e.g., *knowledgeable*); some are certainly not (e.g., *polite*). But what about *relevant, lucid* or *insightful*? It's less clear.

In this article I argue for a perhaps initially surprising conclusion:

> **Gricean epistemic norms** There are epistemic norms corresponding to all of Grice's (1975; 1989) maxims.

Epistemically good assertion, I'll argue, is not only knowledgeable, but, at least *ceteris paribus*, relevant, clear, and perspicuous. In certain cases, one breaches a genuinely epistemic norm by saying something utterly irrelevant or needlessly obscure. This goes against the grain somewhat; theorists interested in the epistemic normativity of assertion, as I'll highlight, have for the most part converged over a view I call Quality-exceptionalism, the idea that there is (at most) an epistemic norm corresponding to Grice's maxim of Quality, but not Quantity, Relation, nor Manner. If a speaker communicates a proposition for which they have sufficient evidence, either

explicitly by way of outright assertion or implicitly by way of pragmatic implicature, then they are off the hook, at least from the point of view of epistemology.

While initially appealing, I'll argue that this view is mistaken. My strategy will be to reach the surprising and controversial thesis just outlined from a comparatively mundane and uncontroversial starting point. Here it is:

> **Boundedness** Our limitations – both psychological and environmental – are normatively relevant; they make a difference to what and how creatures like us ought to think and believe.[67]

I'll say more about both theses – what they each amount to, where they come from, why the first is controversial and the second is not – in due course. But the main aim of this paper is to sketch the route from **Boundedness** to **Gricean epistemic norms** (hereafter, just **GENs**). If you accept **Boundedness**, as many do, you ought also to accept **GENs**.

Here's the plan. In §2 I'll lay out some cases of what I call *merely conversational* failings – which is to say non-Quality violations – and explain the standard reaction to them, given extant accounts of the epistemic norm of assertion. Then, in §3, I'll present a range of arguments for **GENs** – some reasons to think that the standard response is mistaken; these are genuinely epistemic failings. In §4 I'll respond to some objections, discussion of which will help to clarify the key thesis. Along the way, I'll say more about **GENs**, how to embrace them, and where all of this leaves us.

§2 Merely conversational failings

§2.1 Examples

Consider the following exchange:

EYES Ollie asks Rory: "Where's the talk?" Rory responds: "In the Reid Room. It's upstairs, second door on the left, and the distance between my eyes is 62mm."

The distance between Rory's eyes is indeed 62mm. This information is utterly irrelevant to Ollie: Rory does not intend to subtly communicate some link between their eyes and the location of today's talk, and, at least as far as Rory knows, Ollie has no further interest in or use for this information. It is just what it looks like: true, perhaps knowledgeable, but irrelevant.

[67] Stronger implementations of the idea that epistemic norms must be theorised with cognitive limitations in mind are associated with the bounded rationality programme (see, e.g., Simon 1955; Gigerenzer & Selten 2001 for canonical formulations. See Greco 2023; Thorstad 2024 cf. Pasnau 2013; Carr 2021 for application to epistemology and discussion). The appeal to **Boundedness** in what follows is considerably weaker. It is closer in spirit to a general epistemic "ought implies can" constraint: epistemic norms govern cognitively limited and non-ideal agents, with limited access, attention and processing capacities. Theorising about epistemic norms should not idealise and abstract away from these limitations. This weaker background assumption is operative across a myriad of different debates in epistemology, and taken for granted by otherwise divergent approaches, including (among many others) virtue epistemology (e.g. Sosa 1991; Greco 2010); knowledge-first epistemology (e.g., Williamson 2000); reliabilism (e.g., Goldman 1986); epistemic utility theory (Pettigrew 2022); zetetic epistemology (e.g., Friedman 2019, 2020; Kelp 2018b, 2021a, 2021b).

Consider a further case:

RENT Chris has recently moved into a new place and receives the following letter from Lisa, their new landlord: "Notwithstanding any provision to the contrary contained herein, the party of the first part (hereinafter referred to as 'the Grantor') shall, on or before the final business day of the current calendar month, remit payment in full, free of any and all encumbrances or claims whatsoever."

Nothing remotely substantial is added by the painstaking detail in Lisa's letter; they might just as well have written: "Rent is due by the end of the month" or some equivalent. Nonetheless, be that as it may, everything they utter is true. If there are objections to be made to Lisa's assertion(s), they will not be on grounds of a lack of evidential support.

Consider one final case:

OVERLOAD Ali, in conversation with his 4-year-old brother, points out that: "In any sufficiently powerful formal system (like arithmetic), there are true statements that cannot be proven within the system."

Ali, let's stipulate, has no grounds for thinking that his 4-year-old brother is a mathematical prodigy; just like most children his age, he has recently been introduced to basic arithmetic. Nonetheless, Ali once again knows what he says. His assertion may be somewhat predictably inappropriate - poorly calibrated to his audience – but epistemically speaking it looks to be just fine.

§2.2 Merely conversational failings

I give these cases only to bring into clear view what I take to be the received view in the literature when it comes to the interrelation of *conversational* and *epistemic* normativity. Let's now unpack this a little further. I'll put the standard view in terms of Grice's influential theory of conversation:[68]

> **Quality-exceptionalism** Of the Gricean conversational maxims, only Quality corresponds to a genuine epistemic norm; Quantity, Relation and Manner are merely conversational.

According to Quality-exceptionalism, a speaker breaches a distinctively epistemic norm in asserting iff they violate Quality. To better understand the view, let's remind ourselves of Grice's (1975) maxims:

[68] It is worth emphasising that nothing in this article turns on any substantive commitment to Grice's theory of conversation. The Gricean framework is deployed here as a particularly clear articulation of the epistemology-pragmatics divide. Our concern is with the interrelation between the epistemic and the conversational, and Grice's well-known maxims provide a useful way of making explicit where discussions of the epistemic normativity of assertion have traditionally drawn the relevant lines.

> ***Quantity*** Make one's contribution no more or less informative than is required for the current purposes of the talk exchange.
> ***Quality**** Do not say either what you believe to be false or lack sufficient evidence for.
> ***Relation*** Be relevant.
> ***Manner*** Be perspicuous; avoid unnecessary obscurity or ambiguity.[69]

The ***** indicates that Quality, unlike the others, has epistemic import, which means (irrespective of how it should be understood *qua* conversational maxim) that it corresponds to a legitimate, epistemic norm of assertion. For the purposes of this paper, what we're interested in is epistemology, not pragmatics; Quality-exceptionalism is an epistemological view about where to place the *****. We will thus set aside the various questions and debates that Grice's view provoked in pragmatics and philosophy of language, regarding how best to specify and unpack the conversational maxims, whether they really capture all that Grice intended, whether Quantity, Relation, and Manner can in fact be subsumed in one, and so on.

It is uncontroversial that some rule corresponding to Grice's maxim of Quality is a genuine epistemic norm of assertion, and that the other, conversational maxims are just that: merely conversational maxims. They do not also correspond to legitimate epistemic constraints on assertion. Rather, they are the kind of rules and imperatives one abides by *insofar* as one is engaged in rational, cooperative inquiry to some shared conversational end (Grice 1989; Stalnaker 1978) or *insofar* as one can be assumed to be engaging in sincere, joint pursuit of the QUD (Roberts 1996; 2018). By all accounts, they are still *rational* failings of sorts; the literature on conversation has provided ample resources for explaining how violations of conversational maxims are within the remit of rationality and normativity more generally. But epistemologists have been too quick, I'll argue, to partition these discussions off from their own – overzealous in their assumption that the normativity at play in conversational failings is non-epistemic in kind.

The view I call Quality-exceptionalism goes back at least as far as Williamson's (1996: 489) seminal work on the epistemic normativity of assertion, is endorsed explicitly in both Brown (2010: 550) and Lackey (2011: 265), though I suspect its endorsement spans further and wider.[70] Epistemologists, after all, are interested in distinctively epistemic, and not social, prudential, nor moral normativity. We should thus be careful to draw sharp lines our target domain, lest we fall prey to various, nefarious forms of encroachment. What I will argue, however, is that while we are right to restrict our discussion squarely to the epistemic domain, we have drawn the lines in the wrong places. Given familiar and well-worn resources for theorising about distinctively epistemic

[69] I set aside debate surrounding how to specify these maxims, because as noted in fn. 68, Grice's theory of conversation per se is not our concern here. Some (e.g., Montgomery 2014; Goldberg 2015b) understand the long debate regarding the epistemic norm of assertion as simply a debate about how to specify Grice's Quality maxim. Benton (2016) argues that Quality is best understood in terms of knowledge. Stotts (2024) argues that the conversational maxims are best understood as social norms. Relevance theorists (e.g. Sperber & Wilson 1986; Carston 2002) argue that the entire conversational domain is to be explained in terms of Relation. Nothing I say here turns substantively on any of those debates.

[70] Though in a slightly different context, see also Goldberg (2024: 33) who, in detailing the normative contours of a conversation, notes that while epistemic normativity 'comes into play', "… epistemic rationality and epistemic norms play […] either *subsidiary* or *derivative* functions in conversation."

normativity, I'll set out to show that these are precisely the kinds of failings that ought to shape and structure our epistemological discussion.

§3 Gricean epistemic norms

Recall the principle of Boundedness:

> **Boundedness** Our cognitive and environmental limitations make a difference to what and how creatures like us ought to think and believe.

The thrust of my argument will be that agents like us, bounded as we are, should no more epistemically regulate our practice of assertion with a norm like Quality than with ones like Quantity, Relation, and Manner.[71]

But first: what, exactly, are epistemic norms? How do we identify and individuate them? A natural starting point is as follows: epistemic norms are just norms with epistemic content. If a norm countenances epistemic notions like truth, knowledge and understanding, then, one might think, it is an epistemic norm. Unfortunately, as many have pointed out, this won't work.[72] There are in fact a myriad of norms with epistemic content that are clearly not epistemic norms (Simion 2018; Kauppinen 2018). To take a few examples:

> **Swimming norm** Do not jump in the pool unless you *know* how to swim.
> **Investment norm** Do not invest unless you *know* the company will turn a profit.
> **Religious norm** *Believe* your pastor.
> **Professional norm** Do not *assert* that your boss is bald.[73]

These norms all countenance epistemic notions, and, what's more, conforming to them places substantive demands on one's epistemic standing or epistemic activities in various ways. Clearly, however, they are not epistemic norms in the thicker sense that epistemologists have in their sights. They are not, for instance, epistemic in the way that the following norms are epistemic:

> **Evidential norm** Do not believe that for which you lack sufficient evidence.
> **Knowledge norm** One should: assert that p only if one knows that p.

Some norms thus share content with epistemic norms – they have to do with one's epistemic states and countenance epistemic notions – without thereby being epistemic. Other norms, like **Religious norm** and **Professional norm** are epistemic in the sense that they evaluate

---

[71] A bolder use of **Boundedness** would be to suggest that we can *derive* **GENs** from this principle. My claim is weaker: if we grant **Boundedness**, as many are inclined to do, then we can explain why norms pertaining to relevance, informativeness and manner have direct, epistemic import for creatures like us. On this view, it is a mistake to suppose – as much of the literature has – that such imperatives are merely conversational or pragmatic, i.e., and not also epistemic.

[72] See, in particular, Coffman (2014), Kauppinen (2018) and Simion (2018).

[73] This example is from Brown (2010), wherein it is argued that such cases demonstrate that knowledge is not *sufficient* for proper assertion. This is not my purpose in deploying the example, and in fact I disagree with Brown on this point, though for reasons that would take us too far afield. The example, like all others in this list, is just an example of a non-epistemic norm with epistemic content.

epistemically evaluable items, like assertions, and yet are no more epistemic as a result. What this shows is that only *some* norms with epistemic contents are epistemic, and only *some* norms governing epistemic activities are epistemic (Kauppinen 2018: 22). In what follows I use 'norms with epistemic content' to refer disjunctively either to norms such as **Swimming norm** and **Investment norm**, i.e., non-epistemic norms with epistemic content, or norms such as **Religious norm** and **Professional norm**, i.e., norms pertaining to epistemic activities but hailing from different normative domains.

If epistemic norms proper are not simply norms with epistemic content, then what are they? Two prominent diagnostic tools have been offered in the literature: epistemic norms render one subject to *distinctively epistemic modes of accountability*; and epistemic norms are rationalised or justified *by association with epistemic goals*. To launch my case for **GENs**, I'll argue that they fit the bill on both counts.

§3.1 Epistemic accountability

There is a tight link between epistemic norms and epistemic accountability practices.[74] Epistemic norms, unlike norms with epistemic content, render agents liable to *distinctively epistemic ways of holding accountable*, to use Kauppinen's (2018: 22) terminology. Such authors endorse some variant of the following diagnostic principle:

> **Practice Indicates Norms** If there is a legitimate practice of epistemically criticizing agents for whether and how they φ, then we have reason to think that there are epistemic norms on φ-ing (Flores & Woodard 2023: 10).[75]

On this view, we can get a grip on whether a norm is at play, and whether that norm is epistemic, by looking at our reactions when it is violated (Flores & Woodard 2023: 11). This, it is worth emphasising, should be relatively unsurprising, given the important role norms are supposed to play in interpersonal regulation more generally (*Ibid.*).[76] The immediate next question concerns how to discern the *epistemic* from the *non-epistemic* modes of criticism. Happily, those working in this area have made progress on this question in recent times. On Kauppinen's (2018) view, a distinctive way in which we hold agents epistemically accountable is by lowering their credibility score; the credence we place in their assertions is lowered, and we are less willing to partner with them in inquiry (Kauppinen 2018: 12). Boult emphasises the importance of distinctively epistemic *blaming* practices (2021; 2024). When we epistemically blame someone, for Boult, we modify our epistemic relationship with them; we alter our expectations of the agent – in particular, we alter our expectations as they pertain to their epistemic agency.

With the first diagnostic method in hand, let us return to the case of assertion. Of course, the pertinent question here is not whether we have a legitimate practice of criticising assertions – this much we already know – but rather what exactly we epistemically criticise them *for*. I'll now argue

[74] Among others, see Kauppinen (2018); Boult (2021; 2024); Flores & Woodard (2023). For wider discussion of the relationship between norms and accountability, see Darwall (2014) and Simion & Kelp (2021).
[75] Though see also Simion & Kelp (2021: 24).
[76] See also Graham's (2015) view that epistemic norms are social norms that reinforce socially beneficial epistemic behaviour.

that we hold each other epistemically accountable not only to Quality, but also to Quantity, Relation, and Manner.

Consider, by way of example, a Quality-obsessed speaker. They have as clean a reliability track-record as one could hope to have; they speak only when they are certain beyond all doubt, and never lie nor bend the truth, no matter the stakes. However, they have a different kind of assertoric quirk: they will incessantly speak off-topic, invariably give way too much or way too little information, and give nothing in the way of cues or signals when topic-changes have occurred. I submit that it would be perfectly appropriate, across a range of possible topics and conversations, to deliberately refrain from soliciting such a speaker's testimony. It would be perfectly legitimate to refrain from engaging with or taking seriously their communicative contributions throughout the process of various different kinds of inquiry.[77]

Moreover, I contend, the legitimacy of such epistemic accountability practices remains once we control for practical and ethical upshots. Suppose that there are no practical stakes whatever to the inquiry, the goal is the pure satisfaction of curiosity. And suppose that I deem you to be something of a moral saint - that is, your assertoric quirk is in no way explained by some underlying moral flaw. In such a case, there would be no practical or moral gain from holding you epistemically accountable, yet doing so remains appropriate. It would remain perfectly appropriate for me to modify our epistemic relationship by refraining from soliciting your testimony in the future. All else equal, I will solicit and take up the testimony of someone who stays on topic and calibrates their assertions to their interlocutor.

If legitimate practices of epistemic accountability are a reliable guide to the epistemic norms governing assertion, then it is an important finding that such practices do not track Quality alone. They also track standards concerning relevance, informativeness, and perspicuity. Insofar as accountability practices provide evidence of epistemic norms at work, they therefore support **GENs** over Quality-exceptionalism.

§3.2 Epistemic goal-association

The argument from epistemic accountability invites a further question. Even if our accountability practices track GENs as well as Quality, why think that those practices are themselves epistemically justified? After all, we can engage in epistemic appraisal for non-epistemic or illegitimate reasons – e.g., when I refrain from inviting you to the team meeting because it is a small room and you are

[77] As the literature since Fricker's (2007) *Epistemic Injustice* demonstrates, failing to have one's testimony solicited or taken up in this way, whether rightly or wrongly, can constitute a distinctively epistemic harm. Some recent developments and applications of the concept seem particularly relevant here. For example, Smith & Archer (2020) and Saint-Croix (2022) argue that it matters not only that agents receive the proper degree of *credibility*, but also of *attention*. In particular, they do so in part because of the role of attention in determining one's ability to influence what others believe, and shape the common ground. Hookway (2010) and Friedman (2025) explain epistemic harms in terms of participation in inquiry. Relatedly, Mallory (2026) delineates the notion of *inquisitive injustice*, which involves not merely the ability to have one's assertions believed, but rather the ability to have one's questions taken up and used to shape the discussion. Such developments are more grist for my mill: while we may not hold the Quality-obsessed speaker responsible simply by lowering credence in the truth of their assertion, we do so in a range of other ways – i.e., by not soliciting and/or taking up their testimony in the first place, by blocking it from shaping the discussion. As these authors have shown, refraining from *believing that p on your say-so* is but one of many ways in which I might reduce epistemic trust in you, or hold you epistemically accountable.

simply too tall. The mere existence of a practice of epistemic accountability, absent a story regarding that practice's rationale or justification, does not yet establish that the norms it tracks are epistemic.[78]

It matters, then, not only *that* we have epistemic accountability practices but *why* we do. Simion offers a useful diagnostic: epistemic norms are associated with epistemic goals.[79] To determine whether a norm is genuinely epistemic, we look to its rational locus. If adherence to a norm – no matter its content – is justified by its role in promoting non-epistemic ends, then it is not an epistemic norm. Return to **Swimming norm**, for example. Its enforcement, if legitimate at all, is grounded in life-preservation – a prudential or perhaps moral aim. By contrast, as I'll argue, GENs are justified by virtue of their contribution to distinctively epistemic goals. We hold ourselves and others accountable to standards such as relevance, informativeness and perspicuity because, at least in creatures like us, conformity to them promotes the realisation of our epistemic goals; violations hinder. The normative locus of these norms is not merely because of their role in ensuring smooth social interactions (though, of course, they may also help with this), but also because of their role in ensuring that when speakers assert, hearers tend to initially attain, and then subsequently retain, true beliefs and knowledge.

One may object: in all examples of merely conversational failings, the assertions were knowledgeable. To that extent, the speaker's particular role in the attainment of epistemic goals via assertion is achieved; knowledge is made available. Any grounds for criticism, then, simply cannot be made by association with epistemic goals, as Simion's test requires. At worst, certain facts about the way in which the speaker asserts may make knowledge more difficult, or cumbersome, for their interlocutor to acquire, but this speaks to a patently practical, and thus non-epistemic, standard.

This objection, however, is misplaced – for a number of reasons. A first problem is that it altogether fails to account for **OVERLOAD**, wherein Ali's speaking knowledgeably is not itself sufficient for knowledge to be made available to his interlocutor. As Ali is well aware, his brother has no reasonable chance of taking up this knowledge.[80] But even granting that the other examples

[78] This exemplifies epistemic accountability for *non-epistemic* reasons. To see why we also need to rule out epistemic accountability for *illegitimate* reasons, consider the norm *Do not believe women.* We do not have to engage in science fiction to imagine societies and contexts in which a norm like this one is in full force. The focus is thus restricted to *legitimate* practices only. Minimally, legitimate practices of epistemic criticism must track epistemic goods (Goldberg 2018; Flores & Woodard 2023). I will content myself with this minimal characterisation; my purposes here are to put this tool to work and assess the results, not to develop the view *qua* diagnostic method. Moreover, I have discussed non-epistemic as opposed to illegitimate norms in the main text because, if there is push-back to my view in what follows, I am highly doubtful that it will be because **GENs** are deemed *illegitimate* in the way that the sexist norm above is. See Goldberg (2018: 170) for a fuller account of what makes a practice legitimate.

[79] The epistemic deontologist may initially balk, but they need not: nothing here implies any substantive value-theoretic commitments. The claim here regarding epistemic goals is one of mere association. The teleologist will cash out the association with the goal prior to the norm - we follow the norm in pursuit of the goal; the deontologist will do the opposite - the goal only has value in light of the norm. In either case, the mere association holds, which is all this view countenances.

[80] Similar considerations seem to speak in favour of so-called 'audience-accommodating' norms of assertion, e.g., Pelling (2013; 2014); Hinchman (2013); García-Carpintero (2020); Simion & Kelp (2021); Gaszczyk (2024); though cf. Willard-Kyle (2021).

yield knowledge, this does not yet demonstrate that, in such cases, the evaluative standards underwriting the speaker's misgivings are thereby non-epistemic in nature.

Consider Grice's notion of *flouting*. A speaker flouts rather than violates a maxim when they cooperatively and transparently breach a maxim in order to communicate a proposition other than that which they explicitly assert. We flout maxims all the time; it is often handy to refrain from outright stating what we think. Still, with a little subtlety, we can remain communicatively cooperative. To take Grice's famous example, a philosophy professor might assert **1a** in order to communicate **1b** without having to state it outright:

> **1a** The student is very punctual, and has excellent handwriting.
> **1b** The student is not particularly good at philosophy.

The professor is in breach of Relation, but their breach is flagrant and transparent; it is neither uncooperative nor deceptive. Hence the classification of such breaches as instances of *flouting* as opposed to merely *violating*. The prevalence of flouting helps to explain why criticism of speakers in merely conversational failings, even those in which the hearer *can* take up knowledge of what is said (that is, in **EYES** and **RENT**), may nevertheless be an epistemic criticism – one made in association with epistemic goals. Note that the speaker in each of these cases does not flout, but rather straightforwardly violates, conversational maxims. They conform to Quality, but breach (at least) Relation in **EYES**, and (at least) Quantity in **RENT**. Now, we may ask whether their breach of Relation and Quantity is in fact merely a conversational, or also an epistemic problem.

I think the answer is that it is clearly at least in part an epistemic problem. As I'll show, this violation leaves two options: one leads to false beliefs; the other to wasted cognitive resources. Given Boundedness, *both* are epistemic problems. Take **EYES** as a demonstration. If Ollie falsely takes Rory to be aiming to be cooperative, they might reasonably form a range of false beliefs on the basis of Rory's assertion – i.e., pertaining to the link between the distance between Rory's eyes and the location of today's talk. Otherwise, why would Rory have included this information in this context? In such a case, the problem with Rory's assertion is straightforwardly epistemic, and not merely social.

But alternatively, suppose Ollie was a little more vigilant, wisely managing to avoid any false beliefs. Perhaps they reason through the possible links and correctly finds that there is no such link to be drawn (at least not a plausible one), and thus avoids the first epistemic risk of false beliefs. Still, given **Boundedness**, it is not hard to see why, even in such a case, criticism of Rory's assertion can be made on distinctly epistemic grounds. We do not have unlimited time and cognitive resources to devote to extracting the epistemic good from every proffered assertion, and in this case Rory's assertion placed a cognitive burden on Ollie that would have been avoided had he asserted cooperatively. The prevalence of *flouting* in cooperative communication makes clear why it is rational for Ollie to devote cognitive resources to such reasoning; speakers are often in fact flouting cooperatively and not merely violating, so when faced with a putative violation, it is perfectly rational to devote resources to charitable interpretation. But in such a case, such resources are wasted; disappointingly, Rory did not *flout* but merely *violated*.

Assertions like Rory's systematically risk either false belief or wasted cognitive resources. This undermines the thought that the only problems with such assertions can be traced to the practical or social realm. If what matters is the association between epistemic norms and epistemic goals, then **GENs** interact with epistemic goals in just the way that *bona fide* epistemic norms are supposed to.

§3.3 Twin epistemic goals

In what follows I sharpen the argument from epistemic goals in two main ways. First, in this section, I draw on the widely-acknowledged Jamesian idea of *twin epistemic goals* to refine the argument from epistemic goals; then, in the next section, I will think more closely about how epistemic goals should structure our practice of assertion in particular.

Thus far, in arguing that GENs meet the epistemic goal-association test, I've implicitly adopted a fairly rudimentary conception of those goals: roughly, the acquisition of truth and knowledge. But James (1896) famously noted that this broad, epistemic goal is in fact ambiguous between two, competing subgoals: the acquisition of truth and the avoidance of falsity.[81] If overly concerned with the former, one may be pushed toward carelessness; but equally, if overly concerned with the latter, one may be pushed toward cynicism. This well-known Jamesian insight thus gives rise to two distinct standards of epistemic evaluation: *reliability*, which speaks to the ratio of true-to-false beliefs that some system, process, or method outputs; and *productivity*, which speaks to the number of true beliefs (or amount of knowledge) it outputs.[82] Perhaps, given unlimited time and cognitive resources, reliability would be our sole consideration – all that would ultimately matter is that, for any proposition taken up, that proposition is true. But, as **Boundedness** emphasises, it is relevant that we are not in such an ideal scenario. Relative to our epistemic goals, then, we might positively epistemically evaluate some process in two importantly different ways – both pull in opposite directions, but – at least for bounded agents like us – are decidedly epistemic forms of evaluation nonetheless.[83]

Applying this Jamesian idea to assertion, note that Quality speaks only to reliablity. No matter how it is spelled out, this norm plays the same role in relation to the realisation of epistemic goals: it helps to ensure that, for any asserted proposition, that proposition is likely to be true. That is why the norm forbids false or unevidenced assertion, and that is the way in which it is grounded by association with epistemic goals. But reliability, as we've seen, is but one part of the full, epistemic picture. Alongside reliability, we might equally want our epistemic norms to govern our conduct in pursuit of productivity. Our practice of assertion, on this line, should be epistemically regulated such that speakers assert in ways conducive to productivity, and not merely reliability.

---

[81] James sometimes spoke in terms of truth, sometimes in terms of knowledge; sometimes in terms of falsehood, sometimes in terms of error, and sometimes in terms of ignorance. These differences will not matter for our purposes here.

[82] This terminology comes from Kelp et al. (2020). Goldman (1986) uses 'power' instead of 'productivity'. Again, this terminological difference will not matter.

[83] See, e.g., James (1896); Alston (1985); Goldman (1986); Foley (1987); David (2001). For the notion of Jamesian goals as it pertains to debates over testimony and assertion in particular, see Simion & Kelp (2025).

Enter Quantity, Relation, and Manner. Standards like these, I'll argue, speak to *productivity*'s side of the Jamesian epistemic balancing act. A perhaps oversimplified example will demonstrate the point. Suppose I am going to assert propositions p1, p2 … pn. From the perspective of Quality alone, there is no way to 'step outside', as it were, of the isolated, epistemic evaluation of each asserted proposition. But presumably, for at least certain sets of propositions 1-n, there are orders that make better sense than others – there are ways of presenting the propositions that foreground different connections, probe different questions, and suggest different inquiries than others.[84] Of course, none of this would matter if all we were interested in was reliability. If all that mattered was that, for any asserted proposition, that proposition is true, then connections do not matter. So long as every individual proposition is asserted on sufficiently good evidential grounds, there is nothing more to be said or done from the perspective of epistemology. But if productivity also matters, then that is no longer the case. It will be from the perspective of epistemology that we care about the order of the propositions, the way in which they are asserted, the amount of detail given, their relevance to the QUD, and so on. Such considerations play an important role in the maintenance of a healthy balance between the twin epistemic standards of reliability and productivity.

If I am right, then Quality-exceptionalism rests on a mistaken and overly narrow conception of the role that epistemic norms can play in governing assertion from the epistemic point of view. The sole role of epistemic norms of assertion, on this conception, is that of ensuring reliability; of ensuring that, for any proffered assertion, that assertion is sufficiently likely to be true (or perhaps: sufficiently unlikely to be false). However, as we've seen, this is to mischaracterise the different roles that epistemic norms can play – the importantly different ways in which they can govern our conduct toward the realisation of epistemic goals. Once we have a more accurate conception of the role of epistemic norms on the table, we see that Quantity, Relation, and Manner, all plausibly speak to decidedly epistemic standards of evaluation. In particular, they speak to the kinds of *productivity*- as opposed to *reliability*-maintaining measures that are involved in a fuller conception of epistemic evaluation.

§3.4 The functional epistemic practice of assertion

So far, I've argued that GENs are epistemic for three, related reasons: we hold each other epistemically accountable with respect to them; they are rationalised by epistemic goals on a simple conception of those goals; and they are rationalised by epistemic goals on a more sophisticated refinement of that conception. Now, before considering some objections to the view, I'll think more closely about how epistemic goals are supposed to shape and structure our practice of assertion in particular.

[84] It is nowadays quite common to emphasise the epistemic importance of dimensions of speech and language that extend beyond the traditional focus on individual propositions, reliability, and truth. Features pertaining to the mode of presentation, linguistic ordering, and perspectival effects – particularly those pertaining to the distribution of salience and attention – are noteworthy recent examples. For various ways of developing these ideas, see Anderson (1995); Camp (2006; 2017); Fraser (2021); Siegel (2022); Whiteley (2023; 2024); Munton (2021; 2022); El Shazly (forthcoming). The productivity line I sketch here is but possible one way of illuminating their epistemic importance. Another, related line is explored in §3.4.

Why, we may ask, is our practice of assertion a site of epistemic normativity in the first place? That it is in fact such a site is sufficiently indisputable as to make the question sound misplaced. But this indisputability can mask important peculiarities about the precise *way in which* assertion is epistemically evaluable. For one thing, most familiar epistemic norms pertain to mental states (paradigmatically: doxastic states) rather than actions. Pausing to consider why and in what exact way epistemic normativity so uncontroversially extends into the domain of action – at least for this particular action – is thus a valuable exercise, especially when considering which specific norms this extension yields. Secondly, there is, at least on the face of it, a comparatively *immediate* or *intrinsic* story to be told about the epistemic evaluability of mental states such as belief – one that is not available for assertion. For example, beliefs can realise epistemic value in and of themselves; they can constitute knowledge. This tight relationship with knowledge legitimises epistemic constraints on the way that beliefs are formed – it licenses norms such as *believe only on sufficient evidence*, and so on.

But the same kind of story cannot be told for assertion. Assertion, when all goes well, does not itself constitute knowledge. At best, assertion produces – or is at least *apt to produce* – knowledge in others. If it weren't for our social-epistemic need to disseminate knowledge among communities, then there is no need for any epistemic constraints on assertion whatsoever; not even Quality-style constraints. The epistemic evaluability of assertion is thus importantly different from that of belief; it is essentially relational and practice-mediated, dependent on its role in a broader social-epistemic practice of knowledge dissemination. Thus, while it is uncontroversial that epistemic normativity does in fact get its teeth in here, it is crucial to note that it does so from a different angle. If we model the epistemic normativity of assertion too closely on that of our private, mental states, in search of an intrinsic or belief-like account, we risk ruling out not only **GENs** but Quality, too.

Once this difference between belief and assertion is in clear view, the natural question becomes how, exactly, our practice of assertion realises epistemic value given that it does not do so intrinsically. The most widely endorsed answer – both in epistemology and the philosophy of language – is some version of the following view: assertion is a site of epistemic normativity by virtue of its epistemic *function*, *aim* or *telos*.[85] We assert in order to transmit knowledge to hearers. That is why we have the practice in the first place; if we had no need for second-hand knowledge, we would have no need for the practice of assertion. What I will now argue, however, is that this widely popular functionalist rationale for the epistemic normativity of assertion in fact yields more than just a Quality maxim. If we ask which norms govern the practice of assertion from the epistemic point of view – where this is here understood as something like 'facilitates epistemic function-fulfilment' – then Quality is but one part of the picture.

To see why, note that a practice of assertion regulated only for Quality is not yet fulfilling its function of knowledge-transmission. A practice regulated only in terms of Quality is one in which speakers speak knowledgeably, which is a good start. But **Boundedness** makes clear why it is just that: a good start, not the full picture. In an ideal world, wherein the agents engaging in the practice

[85] See, e.g., Austin (1962); Lewis (1969); Stalnaker (1978); Craig (1990); Williams (2002); Garcia-Carpintero (2020); Simion & Kelp (2021). Of course, how exactly to think about this function, aim, or telos, is anything but uncontroversial. It will suffice for my purposes here to run with the general, widely-accepted idea that assertion has an epistemic function, however it is to be theorised.

are limitless in ways that we are not, Quality might be just fine. After all, in such a world we have unlimited time and cognitive resources to take up any putative piece of knowledge in our environment, if we so choose, including those pieces that are put there by speakers' assertions. The only hurdle to function-fulfilment in such a world is that speakers say the wrong thing – i.e. something false or unjustified – because there are no hurdles on the hearer's side. If it is there, we will take it up.

Of course, as already noted, this is not in fact our situation. Hearers' time and cognitive resources are limited in substantive and normatively relevant ways. For creatures like us, we need further norms governing our practice of assertion from the perspective of epistemic function-fulfilment. In particular, we need further norms that speak to the process of *uptake*; function-fulfilment occurs not merely when knowledge is produced by way of assertion, but when it is then subsequently taken up by a hearer.[86] We may then ask which further norms speak to this side of things. And I think it is clear that, in this regard, **GENs** fit the bill. Function-fulfilment in this domain requires norms that speak to the kinds of information creatures like us can handle. Given our limitations, it is no guarantee that produced-knowledge amounts to transmitted-knowledge. But note that this is precisely what norms such as Relation, Manner, and Quantity speak to. They ensure that the assertion is not only knowledgeable, but that it is of the kind that a cooperative hearer is likely to be able to take up, both *quantitatively* (in the case of Quantity) and *qualitatively* (in the case of Relation and Manner).

To illustrate, note that Quantity places an importantly different kind of constraint on assertion than does Quality. It commands a speaker not only to speak knowledgeably, but to give the right *amount* of information, where 'the right amount' is indexed to the QUD, or the established common ground, and the epistemic goal of knowledge-transmission within those parameters. It builds into its evaluative standard the idea that, even holding the exchanged proposition fixed, there are certain levels of detail that are more or less conducive to epistemic progress within the exchange than others. We can of course debate how exactly to incorporate such evaluations, and weigh them against others, but the point is clear: epistemology need not be silent on these standards simply because they do not straightforwardly bear on the question of whether the assertion is itself knowledgeable. There are other evaluations of assertions that we would and should want to make, given the epistemic function of knowledge-transmission.

And the same considerations apply, I submit, to the qualitative dimensions that Relation and Manner represent. These speak not to *amount* of information likely to enjoy uptake and thereby contribute to knowledge-transmission; rather, they speak to the *types* of information likely to enjoy such uptake. Our bounded minds draw links between things; we do not merely take up every asserted proposition in isolation. We see connections, construct mental models, deploy heuristics, and so on, and we do so in order to better cognise and retain information about the world. It follows, then, that we are not and should not be indifferent to the *order in which* information is presented to us, the *links drawn* (whether explicitly or implicitly) between propositions in a

[86] This, too, is in-keeping with the rationale for *audience*-accommodating norms of assertion mentioned in fn. 80.

discourse, and so on.[87] But Quality is silent on such matters; it asks, for whatever proposition was asserted, only whether the speaker's epistemic standing with respect to said proposition was sufficiently solid. To fully account for function-fulfilment, however, we need evaluations that step outside of such isolated evaluations of individual assertions, and this is precisely what **GENs** do.

It is important to be clear on what does and does not follow from this functionalist rationale. One might think that even if assertion is epistemically evaluable in virtue of its role in a social epistemic practice, a Quality-style norm suffices to govern that practice from the epistemic point of view. On this view, all that is required for epistemic success is that what is asserted is known; any further considerations concerning uptake, relevance, or intelligibility concern at best the efficiency of communication.

But this line mischaracterises what it is for assertion to fulfil its epistemic function. Once we more accurately identify precisely where epistemic normativity and assertion interact – namely at the level of the social, functional practice – we see that the relevant normative standard for such a practice is not merely *proffered knowledge*, but also its subsequent *uptake*. It is this process, considered as a whole, whose success represents the gold epistemic standard by which this practice is epistemically evaluated. A knowledgeable assertion that is wholly irrelevant to the inquiry at hand, pitched at the wrong level of detail, or needlessly obscure, may in some attenuated sense make knowledge available, yet fail to realise the epistemic function in virtue of which assertion is norm-governed at all. Quality is therefore epistemically indispensable yet epistemically incomplete: it constrains *what is asserted* but is silent on whether what is asserted is successfully transmitted. If epistemic normativity gets its teeth into assertion at the level of the practice, then it is the whole process that matters.

In sum, it is very plausible to think about assertion's epistemic normativity in functionalist terms, and it is very plausible to think about its function in terms of knowledge-transmission. Once you do this, however, you do not merely get Quality but also **GENs**.

§3.5 Assertion and understanding

Epistemic goals have played a central role in the argument so far. But in all cases we have deployed a fairly narrow conception of what the relevant epistemic goals are. In §§3.2-3.4 it was assumed that our main epistemic goal is, roughly, that of knowledge, or perhaps true belief. I gave three different arguments for thinking that, on that broad knowledge-oriented picture, **GENs** hook up to the practice of assertion in the ways we'd expect if they were *bona fide* epistemic norms. Now, in my final two arguments, I abandon the simple knowledge-centric conception of epistemic goals. Moving beyond the simple picture we've adopted so far, new avenues to **GENs** come to the fore.

[87] Particularly relevant here is Ella Whiteley's (2023; 2024) work on *order-based salience patterns in language*, wherein she highlights that even the most seemingly inane features of our language – such as the order of words in a sentence – can have profound and normatively significant effects with respect to the way we come to think about the topic at hand. See also Rachel Fraser's (2021) notion of *narrative testimony*, wherein she details the distinct epistemic (and *perspectival*, in Camp's (2006; 2017) sense) effects of narratively structured testimony. Even the same basic propositions, she argues, can have vastly different epistemic effects depending on their role and place in a structured discourse.

Perhaps the most natural and obvious departure from the simple knowledge-centric picture we have adopted thus far countenances the epistemic goal of *understanding*. So far I've argued that **GENs** - in a number of different ways - help us to regulate our practice of assertion in pursuit of knowledge, and, to that extent, are epistemic. But, very plausibly, it matters not only that we epistemically regulate our practices in pursuit of knowledge, but also in pursuit of understanding. In simple terms, we (epistemically) ought to assert in ways that not only bring others to know but also to understand.

Now, the literature on the epistemology of understanding is long and complex, as is the subliterature on *testimonial* understanding. Happily, however, we can sidestep most of the controversy, for the point to which I will draw our attention is amenable to multiple, competing ways of filling out the various details. Whether or not testimonial understanding is possible, and whether or not understanding is itself epistemically distinct (i.e., from knowledge), what I will show is that regulating our practice of assertion in accordance with the epistemic goal of understanding requires further norms than Quality. On perhaps the most natural and straightforward way of expanding our conception of the epistemic goals that shape and structure our epistemic norms, then, **GENs** are a good fit.

How ought we regulate the practice of assertion, if what we want from it is that hearers come not only to *know* but, at least sometimes, also to *understand*? There is surprisingly little work done on this question, and this is despite the fact that there has been an abundance of work on the possibility of transmitting or communicating understanding via testimony. But whether or not you think it possible to directly transmit understanding via testimony, it is an interesting question in its own right whether the epistemic value of understanding places distinct normative constraints on our practice of assertion.

What, beyond knowledge, is required for understanding? On most if not all accounts, what I'll call *connectedness* plays a crucial role. When one understands, one does not merely know all the relevant facts about some object or phenomenon, but one also *grasps* how those facts connect to one another - how they interrelate in explanation. Now, the notion of grasping, and its role in understanding, is controversial. Some think of grasping as an entirely non-propositional standing; to grasp a phenomenon in the way that is constitutive of understanding is to *see* how things fit together and interrelate. But even for reductionists, who think that understanding ultimately reduces to knowledge, the relevant kind of knowledge one needs in order to understand is knowledge *of connections* (Kelp 2021a). One must not only have knowledge of facts, but also knowledge of how these facts hang together in explanation. The notion of connectedness, then, whether mediated by a non-propositional notion of grasping or by knowledge of connections, is one of the crucial ingredients of understanding.

We are now in a position to see why incorporating the epistemic goal of understanding delivers further norms on assertion than Quality. Recall that, as we saw in the previous sections, Quality is a fundamentally isolated standard of evaluation, meaning that it exclusively evaluates individual asserted propositions. This caused problems - detailed in the last section - for accommodating the epistemic value of productivity, and it causes problems here, too, for accommodating the epistemic value of understanding. Even on the most reductive accounts, understanding requires not merely

more knowledge but knowledge that is interconnected in the right way. Connectedness evaluations, like productivity evaluations, pertain not to isolated propositions but rather to *bodies* of propositions. Quality simply does not provide the resources for what I am calling a connectedness evaluation. Quality, no matter how it is spelled out, evaluates the speaker's epistemic standing with respect to the asserted propositional contents; it is met when speakers enjoy the appropriate epistemic standing for that which they assert, and violated when they lack it.

If, however, we want to evaluate assertions not only for whether they bring about knowledge of what is said but also whether they bring about understanding, then such Quality-style norms will not do. One might score perfectly well on Quality-style norms - asserting only what one has unimpeachable epistemic standing for - all day long and yet never produce anything that is connected in a way that conduces to understanding. We may oversimplify for a moment to demonstrate the point. On a reductive account of understanding the rough picture is as follows. There is a body of propositions p1-pn that constitute the basic facts regarding the target phenomenon P, then a separate body of propositions c1-cn that constitute the list of relevant connections between those facts. If I know p1-pn *and* c1-cn then I understand P. From the perspective of Quality, there is no epistemic problem whatsoever with randomly and disconnectedly asserting any true proposition regarding P. All that matters is that, when you assert, you do so with sufficiently good epistemic basis for that which you assert. The norm simply doesn't have the scope to explain the role that c- as opposed to p-propositions play in understanding. If we care about understanding, then even on a reductive account - i.e., the kind of account that represents the smallest departure from the traditional knowledge-centric picture - we'll need further norms in order to evaluate assertions by reference to the bodies of knowledge they contribute to, the connectedness among those bodies, and whether they amount to understanding.

The immediate question that arises, then, has to do with which specific kinds of norms we *do* need to make connectedness - as opposed to isolated - epistemic evaluations of assertions. And here, I think, **GENs** are of the ilk we'd need. I will refrain from saying too much here, because I think that if I'm right, and it turns out that **GENs** do indeed have unappreciated epistemic import, then we need to do significant work thinking about how exactly to specify them. This is not work I have space to even begin in this article. However, what I can show is that, structurally, they fit the bill. They make the kinds of evaluations we'd want to make if we were in the business of evaluating assertions for conduciveness not only to knowledge but also to understanding.

Notice that Quantity, Relation, and Manner, are all anchored to the conversational context. The evaluations they trade in are importantly different from Quality insofar as they are dependent on not only the contents of what is asserted but also the QUD (Roberts 1996; 2018), the common ground (Stalnaker 1978), or the 'accepted purposes of the talk exchange', to use Grice's (1975; 1989) own preferred terminology. One cannot evaluate an assertion as relevant or irrelevant in and of itself; it is adjudged relevant or irrelevant relative to the topic at hand. Same goes for specifying what just the right amount of quantity looks like - what constitutes too much or too little. Whether the norm is met or violated will depend, in part, on the conversational context. Now, at least when the topic at hand is something that can be the proper object not only of knowledge but also of understanding, we see clearly why these further norms - speaking to the relevance, quantity and clarity of information given - will have epistemic, and not merely conversational import. These are

the kinds of norms that track whether what one asserts is conducive to the hearer's coming to understand.

As noted, there is much further work to be done in thinking about what exactly the epistemic versions Quantity, Relation, and Manner will amount to. My purpose here, however, has just been to point out that the kinds of standards they speak to make them apt for tracking connectedness in a way that Quality-style norms could not. To that extent, **GENs** look like the kinds of norms we'd epistemically govern assertion with were we concerned not merely with knowledge but also with understanding.

§3.6 Assertion and inquiry

A recent turn in epistemology provides the makings of one final argument for **GENs**, or so I will now argue. Traditional epistemology has focussed primarily on one aspect of our conduct: our beliefs and belief-like attitudes. The core locus of epistemic evaluation, as Friedman (2019) puts it, is the doxastic; whether our beliefs are justified, coherent, well-evidenced, and so on. In recent times, however, there has been a surge of resistance against this doxastic paradigm. Increasingly many epistemologists have argued that whilst belief is important, it is but one aspect of our epistemic lives. Epistemology, properly construed, should be no more concerned with our practice of forming and maintaining beliefs than it should be with that of distributing attention, of assigning salience, of gathering evidence, of taking up and pursuing lines of inquiry.[88]

It is worth noting that this epistemological turn is not entirely independent from our principle of **Boundedness**. After all, we epistemic agents are limited, both in the time that we have to devote to particular inquiries, and the cognitive resources we put to use throughout the process. Given this, it is very clear why it matters not only what beliefs we end up with - how we respond to the evidence that we currently have - but also how we restrict and distribute the limited cognitive resources that we have. If my mind and memory were unlimited in ways they are not, it may not matter how exactly I restrict my focus in the process of inquiry - which questions I take up and the decisions I make in strategizing about how best to pursue those questions. But given that I am in fact limited in these ways, epistemology need not and plausibly should not be silent on these elements of inquiry.

Now, to see our final argument, the argument from inquiry, let us think about how this recent turn in epistemology, if taken up, would manifest itself in theorising about the epistemic normativity of assertion. One important difference would be the displacement of belief at the normative locus. Just as epistemology in general ought not to focus exclusively on belief, at least according to proponents of the zetetic turn, neither should the epistemic normativity of assertion focus exclusively on the production of beliefs via assertion. Insofar as assertions are within the remit of epistemic evaluation, it is not necessarily due to their effects on hearers' beliefs, but rather, more generally, their effects on hearers' inquiries. In a slogan: asserters influence hearers' inquiries, not

[88] Some examples include the attention literature (Watzl 2022; Wu 2024; Saint-Croix forthcoming; Gardiner 2022), the salience literature (Munton 2021; Whiteley 2022), the inquiry literature (Friedman 2019, 2020; Kelp 2018b, 2021a, 2021b; Flores & Woodard 2023) and the narrative and perspectives literature (Camp 2017, 2019; Fraser 2021).

just their beliefs. Specifically doxastic effects of assertions need not take centre stage, given a zetetic approach.

But once we take up this vantage point in thinking about the epistemic norms that govern assertion, it is hard to see why Quality-style norms would be all that we need. Quality-exceptionalism, in other words, makes sense only if the doxastic paradigm is assumed to be operative, and the zetetic turn is presumed to be a non-starter from the outset. But to the extent that the zetetic turn is of increasing popularity, Quality-exceptionalism, is in trouble. On the zetetic turn, Quality polices but one of many epistemic dimensions of assertion. In particular, it polices the impact of assertions solely on the closure of inquiry: the cases in which some hearer H, on the basis of S's assertion that p, closes inquiry into whether p. If you think that all that matters, epistemically speaking, is our beliefs, then it follows that the epistemic import of assertion will centre around its impact on hearers' beliefs. But if you drop this assumption, as many do, then it is not clear why the epistemic import of assertions need necessarily focus on their impact on hearers' beliefs.

Suppose you assert in ways that systematically, predictably lead me to gather the kind of evidence that I do, and ignore other kinds of evidence. Or suppose that you assert only information relevant to X rather than Y element of the phenomenon under discussion. Or suppose you assert in a way that entirely co-opts my focus and makes salient only one particular aspect of said phenomenon. On the doxastic paradigm, these are not the kinds of evaluations of assertions that epistemology is in the business of making. Epistemology within this tradition cares about your doxastic response to the evidence you have, and doesn't particularly care about how you got that evidence. But the same is not necessarily true on the zetetic turn. If you co-opt my attention in this rather than that way, if you shift my focus to this as opposed to that question, if you emphasise these as opposed to those features, then your assertion has concretely influenced my inquiry, even if it hasn't straightforwardly changed the justificatory status of any resulting assertion-based beliefs.

In sum, a zetetic approach to epistemology has become increasingly hard to ignore. But from the inquiry-based perspective, it is hard to see why the epistemic branch of norms governing assertion would start and end with Quality-style norms. Evaluations of assertions as relevant or irrelevant to a topic or question, or as too much or too little information relative to the hearer's capacities, and so on, look straightforwardly like epistemic modes of evaluation from the zetetic point of view.[89] How exactly to specify and define these epistemic version of Quantity, Relation, and Manner, will be a thorny matter for another day; after all, everyone is more or less in agreement that Quality corresponds to a bona fide epistemic norm, and yet there is anything but agreement over how it is to be specified. The same will doubtless be true for the other norms. What is important is that the zetetic turn in epistemology represents another, independent line of argument in favour of facing up to these difficult questions. We need to face up to these questions regarding how best to specify and define these epistemic norms because it is no longer clear that they can be comfortably eschewed to normative domains next door. They track zetetic even if not

[89] The claim that epistemic evaluation is question-relative is near-consensus in zetetic epistemology: inquiries are individuated by their focal question, and the norms governing them are oriented toward its closure. See, e.g., Friedman (2019; 2020); Kelp (2018b; 2021b); Haziza (2023); Guilielmo (2024); Falbo (2025); Willard-Kyle et al. (2025).

traditional, doxastic evaluative standards, and increasingly many people are inclined to a view that would render them thereby *epistemic*.

§4 Objections and replies

I have argued that on various, plausible ways of thinking about both epistemic normativity in general and our practice of assertion in particular, Quality-exceptionalism is false; assertion is epistemically governed not only Quality, but also Quantity, Relation and Manner. Let's remind ourselves of the maxims:

> ***Quantity**** Make one's contribution no more or less informative than is required for the current purposes of the talk exchange.
> ***Quality**** Do not say either what you believe to be false or lack sufficient evidence for.
> ***Relation**** Be relevant.
> ***Manner**** Be perspicuous; avoid unnecessary obscurity or ambiguity.

The main purpose of this paper has been to argue for the epistemic import-signalling * to be added to Quantity, Relation, and Manner. All else equal, an assertion epistemically ought to be (i) *no more or less informative than required;* (ii) *true* or *knowledgeable*; (iii) *relevant*; and (iv) *perspicuous*. Given various popular and well-established ways of thinking more generally about epistemic norms and their interaction with the practice of assertion, we should refrain from drawing a sharp line around the Quality-style epistemic norms, on the one hand, and the conversational, non-epistemic ones, on the other.

On my proposed way of redrawing the normative terrain, several neighbouring debates come into view. For instance, recent work has developed Gricean norms at the level of inquiry itself, articulating individual-level zetetic norms and virtues governing the pursuit and use of information (e.g., Flamini 2025). Even if this is right, however, it is a further and non-trivial question whether and in what way such norms carry over to assertion. The relationship between the epistemic normativity of assertion and that of individual action or belief is, after all, itself hotly contested; it cannot simply be assumed that the norms that govern our individual conduct in inquiry also automatically govern speech acts. Relatedly, some (e.g., Carter 2024; Rimoldi 2024) have advocated for modelling conversation on shared inquiry – in the spirit of Gilbert (2009) and Bratman's (2014) theories of joint action. Yet even if conversation can be modelled as joint inquiry, it is far from clear that this could ground epistemic norms of assertion like GENs. After all, while some conversations may fit this model (i.e., explicitly cooperative, inquiry-directed ones), many conversations do not.[90] Oftentimes conversations take place within hostile, non-cooperative, and otherwise non-ideal contexts.[91] Finally, there is live disagreement over the goal of conversation

[90] An additional complication concerns the extent to which Grice's theory presupposes idealised cooperative environments. See, e.g., Harnish (1991) [1997]: 357 fn. 29; Grice (1989: 368-370); Stokke (2016); Cappelen and Dever (2021); Berstler (2025) for discussion. The present proposal does not depend on conversation approximating an idealised cooperative setting. If anything, the epistemic import of the maxims, as I've argued here, derives in part precisely from non-ideal agents and environments – most notably, the kind of cognitive and environmental boundedness discussed above.

[91] For illuminating discussion on this point, see, among others: Saul (2018); Camp (2018); Peet (2021); Ellefsen (2021); and various chapters in Khoo & Sterken (2021).

itself: some take it to centre on knowledge exchange (e.g., Szabó 2020), others on coordinated cognitive focus (e.g., Dickie 2020). The revisionary picture I have advocated for regarding the relationship between epistemic and conversational normativity may be more or less surprising depending on whether and in what sense the goal of conversation should be understood as an epistemic one.

These views all raise important questions about how assertion, conversation and inquiry interrelate. The claim defended here is simply that, whatever the ultimate story regarding inquiry and conversation, all Gricean conversational maxims govern assertion from the epistemic point of view. Fully disentangling how these views interact with the present thesis would require significant further work. I mention them only to register that, while closely related, their bearing on my argument is neither straightforward nor decisive.

§4.1 Overinflation

With the view now clearly on the table, I turn to some of its most pressing objections.

One might worry that the considerations marshalled in favour of Gricean epistemic norms prove too much. The arguments from epistemic accountability, goals, and function all appear to generalise in ways that would support a wide range of further norms governing assertion. Norms such as *assert only when well-rested*, *assert only when fed*, or *assert only after one's morning coffee* might also license epistemic accountability, or ensure or conduce to the realisation of epistemic goals and the fulfilment of epistemic function. If such considerations suffice to establish that norms governing relevance, quantity, and manner are epistemic, then it is difficult to see why these so-called 'sandwich norms' should not equally count as epistemic.[92] Since this would amount to an implausible overinflation of the epistemic domain, the natural conclusion is that the conception of epistemic normativity at work here – that is, the conception that rules *in* **GENs** - is too permissive. This worry is particularly pressing as it seems to arise for each of the argumentative routes taken in §3. Here I argue that appearances are misleading. Once we attend carefully to the precise role these arguments are intended to play, we see that my view yields the correct predictions: **GENs** are epistemic; sandwich norms are not.

This worry can be met by carefully distinguishing norms that constitute epistemic evaluations from conditions that promote epistemic performance. To bring the distinction into clear view, let's look more closely at our epistemic accountability practices. My argument was: to the extent that epistemic accountability practices are supposed to function as a window to the epistemic realm – as tell-tale signs of the substantive norms that underlie them – they support **GENs**. Assertions are epistemically criticisable not just by virtue of breaching Quality, but also Quantity, Relation, and Manner, at least all else equal. But the same is not so for assertions that breach sandwich norms. To see this, let us grant that assertions made prior to one's morning coffee are indeed criticisable. Notice, however, that the precise locus of epistemic criticism in such cases is not the lack of coffee itself but rather the faulty assertion that followed. Absent any faulty assertion, the criticism is not

[92] This name comes from the inquiry literature, wherein a version of this objection has been much-discussed. See Singer (2019); Thorstad (2021; 2022), Singer & Aronowitz (2022), Falbo (2023) and Flores & Woodard (2023) for discussion.

appropriate. **GEN** violations are importantly different. In my cases, the criticism is licensed by the irrelevant, unnecessarily complex or unclear assertion itself. Once an irrelevant assertion is made, we need not wait around to find out whether it leads to downstream Quality violations, too; epistemic criticism is rendered appropriate by the irrelevant assertion itself. In this way, my view does not overinflate on grounds of epistemic accountability. The line I deployed yields the correct prediction: **GENs** are in and of themselves standards for epistemically evaluating assertions, whereas sandwich norms are at best conditions under which they tend to meet such standards. A lack of coffee can explain epistemically bad assertion; irrelevance can constitute it.

The same distinction clarifies the role of epistemic *goals* and *function* in the foregoing arguments. Considerations about epistemic goals and function explain why relevance, clarity, and informativeness represent genuinely epistemic standards. But not every condition that promotes epistemic success thereby constitutes such a standard. Sandwich-style norms may facilitate the realisation of epistemic goals; GENs, by contrast, specify the standards by reference to which assertions are themselves epistemically assessed.

§4.2 Centrality

Another worry runs as follows. For any activity, there are certain norms one must follow *given our limitations*, and others that one must follow *regardless*. It looks as though Quality, and not **GENs**, fall in the latter camp. **GENs**, at best, are norms governing the human practice of making assertions; whereas Quality governs the practice itself. In this way, Quality has special status in the normativity of assertion.

First, a clarificatory point in response. I have not argued that **GENs** have *equal epistemic status* with Quality, and there is no general requirement that all epistemic norms are equal in this way. Indeed, I have purposefully left open the question of how to balance these norms against one another. For all I've said, it may well be that Quality does have special status in the epistemic normativity of assertion – perhaps even for precisely the reasons the objection suggests.

The difference between my view and that of the Quality-exceptionalist is that putatively conversational maxims in fact have *direct, epistemic* import: they govern assertion from the epistemic point of view. If I am right, the epistemic normativity of assertion is some function of the assertion's performance across multiple maxims, not just Quality. Crucially, however, this claim is compatible with a range of ways of spelling out that function. It may be spelled out *hierarchically* such that one maxim (plausibly Quality) must first be satisfied for the others to come into play. Alternatively, we might instead opt for a *threshold* view, according to which each maxim must clear some minimum threshold before further improvements matter epistemically. Or the standards may be *interactionist*, insofar as whether some maxim is met – or what it takes to meet it – depends at least in part on performance on other maxims. I will not attempt to adjudicate between these options here. The point is simply that once Quality-exceptionalism is rejected, questions about how these maxims interact arise *within* epistemology itself, rather than across distinct normative domains.

That I have not proposed a precise formula for balancing **GENs** does not undermine the significance of this result. As Goldman pointed (albeit in a different context), we may take comfort from law and ethics, where noting that there are competing values to be balanced (individual rights with social justice, for example) does not necessarily mean that there is a neat or precise formula for doing so. (Goldman 1986: 125) The same is true here: the absence of a tidy balancing rule fails to support the claim that these constraints on assertion belong to fundamentally different normative domains.

The onus is thus on my opponent to demonstrate not that there is a difference in *degree* between Quality and the other maxims, but rather a difference in *kind*. More specifically, a difference in kind that can help salvage the traditional Quality-exceptionalist view, according to which these norms are of altogether different kinds by virtue of their different normative statuses. The plausible suggestion that they ought to be weighted differently does not achieve this task.

§4.3 Categoricity

One way of developing the objection might go as follows: epistemic norms hold categorically, all across the modal space; **GENs** do not. **GENs** thus differ from genuine epistemic norms not merely in *degree*, or *strength*, but rather in *kind*.

We can (and should) agree that epistemic norms are not contingent on the particular goals, beliefs, desires and preferences of individual agents. In that sense, they are categorical. However, it is not true that they are categorical in the stronger sense that they are entirely independent of facts about humans as a species and the environments – social and epistemic – in which their knowledge-acquiring faculties and practices operate. Suppose that the world changes such that process X, which throughout history has systematically produced knowledge in creatures like us, no longer does so. Many will agree that such changes can alter the normative, epistemic landscape. Given this change in our environment, broadly construed, it is no longer true that we epistemically ought to form beliefs that way. And we can imagine various different iterations and tweaks on the same broad idea – as new technologies, new ways of accessing and disseminating information, and new modes of communication, all become available.

There is thus space here to agree with the objection in spirit, whilst, equally, allowing that epistemic norms often arise from and remain beholden to our environment. Our social environment is no different. The sense in which GENs are contingent is that they are derived from facts about the ways in which creatures like us interact, retain information, attain knowledge and understanding from communication with others – and these are simply facts about our social environment. Far from being subject to the whimsical preferences and idiosyncrasies of individual agents, these norms derive from the ways that creatures like us systematically and stably bring about epistemic value via our engagement in and regulation of the social, epistemic practice of assertion.

§4.4 Consequentialism

One final objection is worth considering. Given that **GENs** have been motivated in part by appeal to epistemic goals and epistemic function, one might worry that they commit us to a controversial form of epistemic consequentialism.

Two points are worth making in response. First, it may be true that, at a high level of abstraction, the rationale for these norms is indeed consequentialist in nature. But then, at that same level of abstraction, consequentialism is not particularly objectionable. Much of the pertinence of the analogy to consequentialism in ethics is that it would indeed be highly implausible if one's epistemological view resembled an epistemic version of *act*-consequentialism. It cannot be that what makes some state or action epistemically good is that it brings about good epistemic consequences. But if the rationale for **GENs** is consequentialist at all, it is *rule*- rather than *act*-consequentialist. Given my arguments, it is at most consequentialist in the sense that these are the rules creatures like us should regulate assertion with in order to best realise epistemic value via that practice. Now, with the particular sense in which the view is consequentialist now clearly on the table, the objection loses much of its bite. For this kind of consequentialism is the structure shared by many, widely popular views in contemporary epistemology.[93] At worst, then, the view here espoused is merely a companion in guilt on this charge with such views.

The second point worth making in response, this time more specific to assertion, is that if consequentialism is to be palatable and benign anywhere within the epistemic realm, the domain of assertion is precisely where it should be. The reason for this is that, as noted in §3.4, while it is obviously true that assertion is epistemically evaluable, the precise way in which it is so is itself less intrinsic – and, I submit, at least *more* consequentialist – than other sites of epistemic normativity. To see why, note that (again, as already noted), at least on a wide range of common ways to theorise the epistemic normativity of assertion, both in epistemology and the philosophy of language, the practice is epistemically evaluable broadly owing to its epistemic *function* or *aim*. This function, as we saw, however exactly it was construed, made reference to effects (or likely effects) in hearers. In other words, assertions are epistemically evaluable not because knowledgeable assertions in and of themselves realise epistemic value in the way that, for instance, you might think knowledgeable beliefs do. Instead, it is due to the inextricable link that such assertions have with the realisation of epistemic value in their consumers. I should assert in this rather than that way, on this line, in part due to the systematic *effects* of my so doing on the resulting mental states of my hearer(s). I mention this only to point out that such a rationale, which is a common and widely endorsed way of theorising assertion, is already a great deal more consequentialist in structure than, say, views on the epistemic normativity of belief.

GENs, in sum, if consequentialist at all, are so only in a way that is both (i) common in epistemology in general; and (ii) to be expected given that they are norms *of assertion* in particular.

§5 Conclusion

I've argued that our practice of assertion epistemically regulated not only by a Quality-style maxim but also by **GENs**. To that extent, a range of norms hitherto shunned to the non-epistemic

[93] Among others, this structure is shared by versions of virtue epistemology (e.g., Sosa 1991; Zagzebski 1996; Greco 2010; Carter 2016; Carter & Willard-Kyle 2025), reliabilism (e.g. Goldman 1979), zetetic epistemology (e.g., Hookway 2006; Friedman 2019, 2020; Kelp 2018b, 2021a, 2021b), proper functionalism (e.g., Graham 2012, 2014; Simion 2019; Simion and Kelp 2021), epistemic utility theory (e.g., Pettigrew 2022), and feasibilism (e.g., Lasonen-Aarnio 2024; 2025). For sceptical reaction to any such structure, see Firth 1998 [1978]; Kelly 2003; Berker 2013.

normative domain of conversation are, in fact, epistemic after all. If I'm right, this is no mere terminological dispute; it concretely alters the normative landscape as it pertains to our practice of assertion. We have to re-think the boundary between *epistemic* and *pragmatic* normativity – the relationship is far closer and more tightly intertwined than we had supposed.

Chapter IV.

Trust, Reasons, and Rational Control: A Defence of the Belief-like View

Abstract: Trust is deeply and unquestionably normative: we trust for reasons, revise trust in light of countervailing evidence, and hold one another accountable when trust is misplaced or illegitimately withheld. Yet the nature of trust's relationship to reasons and rational control is sharply contested – on belief-like views it is beyond the realm of voluntary control and answers only to epistemic reasons; on action-like views, by contrast, it can be directly willed for non-epistemic reasons. The debate has centred on cases of therapeutic trust, in which trusting someone despite decisive counterevidence appears to be precisely what moral or practical reasons can warrant. The result is a dialectical impasse: intuitions pull both ways, and verdicts tend to track antecedent commitments about trust's nature rather than illuminating it. This article aims to move the debate beyond that impasse. Drawing on theory-neutral constraints on reasons-responsive and norm-governed attitudes, I present a novel case for the belief-like view, one that refrains from relying on theoretically heavyweight commitments or contested intuitions. But a central contribution lies in the following, emergent upshot of that argument: belief-like views have been thought unable to account for trust's characteristic agency – the sense that we are not merely passive in whom we trust. By appeal to *attitudinal*, as opposed to *volitional*, rational control, I show that a belief-like view can give a satisfying account of the agency involved in trust without collapsing into voluntarism. The picture that emerges captures what was enticing about its opponent's view without incurring its theoretical costs.

§1 Introduction

I may perform some action, φ, because φ-ing is fun, because it figures in my plans, because it is morally required, or simply because I feel like it. In this way, agency manifests itself in action in a direct, unmediated way; at least in principle, I can φ for any reason whatsoever that I deem to render φ-ing worth the effort. Agency as it pertains to belief, on the other hand, is significantly more constrained. Indeed, according to the dominant view, there is *but one* species of reasons that

can directly support belief: epistemic reasons.[94] It doesn't matter what incentive you provide me with for believing that *p*, if I take the evidence to suggest not-*p*, my hands are tied. This gives rise to a range of perplexing questions regarding reasons, control, and the ways in which these notions each interact with belief and action respectively. The central question of this article, however, concerns *trust*: Is the attitude of trust *action-like*, or rather *belief-like*, in its relation to reasons and control? The literature on this topic is split into two broad camps, which I'll call the *action-like view* (ALV) and the *belief-like view* (BLV) respectively.[95] On the former, trust behaves much like action: it is subject to voluntary control, and responsive to all sorts of reasons. On the latter, trust is more like *belief*: it is beyond the remit of voluntary control and rationally answerable only to epistemic reasons.[96]

A crucial flashpoint of this debate has been cases of so-called 'therapeutic' trust – putative instances of willed trust.[97] If therapeutic trust exemplifies real, full-fledged trust, then ALV follows. Trust cannot be belief-like in its relation to reasons and control. If, on the other hand, therapeutic trust is a distinct attitude from full-fledged trust, then the most prominent counterexample to BLV is dispelled. But such cases are notoriously tricky, and the lesson to be drawn from them is far from clear; intuitions are starkly opposed and difficult to reconcile. As Faulkner puts it: '… the phenomenology is not decisive.' (2025: 87) The result is a dialectical impasse: without recruiting weighty theoretical commitments, or retreating to the aforementioned murky phenomenology, no route beyond it is obviously forthcoming.

In this article I aim to pave the path beyond the impasse over therapeutic trust. Both ALV and BLV will agree that the attitude of trust – whether or not therapeutic instances exemplify it – is both reasons-responsive and norm-governed. We trust for reasons, revise trust in light of countervailing evidence, and hold one another accountable when trust is misplaced or illegitimately withheld. With this widely held assumption on the table, a range of theory-neutral constraints pertaining to the attitude's *rational governance,* its *evidence-responsiveness*, its *patterns of accountability*, all come to the fore. I'll argue that these constraints can be fruitfully deployed as theory-neutral adjudicative resources for evaluating ALV and BLV. If we want to maintain and make sense of the platitude that trust is normatively valanced, then comparing the diverging accounts that ALV and BLV provide of trust's rational profile represents a way to make progress on the debate between them. Ultimately, I'll argue that these considerations bear asymmetrically on this debate, in favour of BLV.

---

[94] See, e.g., Kelly (2003); Shah (2006); Berker (2018); Schmidt (2024); cf. Maguire & Woods (2020); McCormick (2015); Rinard (2017); Meylan (2020).

[95] For the ALV, see, among others: Holton (1994); Jones (1996); McGeer (2008); Frost-Arnold (2014); Carter (2024a; 2024b); Faulkner (2007; 2025). For the BLV, see: Hieronymi (2008); McMyler (2011); Keren (2014; 2020); Hawley (2019).

[96] It is worth making explicit that neither the belief-like nor the action-like view entail a straightforward equation between trust and belief or action in their respective metaphysics of trust. If trust is action-like in its relation to reasons and control, then it can be willed, directly, for non-epistemic reasons; if it is belief-like, then it cannot. This is point of contention I have in my sights in what follows. How the proponents of either view prefer to fill out the relevant details in their broader view of trust's nature will not be my concern. I'll say more about each in the next section.

[97] The debate divides predictably: Holton (1994), Faulkner (2007; 2025), McGeer (2008), and Carter (2024b) treat such cases as conclusive evidence that trust can, at least sometimes, be willed for practical reasons; Hieronymi (2008) and Keren (2020: 116-7) contest this, arguing therapeutic cases most plausibly exemplify something less than full-fledged trust. That verdicts track theoretical commitments in this way is precisely the methodological problem §2.2 diagnoses.

Here, in more detail, is how I proceed. In §2 I explain the case of therapeutic trust and its significance in trust's broader philosophical context. I introduce the two main views on the market, explain where exactly they diverge, and demonstrate the dialectical stalemate that has resulted. In §3 I import theory-neutral constraints from wider discussions in epistemology, ethics, and the philosophy of action, pertaining to the normativity of various types of attitudes. Therapeutic and non-therapeutic instances of trust vary in ways that I argue would be surprising if they were, as ALV maintains, instances of the same type of attitude. Lastly, in §4, I aim to demonstrate that BLV can accommodate much of what made ALV attractive in the first place. More specifically, I develop the BLV in two main ways. First, I show that it can give a satisfying account of the agency and control involved in trust – in particular, by locating it at the level of rational, or attitudinal control rather than direct volition. Then, I show that it can offer a principled explanation of why therapeutic cases have exerted such pressure on theories of trust. If I am right, then not only do we have a principled escape route from an entrenched dialectical impasse; we also see that the BLV can capture what was enticing about its opponent's view without incurring its theoretical costs.

§2 The significance of therapeutic trust

§2.1 Holton's Shopkeeper

On the face of it, it may seem as though trust is straightforwardly belief-like in its relation to reasons and control. After all, while it may be overwhelmingly prudent for me to trust my sociopathic boss, I cannot on this basis simply choose to trust him. Indeed, it seems that I have no more control over whether to trust in such a case than I am typically thought to have over my beliefs, given the thoroughgoing unpopularity of doxastic voluntarism. So far, so good for a belief-like view.

But on this seemingly smooth journey from pre-theoretical intuitions to a belief-like view of trust we quickly confront a major bump in the road. Consider Holton's (1994) shopkeeper case to demonstrate:

> **SHOPKEEPER** Maureen runs a small shop. She discovers that one of her employees, Matthew, has been convicted of petty theft. She decides, nevertheless, to trust him with the till. Doing so, she hopes, is the best way to draw him back into the moral community and promote trustworthiness in him.[98]

Cases such as this represent the flashpoint of the debate between what I'm calling *action*-like and *belief*-like accounts of the nature and normativity of trust. Such cases are by far the most prominent and persistent objections levelled at BLV from the ALV camp. That one can really choose to

[98] This is lifted almost verbatim from Holton's well-known case (Holton 1994: 63), with names added for ease of reference. For similar cases, see Horsburgh (1960); Faulkner (2007; 2025); McGeer (2008); Frost-Arnold (2014); Carter (2024b). Frost-Arnold (2014) has argued for ALV on the basis of a broader range of cases beyond therapeutic trust. I focus on TT here given that it has received by far the most attention in the literature, including, notably, from proponents of ALV themselves.

contravene the evidence and nonetheless trust in such a case is supposed to demonstrate that trust, in this important respect, cannot be belief-like. As Holton puts it:

> "It might be that if you really believe he will steal, you will not be able to trust him. *But you can trust him without believing that he will not.*" (Holton 1994: 63, emphasis added)

McGeer (2008) and Faulkner (2025) read Holton's case in precisely the same way:

> "Yet there do seem to be *genuine cases of deciding to trust* others, deciding to invest our trust in them, even if we remain uncertain, or even anxious, about the consequences of so trusting [Holton 1994]." (McGeer 2008: 241, emphasis added)

> "It seems that the shop owner can nevertheless choose to rely on his new employee to take charge of the till and *in choosing to do this can decide to trust* his employee not to steal from the till." (Faulkner 2025: 66, emphasis added)

These cases – according to these authors – demonstrate that real, bona fide trust at least sometimes obtains in conditions in which belief could not.[99] This, in turn, demonstrates that trust – unlike belief – can at least in certain circumstances be willed directly, for practical or moral reasons. Faulkner (2007) is explicit on this point:

> "She can choose to accept her new employee's statement that he will not steal without assessing his testimony as evidence for its truth, and ignoring the evidence she possesses for its falsity... The reformer does not need evidence to support this presumption – she need not have assessed his character, for instance – *she simply chooses to trust.*" (Faulkner 2007: 885-886, emphasis added)

Indeed, on Faulkner's view, "we have greater voluntary control over trust that belief because we have greater voluntary control over acceptance than belief." (Faulkner 2007: 886) On this view, then, it is trust's amenability to voluntary control – and its consequent openness to practical and moral reasons – that mark its decisive rational divergence from belief.

Properly understood, however, the claim is not that we can trust arbitrarily, for any reason whatsoever, and with no constraints whatsoever.[100] It is rather that the kinds of rational constraints on trusting are closer to those of action than those of belief. There are certain actions that I simply cannot perform; perhaps because I cannot muster the strength, or perhaps not without great cost. But these constraints are importantly different than those that bear on belief. No practical or moral reason, no matter how weighty, can rationally bear on belief, and if this is true it is made so just by

[99] See also J. A. Carter's (2024b) desiderata for a satisfying analysis of therapeutic trust: (i) that it counts therapeutic trust as genuine trust (2024b: 1); and (ii) that it does not reduce therapeutic trust to merely normatively bad or ill-advised trust (2024b: 6). Therapeutic trust, then, on Carter's view, at least sometimes amounts to full-fledged, rational trust. As a result, our theories of trust's nature need to make space for it.

[100] Both Holton (1994: 69) and Faulkner (2007: 886; 2025: 66) acknowledge that there are constraints on voluntary trust; the point is that such constraints are significantly less stringent – indeed, categorically different in kind – than those of belief.

virtue of the kind of activity that belief is.[101] Practical and moral reasons, as it is often put, are *the wrong kinds of reasons* for belief.[102] That we can – in certain circumstances – directly will trust into existence for practical or moral reasons, then, constitutes a significant divergence in its normative profile from that of belief.

As noted at the outset, rational appraisal of trust is a natural and ubiquitous practice – we frequently criticise agents for whether and how they place their trust. But, as we now see, ALV and BLV give substantively different accounts of what it is that we are doing we engage in such rational appraisal. If **SHOPKEEPER** exemplifies full-fledged trust, then when we interrogate the rationality of someone's trusting, we do so under the presupposition that it was under their voluntary control – in the way we might assess an action. If it does not, then our normative questions look importantly different: we might ask why you trusted given the evidence you had, for example, or given everything that you already knew about the trustee. Such questions are meaningful, but they abandon the presupposition that the locus of evaluation is something you directly chose to do.

In what follows I seek to shed light on the question of what it is that we're doing when we rationally evaluate trust. But first thing's first: I will show why, at least as things stand in the literature, we should not expect much progress in the debate between ALV and BLV.

§2.2 A dialectical stalemate

As we have seen, **SHOPKEEPER** represents a crucial fork in the road: if this is full-fledged trust, significant pressure is placed on BLV; if it turns out to be something less, or even just something different, then the view is on firm ground. But, as I noted at the outset, intuitions are notoriously fuzzy here, and the phenomenology indecisive. The result is something of a dialectical stalemate: it is hard to see past these fuzzy intuitions and their entanglement with heavyweight theoretical commitments.

The exchange between Hieronymi (2008) and Carter (2024b) demonstrates the persistence of this stalemate. Hieronymi argues that theories of trust need not subsume and explain therapeutic instances because what they exemplify is not *full-fledged* trust. In support of her view, Hieronymi cites the notion of the *felicity of complaints* in therapeutic instances. Were the trustee to discover, having been therapeutically trusted, that their would-be trustor had therapeutic considerations in mind, they could legitimately complain: "You didn't really trust me." (Hieronymi 2008: 230) The felicity of such a complaint grounds the demotion of therapeutic instances to a category below full-fledged trust. Our theories need not accommodate such cases because they are not pure instances of the attitudinal category in question. Carter (2024b), however, disagrees. In such cases, he points out that one might just as naturally *praise* or *thank* our would-be trustor upon discovering their therapeutic means: "Wow, you trusted me without believing – you must have *really* trusted me!" (Carter 2024b: 3) On Carter's view, complaints are no more appropriate than praise in therapeutic cases, and he thus develops an account on which such cases count as *bona fide*, or *full-fledged* trust.

[101] Cf. Maguire & Woods (2020).
[102] See Hieronymi (2006); Way (2012) for canonical developments of the wrong kind of reason constraint.

What is noteworthy about this exchange is not merely that intuitions diverge, but that they do so in ways that track antecedent commitments about trust's relationship with reasons and control. Those inclined to think that trust's rational governance is belief-like will naturally approach therapeutic cases with scepticism, while those inclined toward a more action-like account are primed instead to treat them as paradigmatic and revealing. This difficulty persists even when we narrow our focus to particular moments within such cases. For instance, in responding to Hieronymi's contention that cases such as **SHOPKEEPER** are better thought of in terms of *entrusting* (an action) rather than full-fledged trust (the attitude we seek to analyse), Faulkner (2025: 66-7) argues that this mislocates the phenomenon. While it may be an accurate description of the particular moment at which Maureen hands over the keys, there is a prior mental event – perhaps the night before, when Maureen ultimately decides to therapeutically trust Matthew for all the reasons described – that looks like trust proper. But this doesn't resolve the indeterminacy any more than it kicks it down the road. Our intuitions about Maureen's attitude, narrowed down to the night before handing over the keys, are no more settled, no more lucid, and no less hostage to antecedent theoretical commitments, than our intuitions about the case considered as a whole.

In this way, intuitions surrounding such cases risk becoming entrenched. But if this diagnosis is right, it suggests a methodological error. If intuitions about therapeutic trust are both indecisive and theoretically heavyweight, then they are ill-suited to the crucial role they currently play in the debate. In the next section I thus propose a methodological shift: I pursue an analysis of trust's relation to reasons and rational control independently of any particular intuitions in cases like **SHOPKEEPER**, and independently of any background theory regarding trust's nature.

§3 Theory-neutral constraints

In what follows I assume that trust is rationally evaluable, reasons responsive and norm governed. This assumption is deliberately modest.[103] It does not settle what kinds of reasons govern trust, nor how they do so. It merely assumes that trust is an attitude whose adoption, maintenance, and revision are subject to standards of rational and evaluative assessment.[104]

With this assumption in place, a familiar methodological approach suggests itself. Across epistemology, ethics, and philosophy of action, attitudes are at least partly characterised by their rational role: by the (§3.1) *kinds of considerations* by which they are rationally governed, by the (§3.2) *patterns of evidence-responsiveness* to which they are subject, and by the (§3.3) *forms of normative criticism* to which they are answerable. Hope, desire, belief, intention, and acceptance differ not merely in their contents, but in the role they play in deliberation and reasoning. These differences provide principled resources for distinguishing kinds of attitudes.

[103] It is noteworthy that even Faulkner (2007), one of ALV's staunchest proponents, devotes considerable attention to specifying when affective trust is and is not justified – taking for granted, that is, that trust falls within the space of rational evaluation even on a view that denies it is belief-like.

[104] The argument in what follows exploits a shared commitment: both ALV and BLV agree that trust can be the subject of reasons and deliberation (whether practical or epistemic), can be justified or unjustified, reasonable or unreasonable, and so on. This shared ground is what gives the theory-neutral methodology its purchase – a point worth noting given that Eliminativism about reasons for attitudes is a live option elsewhere (see, e.g., Maguire (2018) and Schultz (2025) on affective attitudes), and would foreclose this approach entirely if it were in play for trust.

The crux of the matter is as follows. A fundamental disagreement between ALV and BLV is that, in reaction to cases like **SHOPKEEPER**, ALV stipulates *one* type of attitude: trust, of which both therapeutic (TT) and non-therapeutic (NTT) instances are examples. BLV, on the other hand, reacts by stipulating *two* types of attitude: full-fledged trust, which subsumes only NTT and relegates TT to some other, derivative category. If ALV is right, we would expect TT and NTT to exhibit a broadly unified rational profile. And yet, across each dimension of rational role, I'll argue that they do not. They each rationally behave in ways that sit more naturally with BLV's *distinct*-attitude verdict.

§3.1 Unity of considerations

Let's now consider the first, most general dimension of rational role, which has to do with the kinds of considerations that constitutively govern an attitude. The argument will start at this general level, and from there will get increasingly specific as we proceed. Here is a pre-theoretically neat and plausible idea that seems to obtain across (at the very least) a wide array of attitudes:

> **Unity of considerations** Attitude types are partly individuated by the kinds of considerations that rationally bear on them.[105]

Instances of desire, for instance, will vary in important and interesting ways once the particular contents and contexts are filled out. But part of what makes each instance a token of the same broader type of attitude is that it is sensitive to the kinds of considerations sensitivity to which is partly constitutive of that type of attitude. Same goes for belief, intention, and so on. Nothing hangs here on filling out the details regarding the particular kinds of considerations for any particular kind of attitude. The point is purposefully less committal: one helpful tool for individuating attitude-types draws on the kinds of consideration by which they are constitutively governed.

To bring to light why I think that **Unity of considerations** supports BLV's distinct attitude verdict, note one point of agreement between all parties to the debate: epistemic considerations *very frequently* rationally bear on trust.[106] Everyone can agree that, at least in the vast majority of garden variety instances, epistemic reasons rationally bear on trust. One's trust in their sociopathic boss, who in the past has only ever let them down, is rationally criticisable on epistemic grounds: you have decisive evidence counting against trusting him.

Let's call this data point *rational epistemic dominance*: that epistemic considerations explain the rationality of trust across (at the very least) a wide array of cases. So far, so uncontroversial. But now note the problem that this thought poses for ALV. In all of these garden-variety cases, trust is irrational precisely because the epistemic reasons bear against it – and this holds irrespective of

[105] The idea that attitude types are individuated by the reasons that bear on them has deep roots in the reasons literature – see Hieronymi (2006); Way (2012); and, for a comprehensive overview, Schultz (2024). It is worth noting that Hieronymi (2008) herself has deployed related considerations specifically in the trust debate, though without the systematic, theory-neutral development pursued here and, as outlined in the last section, still with appeal to intuitions across contested cases. The present methodology aims to show that once these considerations are deployed independently of any prior commitment to BLV or ALV, they bear asymmetrically on the debate – in favour of the former.

[106] See Keren (2014; 2020: 111-113) for extended discussion of the parallelisms between trust and belief.

whatever practical reasons point in the opposite direction. You can offer me £100 to trust my sociopathic boss and it won't make the trust any more rational *qua* trusting attitude. You can even increase my incentive tenfold: still no difference. Moreover, structurally analogous cases abound – any instance in which counterevidence seems to render trust irrational, irrespective of whatever practical or moral reasons obtain, tells the same story. But across all such cases, are those practical reasons simply systematically outweighed by the epistemic ones? What would explain this rational epistemic dominance? Notice how peculiar ALV's position is on this matter: practical normativity, on their view, is always potentially in play in instances of trust, it just systematically loses out to the epistemic.

This is, of course, no knock-down objection. But it marks an underappreciated asymmetry. BLV straightforwardly predicts and explains the rational epistemic dominance data: therapeutic trust is not real trust, so practical reasons never rationally bear on the attitude. ALV, by contrast, is forced to say that practical reasons are rationally in-play but systematically outweighed – an awkward coincidence rather than a principled prediction. Note the subtlety here: both sides will disagree about whether therapeutic instances exhibit real trust. My point is that they should agree on the data point of rational epistemic dominance. From there, BLV has a comparatively neat story; ALV does not.[107] There we have our first asymmetry.

§3.2 Revision dynamics

To see the next asymmetry, let us now consider how TT and NTT each respond to evidence. Consider a slight development of **SHOPKEEPER**, call it **SHOPKEEPER***. Suppose Maureen has already hired Matthew, and knows relatively little about him. Now consider the particular moment at which she obtains evidence E, where E = *that Matthew has a criminal history*. My contention is that E counts decisively against NTT such that, upon obtaining E Maureen is precluded from rationally (non-therapeutically) trusting, and the same is emphatically not so with respect to TT. In fact, it looks as though the obtaining of E is precisely what grounds or makes TT possible. Without this evidence regarding Matthew, there is no case to be made that this is an instance of therapeutic trust, or trust that is directly willed for practical or moral reasons. It is just plain old non-therapeutic trust prior to the uncovering of E. If TT and NTT are one and the same type of attitude then we would not expect that a single body of evidence could simultaneously (i) constitute a decisive reason for abandonment, and (ii) ground the formation of that very same attitude.

The reader may be reminded at this point of the permissivism versus uniqueness debate in epistemology.[108] The former is the view that the same body of evidence E can rationally mandate more than one doxastic attitude, e.g., belief and suspension, whereas the latter is the view that the same body of evidence can mandate at most one doxastic attitude. One may ask: are ALV's revision dynamics not exactly what one would expect if permissivism were true? The problem for

[107] The ALV theorist might be tempted here by a kind of contextualism: only epistemic considerations bear directly on trust, but practical and moral considerations can set the evidential threshold in a given scenario. The revision dynamics discussed in §3.2 address this: I show there that practical and moral factors bear on TT in ways that are considerably more divergent from NTT than a threshold-shifting picture would predict.

[108] See White (2005) and Feldman (2007) for canonical Uniqueness defence, and Kelly (2013) and Schoenfield (2014) for Permissivism defence. For helpful overviews see Kopec & Titelbaum (2016) and Smith (2025).

ALV, however, is not mere permissivism. Permissivism is the view that one body of evidence may speak in favour of more than one attitude; the problem here is that the very same body of evidence that rationally mandates the abandonment of non-therapeutic trust simultaneously grounds and makes therapeutic trust possible.

Supposing permissivism is true, it may well be true that the same body of evidence could warrant both trust and therapeutic trust (if they are different attitudes after all), or trust and optimism (if those are different attitudes), but this won't help with the specific problem I've pressed here. We would not expect one and the same body of evidence to play the rational role it plays in **SHOPKEEPER*** if TT and NTT are the same attitude. Unified attitudes may permit a range of responses with respect to the strength or weight placed in the various considerations that bear on its adoption or continuation, but we would not ordinarily expect them to exhibit and permit entirely incompatible rational revision patterns in response to the exact same body of evidence. By contrast, if TT and NTT are two different kinds of attitude, things are exactly as we should expect: two different revision profiles because different rational attitudes, answering to different kinds of considerations.

The problem for ALV here is not merely that epistemic considerations clash with practical or moral considerations in deliberation; given the prevalence of normative clashes all across the normative landscape, this much would be exactly as expected. Instead, the tension arises because one and the same body of evidence seems to count decisively against NTT whilst grounding TT. This is not mere normative clash throughout the process of deliberation, it is an internal tension in the kinds of consideration that support trusting as such. This tension only arises on a *same-attitude* verdict, and it is a peculiarly deep tension – not merely that E is weighed differently by two attitudes, but that E plays structurally opposite rational roles simultaneously. This is an awkward feature to accommodate in one's account of the rationality of trust if one wishes to maintain that the target attitude is exemplified in both TT and NTT cases.

§3.2.1 Trust and evidence

Suppose the proponent of ALV concedes that the rational profile of trust, on their view, is indeed idiosyncratic in the ways I've highlighted. Still, they may insist, on BLV it is worse. Consider the following, well-known case:

> **BABYSITTER** A paranoid parent organises a babysitter for their child to allow them to go out, then proceeds to spend the evening out monitoring their babysitter's antics remotely via a 'nanny-cam'.[109]

The paranoid parent is, as Wanderer and Townsend put it, "… not only a lousy date, but also a lousy trustor; in performing the seemingly rational act of broadening the evidential base relevant to her judgements of trustworthiness, she is, precisely, failing to trust the babysitter" (2013: 1). Whether the φ in question is a practical one like looking after a child, or an epistemic one like

[109] For versions of this worry with BLV, see Baker (1987); Jones (1996); McLeod (2002); Faulkner (2007; 2025).

telling the truth, trust as such is in tension with A's seeking independent, non-trust-based, evidential grounding that B will φ.[110]

Here is the problem as it pertains to the present debate: if trust is belief-like in its reasons-responsiveness, it should not embody this inverse relationship with growing evidence. As the evidence increases, the rationality and propriety of trust decreases. This is better explained if ALV is true, given that there is no rational tension in contravening the evidence when *acting on*, *accepting*, or *being hopeful* with respect to some proposition. The tension arises only if, in trusting, one adopts a *cognitive*, or *belief-like* attitude. Might the proponent of ALV thus respond that, on balance, it is their view that ultimately makes better sense of trust's particular relationship to reasons and evidence?

Two versions of the problem are worth distinguishing and treating separately: A's trust feels inappropriate or misplaced when A *knows*, or is *certain*, that B will φ; and whether A trusts B to φ is undermined by A's *monitoring* behaviour with respect to B's φ-ing. I'll argue that BLV can comfortably handle both. Let's take each in turn.

Once A is certain that B will φ, trust is inappropriate or misplaced. If trust is *belief*-like in its reasons-responsiveness, we would not expect increasing evidence to undermine it in this way; the spectrum of *increasing* evidence, after all, should only rationally *improve* belief. But the BLV theorist can acknowledge this felt tension while denying that it jeopardises their view. Just as there is a lower bound on trust (less than belief), there is also an upper bound (perhaps knowledge, or certainty). The attitude of trust operates neither in conditions of total certainty nor total absence of evidential warrant; that is just the kind of attitude that it is. This response is perfectly consistent with BLV's position, which, recall, does not *reduce* trust to belief. The view is that trust is belief-like in its relation to reasons and rational control. That view is perfectly compatible with it involving upper and lower thresholds. Once the BLV theorist invokes such thresholds, the problem of trust's being inappropriate in conditions of certainty is dispelled.

Now for the second strand of the objection: trust is rationally undermined by monitoring behaviours. The worry is as follows: if trust is belief-like in its reasons-responsiveness, then it needs to follow the evidence. But following evidence (e.g., by incessantly double-checking on the would-be trustee) is exactly what a trustor does *not* do. What a would-be trustor should do to strengthen their belief is precisely what they should not do to strengthen their trust. Monitoring is simultaneously belief-strengthening and trust-undermining, so BLV and not ALV has a problem.

This objection, however, rests on a mistake. As Simpson (2023: 55) has pointed out: the objection equivocates on what it is for one to 'follow' the evidence. On one reading, 'follows' means something like 'conforms to'. A follows the evidence in this sense just if their attitude is fitting given one's presently available total evidence regarding whether p will, or is trustworthy with respect to, φ. On another reading, 'follows' instead means something more like 'pursues'. A follows the evidence in this sense when A, between t1 and t2, gathers evidence bearing on whether B will

[110] For discussion on the putative tension between monitoring and otherwise seeking evidence, on the one hand, and trusting, on the other, see Dasgupta (1988); McMyler (2011); Nguyen (2022); cf. Gordon (2022); Simpson (2023).

φ – when they inquire so as to improve their epistemic position. The former reading of 'follows' is a property of mental attitudes; the latter is an activity.[111] The proponent of BLV can happily agree that a trusting attitude is incompatible with active inquiry whilst denying that said attitude can diverge from the evidence.

Indeed, Friedman's well-known **Don't Believe and Inquire** norm connecting belief to inquiry would, at least prima facie, seem to offer a satisfying explanation of *why* one rationally ought to not gather evidence. Here it is:

> **Don't Believe and Inquire (DBI)** One ought not inquire into/have an interrogative attitude towards Q at t and believe p^{Q} at t.[112] (Friedman 2019a: 303)

If some variant on DBI describes the correct normative connection between the mental attitude of belief and the practical-epistemic activity of inquiring, then BLV would seem to have a good explanation of why one cannot trust that B will φ whilst inquiring into whether they will.[113] Inquiry betrays a lack of belief; a lack of belief betrays a lack of trust. It is the 'pursues' reading of 'follows' that undermines trust, not the 'conforms to' notion; but it is the latter notion that BLV needs in order to maintain that trust is relevantly *belief*-like in its rational interaction with evidence. BLV can thus accommodate the following two thoughts regarding trust's relationship with reasons and evidence: trust is all-things-considered appropriate only in the absence of knowledge or certainty that B will φ; and trust is incompatible with active inquiry into whether B will φ.

§3.3 Accountability and the norms of trust

Let us now turn to the final dimension of rational role that I will discuss: trust's *accountability practices*. Our practices of holding one another accountable – across all walks of life – are deeply intertwined with the norms that govern us. The following principle from Flores & Woodard (2023: 10) illuminates the thought:

> **Practice Indicates Norms** If there is a legitimate practice of criticising agents for whether and how they φ, then we have reason to think that there are norms on φ-ing.[114]

---

[111] Simpson (2023) himself puts the point in terms of a distinction between a synchronic vs. diachronic interpretation of 'follows'. However, in conversation, Matt Vermaire has convinced me that one's attitudes could also diachronically follow the evidence in the 'conforms to' sense, for example if they update appropriately as the evidence changes. For this reason, I put the point in terms of 'conforms to' vs. 'pursues'.

[112] For discussion of other variants, see the literature surrounding the well-known Ignorance norm (Friedman 2019b; Kelp 2021a, 2021b; Haziza 2023a; Willard-Kyle 2023 cf. Archer 2021; Falbo 2023; Woodard 2024).

[113] One complication here is that the BLV maintains that trust is belief-*like*, and not *mere belief*, and thus it is not immediately obvious whether DBI directly applies. But the monitoring case only requires a norm connecting trust to inquiry that is structurally analogous to DBI – and such a norm is independently plausible given trust's belief-like profile. The details of how exactly trust and inquiry interact fall beyond the scope of this paper, but the synchronic/diachronic distinction does the essential work regardless of how those details are resolved.

[114] For related principles and particular applications thereof see, among others, Darwall (2014), Kauppinen (2018) and Simion & Kelp 2021: 24). The focus is restricted to 'legitimate' practices because such practices are, at best, an imperfect guide to the relevant norms. Suppose, for example, there were a practice of criticising women for speaking out of turn. Such a practice would not ground a norm to the same effect. While I will not do so here, more needs to be said, then, on what exactly makes a norm legitimate. For more, see Goldberg (2018: 170); Flores and Woodard (2023); Simion (2024).

Practices of accountability shed light on the underlying normativity of the domain(s) to which they apply. This much should not be surprising given the overwhelmingly plausible idea that norms must bear some connection to our actual or possible practices of interpersonal regulation (Darwall 2014; Flores & Woodard 2023). Given **Practice Indicates Norms**, if TT and NTT were two different kinds of attitudes, with different rational profiles, we should expect that they are subject to different practices of accountability. As it happens, this is exactly what we find.

Before proceeding, one clarificatory point is worth making. On my view, TT and NTT are, at best, two different types of attitude.[115] Given this, it is not strictly speaking possible for one and the same instance to count as both TT and NTT. In the following cases I am simply asking the reader to consider each case under each description in turn, and ask: what would criticisability for such an attitude look like if it were TT/NTT? Given the principles laid out above, this hypothetical should be informative with regards to the question of whether we are dealing with two instances of the very same attitude, or rather two distinct attitudes in each case.

Consider, once again, Holton's **SHOPKEEPER** case, but this time through the lens of criticisability. We have a trustor, Maureen, a would-be trustee, Matthew, and a body of evidence, E, where E = that Matthew has a criminal history. Considered as an instance of TT, it is hard to see how Maureen is criticisable for 'trusting' as she does. Given her therapeutic aim, and the details of the case, this looks like a paradigmatically rational instance of the kind of attitude that therapeutic trust is. Maureen, owing to her excellent judgement of character, is responsive to the precise aspects of Matthew's character that render him a good therapeutic candidate.

Considered instead as an instance of garden variety, non-therapeutic trust, however, the same is not so. On the contrary, it is hard to see in what sense, given our normal practices of criticising non-therapeutic trustors, Maureen might *not* be criticisable. After all, even if all goes well, Matthew may still steal out of the till for (at least) the first couple of times. Moreover, he will do so for precisely the reasons that Maureen is all too aware of, given his recent criminal history. E, then, combined with Maureen's full awareness of it, seem to straightforwardly rationally undermine the attitude she thereby takes up, at least when considered as a putative instance of NTT. This divergence in patterns of criticisability is what we would expect if TT and NTT were altogether different types of attitude.

But the divergence runs deeper still, for we can easily imagine cases running in the opposite direction, wherein TT would be criticisable but NTT uncriticisable. Consider the following case:

> **DISILLUSIONED HOUSEWIFE** Larry trusts his long-term partner, Dana, to carry out various household chores. He does so in part because, throughout the entirety of their long marriage, she has always done so. However, every instance of his so trusting progressively erodes her trustworthiness (with respect to the φ in question) as she becomes increasingly disillusioned with the emerging dynamic between them. Dana sums up her

[115] Indeed, on my view, TT may well be not an attitude at all but rather an action – namely, entrusting.

> frustration to Larry: "I don't have a problem with doing household chores; I have a problem with *being trusted* to do them."[116]

Consider the first relevant instance after said conversation. Larry has solid evidential reason to think that, on this particular occasion, Dana will indeed do the household chores. It is hard to see any sense in which Larry's outright trust – absent a therapeutic aim or purpose – is rationally criticisable.[117] But, considered as an instance of TT, with a forward-looking, trust-building aim, it is unquestionably highly criticisable. He has every reason to think that the effects of his trusting her will have disastrous consequences with respect to her eventual trustworthiness. The trusting relationship is being damaged precisely by his trusting.

If TT and NTT were one and the same type of attitude, and that attitude is within the realm of reasons, rationality, and normativity, it would be surprising if it were subject to such incompatible (and in fact opposing) patterns of criticisability. By contrast, such patterns are precisely what we would expect to find if TT and NTT were two different types of rational attitudes, governed accordingly by different norms. If, as the previous sections argued, TT and NTT are rationally answerable to different kinds of considerations and subject to different revision dynamics, then we would expect their respective norms to diverge, too. These cases suggest that they do.

To be clear, the point here is not that the differences in our practices of appraisal for each *determine* the underlying metaphysical types. The point is rather that, insofar as one (along with many others) buys the tight connection between norms and underlying normativity, then this divergence favours BLV over ALV. The norms across therapeutic and non-therapeutic cases are systematically different and often in tension, and this stands in need of explanation only if one wishes to maintain that they are nevertheless instances of the same kind of attitude.

That TT and NTT differ in terms of their (i) unity of considerations; (ii) revision dynamics; and (iii) patterns of accountability, is, as I've emphasised throughout, only an explanatory burden on ALV. These are problematic differences only if we want to maintain that trust is a reasons-responsive, norm-governed attitude that is exemplified equally in therapeutic and non-therapeutic cases. If we drop this ambition, and instead adopt BLV's approach, whereby we deny that therapeutic trust exemplifies our target attitude, we can have rational unity in our analysis without tension. These differences speak in favour of BLV's *distinct-attitude* response. The locus of analysis, they may say, is *real,* or *full-fledged* trust, and these differences constitute reasons to think that this is not that.

§3.4 Biting the bullet

Consider the following objection. On one reading of ALV, TT and NTT are interrelated insofar as they are *species of the same genus*: namely, trust. We may gladly bite the bullet, then, with respect to the normative divergence pressed throughout this section. Precisely what differs between TT and NTT is the practical aim or purpose with which the former is constitutively bound up. It is this

[116] Hawley (2014: 7) makes a similar observation – namely that trust itself can sometimes be the wrong attitude.
[117] *Outright* NTT because Dana's testimony may constitute a partial defeater for Larry's confidence that she will φ. The point remains, however: his steadfast maintenance of *outright* NTT is not rationally criticisable in such a case.

distinctive, additional, practical aim that generates the rational idiosyncrasies that TT exhibits. This is all exactly as ALV would predict if the target locus of analysis is the genus trust, and TT is a species of that genus.

But notice that the burden of proof is now on the bullet-biter to provide a principled account of the genus *trust* such that its species admit of such normative divergence. BLV has given principled reasons for maintaining a *distinct genus* verdict; namely, that the putative counterexample, TT, differs from garden-variety instances of trust in deep-lying, rationally constitutive ways. The bullet-biter has conceded this normative divergence – indeed, this is what warrants a *different species* verdict, on their view, even if not a *different genus* verdict. But once normative divergence is conceded, the bullet-biter owes us an account of the genus such that both TT and NTT are species of it. Without such an account, the genus risks overexpanding to include attitudes that everyone wants to exclude – mere reliance for example.[118] The methodology developed in §3.1 shows why that account will be difficult to provide: by **Unity of considerations**, rational profiles are partly constitutive of attitude types. NTT tokens share such a profile, and practical reasons play no role in it. The practical reasons that bear on TT are therefore *additional,* and not *constitutive*, precisely because they do not figure in the profile that unifies every other instance of the attitude type.

Until a satisfying positive account of the genus trust is given such at its species both (i) include TT and NTT; and (ii) happily admit of the rational divergence here detailed, the distinct attitude verdict is on far firmer ground than the alternative.

§4 Explaining the appeal of ALV

Two interconnected thoughts sustain ALV's appeal. First, trust seems agentive – not a mere passive registration of evidence, but something we deliberate over, resolve to do, and are responsible for.[119] It may seem that BLV, in denying voluntary control, is thereby committed to an impoverished account of this characteristic agency.[120] Second, TT and NTT feel deeply intertwined: even if TT does not qualify as full-fledged trust, it shapes behaviour and structures interpersonal relations in ways that closely resemble NTT. In what follows I show that both thoughts can be accommodated within a belief-like framework, drawing on attitudinal control to vindicate trust's agential character, and on a functional perspective on attitudes to explain the closeness between TT and NTT. If I'm right, then BLV can capture what was enticing about its opponent's view without incurring its theoretical costs.

[118] The idea that trust must be distinguished from mere reliance has been ever-present, accepted by all sides of all debates in the literature since (at least) Baier's (1986) seminal work.
[119] Carter (2024a: 37-8) draws a helpful distinction, between *deliberative* and *implicit* trust, that complicates this picture somewhat. Implicit trust is passive rather than agentive insofar as it arises spontaneously or unreflectively. Deliberative trust, on the other hand, is agentive insofar as it is the result of deliberation. This section can be read as developing BLV's account of *deliberative* trust – that is, answering the question as to what agency in a belief-like account of trust looks like, when trust does indeed involve agency.
[120] This worry is quite natural given that prominent accounts of trust foreground its volitional character as central to the agency it involves. See, e.g., Hinchman (2019), as well as the ALV references cited in fn. 95.

§4.1 Agency and control without voluntarism

Current discussions of therapeutic trust encourage a false dichotomy: either trust is action-like and voluntary, and thus deeply agential; or it is belief-like, involuntary, and thus largely passive. But voluntariness and agential control, as the literature on doxastic (in)voluntarism makes clear, are not equivalent.[121] An attitude may not be directly subject to the will and yet still fall within the scope of rational self-governance. To develop BLV, I therefore articulate a conception of the agency involved in trust that lies squarely in the space of *rational control without voluntarism*. The rational control that trust involves is attitudinal, not volitional.

I borrow the term 'attitudinal control' from McHugh's (2012; 2017) influential work on the topic. McHugh's (2017: 2746) starting point is the *epistemic responsibility problem*: how can it be that beliefs are properly subject to prescriptive norms and yet we do not have direct, voluntary control over them? As McHugh puts it, the problem "… turns on the implicit assumption that bodily action is the paradigm for norm-governed, responsible conduct, and that anything that does not resemble bodily action falls into the same category as headaches." (McHugh 2017: 2748) This assumption is mistaken because there is a large, significant agentive realm on the voluntariness spectrum that lies between bodily action, at one end, and headaches, at the other. Intending, for instance, is a paradigmatic agentive state: it is reasons-responsive, subject to prescriptive norms, and something for which we can be held to account. Yet, as Kavka's (1983) toxin case illustrates, intention is not under our voluntary control; I cannot simply intend to φ for practical or moral reasons. The same is true of belief (McHugh 2017: 2759).

Responsibility is thus the crucial ingredient of agentive conduct, but *voluntariness* is not necessary for it. McHugh draws a helpful distinction between *voluntary* and *attitudinal* control, one that we'll presently put to use in the current debate.[122] The latter notion requires not voluntariness but rather reasons-reactivity and -responsiveness. When we can form, suspend, and sustain attitudes *for reasons*, then they are agentive and responsibility-involving even if not straightforwardly voluntary.

Herein lies the makings of a promising and BLV-friendly characterisation of trust's agency. By eschewing voluntariness, BLV risks leaving the impression that trust is thereby headache-like in its relation to agency. But as McHugh and others have helped us to see, there is plenty of scope for a substantive and satisfactory account of trust's agency in the vast terrain between voluntary action and headaches. Once this terrain is acknowledged and properly exploited, a central motivation for ALV loses much of its force. The route from 'trust is agential' to 'trust is voluntary' rests on a mistake: it errantly treats voluntary action as the exclusive or paradigmatic instantiation of agentive conduct.

Trust, on the present account, is non-voluntary but it is reasons-responsive and agency-involving in the same way that belief is. I cannot switch my trust on or off just like that, but whom and how I trust says something *about me.* It says something about the kind of cognitive, moral, and epistemic agent that I am. I trust for reasons, and I can reflect on those reasons. I can deliberate about

[121] Most pertinent here is McHugh (2017), which I return to presently, and Kearl (2023), which articulates a conception of epistemic control without voluntarism in elucidating what epistemic agency amounts to.
[122] Attitudinal control is essentially what Moran (2001) calls 'deliberative' and Hieronymi (2009) calls 'evaluative' control. The differences between each of their views and McHugh's won't matter for present purposes.

whether I've been too conservative in my willingness and openness to trust. I can actively engage myself in the project of altering and developing my trusting practices.[123]

It thus makes perfect sense, and is in fact essential to the practice, that I am held accountable *qua* trustor. This is a crucial tool in the honing of good, responsible trusting behaviour. Criticism functions as a signal that I need to cultivate better trusting habits, even if I can at best influence them only over time, or expose myself to different evidence, different people, and different testimonies. This is the picture of trust's agency that BLV can (and, to my mind, should) develop, and none of it requires voluntary control.

§4.2 Why therapeutic trust feels so close to trust proper

Even if trust can be agentive without being voluntary, a further worry with BLV's approach remains. In particular, irrespective of whether they are the same type of attitude, TT and NTT are undoubtedly closely linked. BLV maintains a distinction in attitude kinds, and so owes an explanation of this interconnection.

A fruitful starting point for BLV, in capturing the connection between TT and NTT, is the considerable precedent that exists for theorising and analysing attitudes teleologically, that is, by reference to their function – the kind of role they plan in planning or theorising – or their aim – their success conditions and rational culmination point.[124] Along these lines, one point is particularly worth emphasising for the proponent of BLV: TT aims at NTT, is normatively governed by it, and, when successful, can give way to it. In other words, the two, while distinct, are thoroughly interconnected. What good TT-ing amounts to is in no small part determined by what good NTT-ing amounts to. Whether or not someone is an appropriate object of TT depends on facts about their (eventual) trustworthiness. Normatively good manifestation of the therapeutic variant is thus entirely bound up with the normativity of its non-therapeutic counterpart.

It is important to note the asymmetric *direction* of normative connection at play here. TT is normatively shaped by NTT as just described; but the reverse is not so. What TT does, in the normative, functional sense, and what it aims at, has entirely to do with what it is that NTT does and aims at, but what NTT does and aims at has nothing to do with TT. The normativity of my non-therapeutic trust in you has nothing to do with whether it was preceded by therapeutic trust or whether it might serve as a basis for future therapeutic trust. Such considerations are simply not relevant to NTT.

There is thus a plausible and coherent story to be told on BLV about the close relationship between TT and NTT. The former is normatively directed toward the latter, and when all goes well can be supplanted by it. If that is right, then it is no surprise that they feel so close, or that the debate has

[123] An interesting potential route for developing this line might draw on the kind of agency involved in what Flores has called one's *epistemic style* – a way of interacting with evidence that express (aspects of) a unified set of epistemic parameters (2021: 8; 2023). Epistemic styles are sites of epistemic agency, and products not only of one's own subjective characteristics but also of the broader environments in which we found ourselves.

[124] See, in particular, Bratman (1987; 2014). For these ideas as they pertain to inquiry, see in particular Friedman (2025; forthcoming) and Paulson (ms). That Bratman-style functionalism might bear fruit in the trust debate should be unsurprising given the precedent in this literature for analysing trust via its function. See, among others, Faulkner (2007; 2025); Simion & Kelp (2023); Bieber & Viehoff (2022).

persisted, or that ALV might have appeal. And yet this positive story of interconnection is perfectly compatible with BLV's verdict, which as §3 demonstrated we have reason to favour. They have different aims, different functional and rational profiles, and are thus meaningfully distinct attitudes. Given familiar methods for analysing and theorising about rational attitudes, then, an account of the close relationship between TT and NTT is perfectly compatible with, and in fact nicely explained by, BLV's distinct-attitude verdict.

§5 Conclusion

In sum, theory-neutral considerations vindicate the belief-like view: it is reasons-responsive and normatively-valenced, but it cannot be willed in the way that action can. Such an account neither denies trust's agentive character nor underplays the interconnection between it and its therapeutic variant.

Chapter V.

Trust's Meno problem: can the doxastic view account for the value of trust?[125]

Abstract: The doxastic view (DV) of trust maintains that trust essentially involves belief. In a recent paper, Arnon Keren gestures toward a new objection to the view, labelled *Trust's Meno Problem* (TMP), which calls into question the DV's ability to explain the widely held intuition that trust has distinct and indispensable value. As of yet, there has been no attempt to take up TMP on behalf of DV. This paper aims to fill precisely this lacuna. I do so in three main stages. In §1 I contextualise and elucidate the problem, to which Keren gestures but does not address in detail. In §2 I disambiguate multiple possible interpretations of TMP, seeking to identify the most philosophically challenging. Finally, in §3, I argue that DV can solve even this interpretation. In order to do so, I make use of the highly plausible claim we find in the work of Katherine Hawley: that trust *pays a compliment* to the trustee. The payoffs of exploring the doxastic view in the context of Trust's Meno Problem are twofold: we better understand the nature of the problem itself, and we see that the doxastic view can give a satisfying answer.

§1 Introduction

We may split the philosophical study of trust into two distinct questions: one ontological, and one axiological. The former concerns trust's nature - what trust essentially is. Trust, it seems, is a kind of reliance, but crucially, it is not *mere* reliance (Goldberg 2020). Addressing the ontological question thus involves providing an analysis of trust that adequately reflects this distinction. Proponents of doxastic views (Adler 1994; Hieronymi 2008; McMyler 2011; Keren 2014, 2020) maintain that trust essentially involves a trusting belief on the part of the trustor, i.e., a belief that the trustee will take care of things as entrusted. Proponents of non-doxastic views (Jones 1996; Baker 1987; McLeod 2002; Faulkner 2007a, 2011), on the other hand, deny that belief is essential to trust, instead arguing for some non-doxastic attitude as the essential component (e.g., a kind of optimism (Jones 1996)).

[125] This chapter is an adapted version of Patrizio (2024b).

Though the two camps thus disagree over the correct answer to the ontological question, they are, for the most part, united in their silence on the axiological question.[126] It is widely agreed that trust is not merely some useful commodity. Rather, trust is, or at least seems to be, distinctly and indispensably valuable for the existence of a society in which individuals can navigate the social world - fostering the kinds of relationships they value, sharing information, engaging in meaningful collaborative projects, and so on. As John Locke put it, trust is the "bond of society" (1663; echoed in Hollis 1998, 4). Sissela Bok similarly writes that, "Whatever matters to human beings, trust is the atmosphere in which it thrives." (1978, 31) Far from being merely a handy tool we might make use of in certain situations, then - like a smartwatch or a hot water bottle - trust, it seems, forms the backdrop of cooperative societies, and solidifies the relationships that shape and structure our lives. But if this is so, then the philosophical study of trust ought to distinguish it from mere reliance not only at the ontological level but at the axiological level, too.

In a recent article, Keren (2020) labels this task *Trust's Meno Problem* (TMP). The Meno Problem for knowledge involves explaining the intuition that knowledge is not merely distinct from, but also more valuable than, mere true belief. What has gone overlooked in the philosophical literature on trust, however, is that a parallel intuition holds for trust as it does for knowledge. If trust is distinctly and indispensably valuable, then we must explain not merely why it is valuable, but why it is *more valuable than that which falls short* (Keren 2020, 117). Moreover, it follows that different accounts of the ontology of trust will provide different accounts of its axiology, and thus that TMP might provide an additional axis along which we can evaluate competing accounts of trust's ontology. (Keren 2020, 117)

TMP thus applies to all competing philosophical theories of trust. However, no sooner does Keren pose TMP than does he suggest that it is particularly worrisome for the *doxastic view* (DV). This is so, the thought goes, because one may *act as if* one has a trusting belief without really having it, and in so doing one may secure the aforementioned value of trust. Insofar as we agree with Keren, then, that TMP represents a novel standard by which we may evaluate competing theories of trust, the search for its solution is most pressing for none other than proponents of the DV. However, this challenge has not yet been explored in any detail, neither by opponents nor proponents of the view. It is this lacuna in the literature that the current paper aims to fill.

Here is how I proceed. §1 elucidates and contextualises the problem as it pertains to the DV. Keren gestures toward it but it is yet to be spelled out in detail. Then, having done so, §2 distinguishes between multiple possible interpretations of the problem, seeking to find the one that presents the most serious challenge to the doxastic view. To this end, I utilise Duncan Pritchard's (2009; 2010; 2011) distinctions between multiple specifications of the original Meno problem. Finally, having clarified the problem and found the most challenging interpretation, in §3 I show that the doxastic view has the resources to provide a satisfying answer. In order to do so, I make use of a highly plausible intuition, found in Hawley (2012; 2019), O'Neil (2012), that well-placed trust represents a way of showing *respect,* or paying a *compliment*, to the trustee. The main payoffs of exploring the doxastic view in the context of Keren's TMP are twofold: we better

[126] A notable exception is Simpson (2012).

understand the nature of the problem itself, and we see that the doxastic view, despite initial appearances, can give a good answer.

§2 The doxastic view and the trust Meno problem

§2.1 Some qualifications

In what follows I will adopt some common assumptions regarding the nature of the attitude whose value we seek to explain. Firstly, I am thinking here of trust as a *three-place, interpersonal* relation, involving two people and a task.[127] (Holton 1994; Jones 1996; Hardin 2002; Hieronymi 2008; Hawley 2012; Carter and Simion 2020; Simion and Kelp 2023) Though we do often speak about cases in which X trusts Y simpliciter, my focus, as is commonplace in the literature, will be on instances in which X trusts Y *to* ϕ. Secondly, and again simply following the dominant view, my focus will be on instances in which X and Y are people. I can rely on objects, but I do not grant them the rich trust I grant people. (Jones 1996; Hawley 2014: 2; Carter and Simion 2020) Thirdly, following Annette Baier's seminal work (1986), I conceive of trust as constitutively involving vulnerability to betrayal. When trusting rather than merely relying, one subjects themselves to the possibility that one will be betrayed, rather than merely let down. (Baier 1986; Dasgupta 1988; Becker 1996) One may well feel let down by a faulty alarm clock, but it is not appropriate to feel *betrayed*. (Baier 1986, 235) Finally, note that the kind of trust whose value we seek to explain here will have its value not unconditionally, but rather only when it is well-placed, or justified. Moreover, notice that narrowing the focus to well-placed rather than any old trust places restrictions on both trustor and trustee. A trusting relation may fall short of the attitude we have in our sights if either (i) the trustor is unjustified in trusting, or (ii) the trustee fails to act *as entrusted*. Plausibly, the distinctive and indispensable value of trust is absent in instances in which I trust you while in possession of positive reasons *not* to trust you.[128] Similarly, if I trust you and you subsequently act as entrusted, but do so out of sheer luck rather than any semblance of recognition, acknowledgement, or responsiveness to my trust, then there is a sense in which you have failed to act *as entrusted*.

§2.2 The doxastic view

In what follows, I take these observations for granted, rather than provide any argument for them.[129] Our focus is on three-place, interpersonal trust, where this is understood as involving vulnerability to betrayal, as well as success conditions on both the trustor and trustee's side (hereafter, just 'trust'). Our answer to the ontological question, the thought goes, ought to be able to explain why trust, so understood, has value over and above that which falls short.

---

[127] See Nguyen (2022) and Coliva (2025) for accounts that abandon the *interpersonal* focus. For lack of space, and so as to walk in lockstep with the vast majority of the philosophical literature on trust thus far, I set such views aside in what follows.

[128] For example, see Carter and Simion (2020) for discussion of "wishful optimism" cases, in which, out of sheer wishful optimism, X trusts Y to carry out some task despite Y's having proven to be extremely unreliable with similar such tasks.

[129] It is important to note that these observations should be uncontroversial insofar as they are shared by most, if not all, competing accounts of the ontology of trust.

We now turn to the doxastic view, of which there are multiple, competing formulations in the literature (Adler 1994; Hieronymi 2008; McMyler 2011; Keren 2014, 2020). It will be beyond the scope of the current paper to settle the question as to which is to be preferred; instead, let's simply establish its key tenets - some common ground occupied by all formulations.

The pivotal claim distinctive of the DV is that trust essentially involves what I will call a "trusting belief" on the part of the trustor (Adler 1994; Hardin 2002; Fricker 2006; Hieronymi 2008; McMyler 2011; Keren 2014, 2020). A trusting belief is here understood as either a belief that (i) the trustee will act as entrusted; or that (ii) the trustee is trustworthy with respect to what she is entrusted to do. The details differ depending on the view, but all insist upon the necessity of such a belief for trust.[130] If X does not *believe* Y to be trustworthy in the relevant sense, then according to the DV, X does not really trust Y.

The popularity of the view is no mystery. For one thing, the DV does an excellent job of explaining trust's systematic relationship with belief. Whatever your preferred ontology of trust, it is hard to deny that trust and belief are intimately related; if X trusts Y to ϕ, why is it that X almost always believes that Y will indeed ϕ, if after all trust does *not* essentially involve belief? Epistemic trust provides a helpful illustration: oftentimes in testimonial exchanges, we trust speakers for the truth and sincerity of their assertion. But why, if trust does not involve belief, does such epistemic trust invariably result in believing what they say? (Keren 2020: 112) Moreover, the doxastic view makes sense of the way in which trust can *give reason* for belief. If, as the DV would have it, X's trusting his colleague, Y, to lock up the office for the day involves a trusting belief, then it is no mystery why X's trusting Y in this respect seems to give X *reason to believe* that Y will do so. Trust provides a reason for belief in such cases precisely because belief can provide a reason for belief (Keren 2020; Carter and Simion 2020). On a non-doxastic view, which denies the necessity of a trusting belief, the intimate link between trust and belief stands in need of further explanation.

Aside from trust's relationship with belief, the doxastic view is also on strong ground with respect to trust and what we might, following Carter and Simion (2020), call mere wishful optimism. Consider a case in which Xavier, knowing full well that Yasmin is recklessly irresponsible, with a disdain for children, nevertheless trusts her to babysit his children. Moreover, he does so out of wishful optimism that she will one day change her ways. It seems that the most natural reading of such a case is that Xavier lost his trust in Yasmin just when he learnt that she was a terribly irresponsible individual, with a disdain for children, and that the attitude he now adopts is somewhat different to trust. The doxastic view does an excellent job of explaining this difference insofar as Xavier stops trusting when he stops believing Yasmin to be trustworthy. Maintaining and explaining this difference is more difficult on non-doxastic accounts, given that Xavier certainly still seems *optimistic* that Yasmin will do so, even if he no longer believes.

The DV is thus, at least prima facie, a position worth defending. The main undertakings one inherits in attempting to do so have been well-rehearsed in the literature to date. In particular, the DV faces a challenge from considerations pertaining to trust's relationship with *evidence* (Baker

[130] Some views, Hieronymi (2008), e.g., make a sufficiency claim. I set aside such 'pure' doxastic views in what follows and focus on the weaker, 'impure' formulation of the view, which involves only a necessity claim.

1987; Jones 1996; McLeod 2002, 2015; Faulkner 2007a, 2011), on the one hand, and trust's relationship with *the will* (Holton 1994; Faulkner 2007a; Simpson 2012), on the other. Whether or not these objections can adequately be addressed by proponents of the DV will be beyond the scope of the current paper.[131] Instead, my focus will be on a novel kind of objection to the view, articulated by Arnon Keren (2020), and dubbed *Trust's Meno Problem*.

§2.3 Trust's Meno Problem

In Plato's *Meno*, Socrates famously raises a quandary about the value of knowledge. If knowledge is, as it seems to be, distinctly and uniquely valuable, then what is it that makes it so? Importantly, this question is not to be divorced from the ontological question regarding the nature of knowledge. Indeed, the two are intimately connected: if one's theory of the nature of knowledge does a poor job of answering the question regarding the value of knowledge, then so much the worse for the theory of its nature. (Kvanvig 2003, quoted in Greco 2006, 91) The *Meno Problem*, as it has come to be known, is, as Greco puts it: "…at the heart of the project of explanation. The task of explaining what knowledge is involves the task of explaining why knowledge is valuable." (2006, 91)

According to Keren (2020), what has gone largely unnoticed is that parallel intuitions to those that generate the original Meno Problem hold for trust as they do for knowledge. The following two intuitions are widely accepted in the literature:

> **Trust Intuition 1**: Trust has distinct and indispensable value.
>
> **Trust Intuition 2**: Trust is not to be equated with mere reliance.

However, if this is right, and trust does indeed possess such value, then just as in the case of knowledge, our theory of its nature had better be able to accommodate and explain this value, and, moreover, this value had better not collapse into the value of mere reliance. This is *Trust's Meno Problem* (TMP).

Keren points out that extant philosophical discussions of trust, which, when they speak of trust's value at all, speak almost exclusively in *instrumental* terms (McLeod 2015), are unlikely to provide an adequate solution to this problem (Keren 2020, 116). For, we can rely on others in ways characteristic of trust without really trusting them, and, presumably, in so doing, realise all of the same instrumental goods associated with trust. Accordingly, TMP consists in explaining why trust is more valuable than relying on others in ways characteristic of trust, if indeed it is. (*Ibid.*)

This challenge, according to Keren, looks prima facie particularly worrisome for proponents of the DV. The reason for this is that proponents of the view, particularly when pressed on objections pertaining to trust's relation with the will, are especially apt to draw a sharp distinction between real trust, on the one hand, and merely acting as if one trusts, on the other.[132] TMP, at least for such views, then consists in:

---

[131] For responses on behalf of the DV, see Hieronymi (2008); Keren (2014), (2020).
[132] See Hieronymi (2008), for example.

> …explaining why this difference should at all matter to us: What is the additional value of relations involving trust when compared with those involving merely acting as if one trusted without believing that the person is trustworthy? Why should we be interested in the very distinction? Is a relation in which I act trustingly because I believe that you are trustworthy more valuable than a relation in which I act in exactly the same way, without having such a belief? If it is indeed more valuable, why is this so? If not, then does not mere reliance, without trust, allow us to enjoy the very same goods that trust allows us to enjoy? In that case, it would seem difficult to maintain that trust is indispensably valuable. (Keren 2020, 118)

If one can merely act *as if* one possesses a trusting belief, and in so doing secure the goods associated with trust, then it is not at all clear what value is added by a trustor's possession of a *bona fide* trusting belief. Following Keren's line, then, though TMP is a problem for all accounts of trust's ontology, it looks particularly pressing for the DV.

§3 Interpreting Trust's Meno Problem

The remainder of the paper is dedicated to arguing that, if there is indeed reason to abandon the DV, it is not on grounds of TMP. However, TMP, much like the original Meno Problem, admits of numerous possible interpretations. The purpose of the current section is to sharpen the problem by identifying its most challenging interpretation - the interpretation that poses the biggest threat to the DV. For, if the DV can solve even this interpretation, then we will have made a good case that there's nothing to see here. Pritchard (2010) helpfully distinguishes between distinct tiers of Meno Problem for knowledge: the *Primary*, *Secondary* and *Tertiary* Meno Problems, as well as the closely related *Swamping Problem*. All of which pose different challenges, and all of which, I believe, have trust analogues which are overlooked in Keren's challenge.

§3.1 Trust's Primary Meno Problem

The minimal reading of the original Meno Problem is that of explaining why knowledge is more valuable than mere true belief. Pritchard calls this the *Primary Meno Problem*. The key thought here is that if we cannot even explain why knowledge is more valuable than mere true belief, then the very project of accounting for the value of knowledge is a lost cause (Pritchard 2010, 6).

Translating this into trust's terms, we see that herein lies a simple trust-analogue. Straight out of our guiding trust intuitions falls the following, which we might call *Trust's Primary Meno Problem:* that of explaining why trust is more valuable than mere reliance. If, in an attempt to explain and accommodate Trust Intuition 1, one were forced to abandon Trust Intuition 2, then one would have fallen at the first hurdle, as it were.

§3.2 Trust's Secondary Meno Problem

Pritchard, however, points out that if one were to solve the Primary Meno Problem for knowledge by pointing out some value-conferring property X, then another problem immediately arises just

in case there are Xs such that mere true belief plus X is nevertheless insufficient for knowledge. The reason for this, according to Pritchard:

> …is that if the distinctive value of knowledge is due to some feature of knowledge which, with true belief, falls short of knowledge, then it seems that what we should seek is not knowledge as such, but rather that which falls short of knowledge […] But if that's right, then why do we regard knowledge as distinctively valuable at all? (2010, 7)

This problem is pertinent in the context of the original Meno Problem because there are states more valuable than mere true belief that are nevertheless still not knowledge - justified true belief, for example. The *Secondary Meno Problem,* according to Pritchard, thus consists in explaining the value of knowledge not merely over-and-above true belief, but anything that falls short of knowledge.

What is important to note, for our purposes, is that the Secondary Meno Problem, too, will arise for trust as it does for knowledge. The reason for this is that, just as with mere true belief and knowledge, there are attitudes more valuable than mere reliance which are nevertheless not trust. Take *reasonable* as opposed to *unreasonable* reliance, for example. I might either reasonably or unreasonably rely on somebody to do something. Plausibly, when I do the former, my doing so has more value than when I do the latter.[133]

If this is right, then even if we could provide an answer to Trust's Primary Meno Problem, we would immediately face what we may now call *Trust's Secondary Meno Problem*. Here it is: if one solves the Primary Meno Problem by citing some additional value-conferring property X in their analysis of trust, then whilst they may have explained the value of trust over-and-above mere reliance, they have not yet explained the value of trust over-and-above *that which falls short of trust.* Neither reasonable nor unreasonable reliance are instances of trust (to see this, note that, unlike trust, neither need incur any vulnerability to betrayal) but that does not mean that the former is not more valuable than the latter. If the distinctive value of trust is due to some feature of trust which, when combined with mere reliance, falls short of real trust, then it seems that what we value is not trust as such, but rather that which falls short of trust (i.e., mere reliance plus value-conferring property X, in this case reasonableness). But, of course, if that is correct, then we have failed to accommodate Trust Intuition 1, that trust has *distinct* and *indispensable* value. (Pritchard 2010, 7)

§3.3 Trust's Tertiary Meno Problem

It seems that in the case of trust as in the case of knowledge, both primary and secondary iterations of the Meno Problem require a solution if we are to adequately account for their respective value.

---

[133] One may resist the analogy here by pointing out that while JTB is not sufficient, it is nevertheless necessary for knowledge, whereas on the other hand, reasonable trust is not even necessary for trust. After all, X may plausibly trust Y to ϕ even if they do so unreasonably. Now, while it is indeed possible to unreasonably trust, what's important to point out is that in doing so, one will have taken up an attitude less valuable than full trust, i.e., the attitude whose value we seek to explain. The attitude that has distinct and indispensable value, and which needs to be distinguished from those attitudes that fall short of having such value, is not merely any old trust; it is, I contend, reasonable, or well-placed, trust. My thanks to an anonymous reviewer at *Philosophical Psychology* for pressing me on this point.

Pritchard, however, points out a further complication, which arises when we note that even an answer to the Secondary Meno Problem does not guarantee a satisfactory account of the distinctive value of knowledge. The reason for this, according to Pritchard, is that the Secondary Meno Problem leaves open the possibility that the difference of value at issue between knowledge and that which falls short is merely one of *degree* rather than of *kind*. (Pritchard 2010, 7) To characterise the difference as a mere difference in degree, according to Pritchard, is to place knowledge on a kind of a continuum with that which falls short. And to do so, the thought goes, is to fail to capture the *distinctiveness* of the value knowledge seems to us to have. (Pritchard 2010, 7-8) After all, if the difference is merely one of degree rather than of kind, then why the long history of epistemological focus on this particular stage of the continuum? (Ibid., 8) Accounting for our intuitions about the value of knowledge over-and-above that which falls short thus requires an explanation that spells out the difference as one of *kind* rather than of *degree*. Call this the *Tertiary Meno Problem*.

Now, trust's analogues to the primary and secondary interpretations of the Meno Problem fell straightforwardly out of Intuitions 1 and 2. Whether or not trust has an analogue to the Tertiary Meno Problem, I think, is less obvious. Perhaps we are let off the hook by the fact that, as we saw at the outset, trust is considered in some sense a *kind of* reliance (Goldberg 2020). Perhaps this gives us reason, then, to focus on trust's Primary and Secondary Meno Problems, and ignore the Tertiary interpretation.

Perhaps. But I think that, on closer inspection, there is at least some prima facie reason to worry that trust has its own version of the Tertiary Meno Problem, too, which would be left unsolved were we to focus our attention exclusively on the Primary and Secondary versions. Note that, even though it is true that trust is seen as a kind of reliance, there is some reason to think that this characterisation may be somewhat imprecise.

Support for this hypothesis comes from the fact that attempts to add that special component, X, to reliance in order that it transforms into trust are oftentimes attempts not merely to slightly tweak the attitude, but rather to transform its entire essence. So much so, one might think, that it may no longer make sense to conceptualise reliance, on the one hand, and reliance plus X, on the other, as mere tokens of the same broad type. Moreover, this is plausible on any competing account of the content of X. Look, for example, at what is supposed to happen to mere reliance once X is added such that X is considered even a contender in the literature. X must add a vulnerability to betrayal component - an evaluatively loaded concept - that is absent in mere reliance. It must also place non-negligible demands on both trustor and trustee that are absent in mere reliance. Moreover, X must transform mere reliance into the kind of attitude capable of playing the kind of role we commonly think that trust plays, i.e., the kind of attitude that can create, as Locke put it, the 'bond of society', or, on Bok's line, the atmosphere in which what we care about can thrive. Some have thought it plausible that trust facilitates or allows for the kinds of human attachments essential to our well-being (Harding 2011; Kirton 2020; Wonderly 2016). If this is right, then perhaps a more accurate reading is that the difference between trust and reliance - much like the difference between knowledge and true belief - is best conceptualised as a difference in *kind* rather than merely in *degree*.

Perhaps not. What is important to note for our current purposes is that, if this reading is at all plausible, then we run the risk of failing to adequately accommodate all of our intuitions regarding the value of trust by focussing on only the Meno Problem's Primary and Secondary interpretations. The reason for this, of course, is that they leave open Trust's Tertiary Meno Problem, whereas solving the Tertiary problem would also deal with the first two problems. Even if one is left unmoved by the observations in the previous paragraph, then, one will surely still agree that, all else equal, if one *could* provide an answer to the Tertiary Meno Problem, then this would be by far the preferable outcome. (Pritchard 2010, 8)

§3.4 Trust's Swamping Problem

On the way to addressing Trust's Tertiary Meno Problem, we must make one last stop to consider the much-discussed *Swamping Problem* for knowledge. As Pritchard (2010, 9) highlights, the Swamping Problem is a variant on the Meno Problem which is not necessarily dealt with in the process of dealing with the other interpretations. Given this, and so as to avoid leaving any stone unturned, best to establish whether it, too, has a trust analogue we ought to consider.

The Swamping Problem[134] can be illustrated with a common example, involving two excellent cups of coffee. Suppose these two cups of coffee are identical in every way - in flavour, appearance, quantity, price and so on. Supposing we value excellent coffee, it follows that we value excellent coffee machines. Notice, however, that once we have the excellent coffee, it is no longer clear that we care about where it came from. In other words, as long as we have the excellent coffee, the fact that it was produced by an excellent machine confers no additional value upon it. (Pritchard 2010, 9) Whatever value is conferred upon a coffee by virtue of its being the product of an excellent machine, this value is 'swamped' by the value conferred upon the coffee by virtue of its being excellent. The broad problem for theories of epistemic value is that if whatever component you add to true beliefs in order to transform them into knowledge (the property of being reliably produced, for instance) is only instrumentally valuable relative to the further good of true belief, then in cases in which you have a true belief, no further value is conferred by the presence of your added component. (Pritchard 2010, 10)

To translate once again into trust's terms, the analogue challenge for philosophical accounts of trust is as follows: if your added component, X, which transforms less-than-trust into trust, adds value that is instrumental relative to the goods associated with trust, then it is no longer clear what value is added by the involvement of trust in instances in which these goods are already present. The worry for philosophical theories of trust can be made more pertinent by noting that trust's value, when it is discussed, is indeed almost exclusively discussed in instrumental terms - in terms of the goods it produces, preserves, or maintains (McLeod 2015). Some spell out trust's value in terms of social goods, of cooperation, cohesion, and so on. Dimock (2020), for example, argues that: "If trust is a cooperation-enabling resource, it is an important kind of social capital in communities." (161) Others similarly argue that trust enables individuals "…to work together for common purposes in groups and organisations." (Fukuyama 1995) Moreover, and as a result of this, some argue that "high-trust" societies have stronger economies and stronger social networks

[134] See Jones (1997), Swinburne (1999, 2000), Riggs (2002), Kvanvig (2003), Zagzebski (2003), Greco (2010), for critical discussion of the Swamping Problem.

in general than "low-trust" societies. (Fukuyama 1995; Inglehart 1999) What is important for our purposes is that whichever extant account of trust's value is correct - and there may be elements of truth in all of them - then, at least insofar as this value is instrumental relative to trust-associated goods, it will immediately give rise to Trust's Swamping Problem.

Suppose that the value of trust is relative to goods pertaining to, roughly, social cohesion. Now, further supposing that we can generate and maintain social cohesion through alternate methods, it is no longer clear why trust's value is *distinct* and *indispensable*. In other words, we would have failed to accommodate Trust Intuition 1. Perhaps it is not; perhaps the goods of trust cannot be generated and maintained in the absence of real trust. But the case for this has not been made, and would need to be, lest we leave Trust's Swamping Problem unsolved.

Now, while it is possible that in handling Trust's Tertiary Meno Problem we will have, in so doing, handled Trust's Swamping Problem, it is not a given. Suppose we show that trust's value over-and-above that which falls short is higher not merely in degree but also in kind. Still, if trust's value is instrumental relative to trust-associated goods, then whenever these are present, the kind of value it has will be lost. Trust's Swamping Problem is thus a distinct problem worthy of further investigation. But, perhaps more worryingly for our present purposes, it looks to sit very nicely with Keren's original reasons for thinking that Trust's Meno Problem was troublesome for the DV in particular. Recall that the problem for the DV consisted in explaining why the difference specifically between holding a trusting belief, on the one hand, and merely acting as if one held one, on the other. Note the following passage, from Keren, in particular:

> Is a relation in which I act trustingly because I believe that you are trustworthy more valuable than a relation in which I act in the exact same way, without having such a belief? If it is, why? *If not, then does not mere reliance provide us with the very same goods that trust does?* If so, trust doesn't seem indispensably valuable. (2020, 118, my italics)

The thought, then, is that one may realise the goods relative to which trust is thought to be valuable by merely relying, just so long as one acts as though they had a belief. We now see, more precisely, the way in which Keren thinks that the worry is one which plagues the DV in particular. If you fill in the X in the equation 'mere reliance plus X equals trust' in the way that the proponent of the DV is apt to, then you must explain why this attitude has more value than another which could, at least in theory, produce the same goods.

In what follows I set aside whether Keren is right that this problem is uniquely worrisome for the DV. Let us grant that it is. Suffice it to note that the disambiguation project of this section has delivered two major findings: (i) if one can solve Trust's Tertiary Meno Problem, one will thereby have solved trust's Primary and Secondary Meno problems; but that, (ii) in so doing, one will not necessarily have solved Trust's Swamping Problem. In order to properly address Keren's challenge then, the proponent of the DV must provide an adequate solution to both Trust's Tertiary Meno Problem, and Trust's Swamping Problem. It is this task to which I now turn.

§4 The Final Value of Trust

One of the main objectives of this paper is already secured: that of disambiguating Trust's Meno Problem, to which Keren (2020) gestures but does not explicate. We now see that it is subject to analogues of Pritchard's distinctions between Primary, Secondary, and Tertiary interpretations, as well as the related Swamping Problem. In this final section I argue that the DV can meet even the most challenging versions of Trust's Meno Problem, which we have identified as both Trust's Tertiary Meno Problem and Trust's Swamping Problem. In order to do so, my strategy will be to explicate the oft-overlooked notion of trust's *non-instrumental* value. In what follows, I will grant that trust's associated goods could indeed be attained by merely *acting as if* one trusted, as opposed to really trusting. Still, I will argue that trust has non-instrumental value, value not derived from its ability to produce further goods, and that the DV does a good job of explaining this value.

§4.1 Two kinds of non-instrumental value

One may be forgiven for thinking that the argument that trust has non-instrumental value is off to a shaky start given that, for one thing, we have built in success conditions to our exploration of its value. In other words, given that we noted at the outset that a plausible account of trust's value need only explain the value it has *when it goes well*, does this not in itself rule out the possibility that trust *per se* has any non-instrumental value?

This worry can easily be addressed by noting a common distinction in the broader axiology literature between two kinds of non-instrumental value: *intrinsic* and *final*. Whilst the fact that the value of trust is hostage to certain success-related constraints might rule out its having *intrinsic* value, it does not rule out its having *final* value. Some good is said to have final value if it is valuable for its own sake - and thus non-instrumentally valuable - but is so in virtue of a *relational* rather than *intrinsic* property. Two intrinsically identical items might, therefore, differ in value in virtue of their relational properties. The common example is of Princess Diana's Dress. (Rabinowicz and Ronnow-Rasmussen 2000, 41) A dress that belonged to Princess Diana might well be considered more valuable than an identical dress that did not. This value cannot be accounted for in intrinsic terms given that they do not differ, we may suppose, in any intrinsic properties. But neither can it be account for in instrumental terms. The value Diana's dress has is not derived by the goods it produces, or what it allows you to do; rather, its value is owed to its relation to Diana.

What is important to note is that the kind of trust whose value we seek to explain is, as we have noted, a *three-place, interpersonal* relation, which holds between two individuals (trustor and trustee) and an action (ϕ). There is thus scope for this kind of relation to have non-instrumental, yet non-intrinsic, final value. The attitude whose value we seek to explain might possess its distinctive value in virtue of its relational rather than intrinsic properties. The value of these relational properties, which obtain between two people and an action when all goes well, might be what distinguishes this attitude from attitudes that fall short.

§4.2 Why think that trust has final value?

If it turns out that trust has final rather than merely instrumental value, then we have a promising new route for solving the most challenging versions of Trust's Meno Problem, or so I will argue. First thing's first, however: why think that trust has such value?

As we have seen, the little discussion there is of the value of trust has focussed almost exclusively on its *instrumental* rather than *non-instrumental* value. Is this, we may ask, because the thought that trust has non-instrumental value is prima facie implausible, or counterintuitive? Not at all. To the contrary: though underexplored in the literature, I think the notion of trust's having non-instrumental value is, in fact, highly intuitively plausible irrespective of your preferred ontology of trust, and stands up on closer scrutiny. Here I elucidate and defend this idea.

Here is a plausible thought we find in Hawley (2012; though we can see similar ideas elsewhere, e.g., O'Neil (2012: 311) and D'Cruz (2019)): trust constitutes a kind of compliment, or respect, to the trustor; distrust, on the other hand, constitutes a kind of insult, or disrespect. Here is Hawley:

> Some of the benefits and harms generated by trust and distrust are practical – affecting our health, wealth, or enjoyment of life. But some strike more directly at who we are, at our personal integrity and autonomy. Distrust is an insult, not just an inconvenience; correspondingly, trust is a compliment. I want my friends, family, and colleagues to trust me, not just because that will make things go more smoothly for me – though it will – but also because I value their high opinion of me, and their level of trust is a key part of their opinion. (2012: 13-14)

Here is another passage we find elsewhere in Hawley's work:

> Trust is not ethically neutral. In the right circumstances, we expect others to trust us, and it is an insult to be distrusted without good reason. Attitudes of trust and distrust carry an evaluative message, and can have far-reaching consequences of great ethical significance. Trustworthiness is an ethically admirable trait, contrasting with both dishonesty and flakiness… (2019: 2)

There seems to me to be something deeply important in these insights from Hawley: that trust, at least when well-placed, pays a kind of a compliment to the trustee. Consider the following case:

> **BAR** Claire is the manager of a bar. She has equal reason to trust her two assistant managers: Lilith and Angie. Neither has ever gone over-and-above for her; but nor has either of them ever let her down. However, for whatever reason, when it comes to important matters, Claire finds herself trusting Lilith and not Angie.

Following Hawley's insights, I think that Claire's trust conveys a kind of a compliment to Lilith. She deems Lilith worthy of trust, at least when it comes to important matters, and that seems to be to say something important, and evaluatively positive, about Lilith as a person. Of course, Lilith may not care about, or even notice, this message. But that does not mean that Claire's trust doesn't carry it.

Let us further suppose that there are no positive downstream effects for Lilith of having been trusted. There is no opportunity for promotions down the line, no benefits of having gained Claire's trust. We may even suppose that doing so will, in fact, be an inconvenience, all things

considered. Nonetheless, I think that, in such a case, Claire's trust alone pays Lilith a kind of a compliment not paid to Angie.

Now, we should be careful not to overstate the point here. It would, of course, be an overstatement to suggest that trust must necessarily leave the trustee feeling flattered, or anything similar. However, I think it is very intuitive that at least part of what sets trust apart from that which falls short is the kind of evaluative compliment that it comes with. Given that trust comes with a vulnerability to betrayal, it follows that it comes with a kind of uncertainty. X is not vulnerable to betrayal if she has a guarantee that Y will indeed ϕ. The kind of compliment which I think trust embodies, extrapolating from Hawley, is that, under this uncertainty, X deems Y to be the kind of person who will indeed ϕ. This, I think, says something positive about Y qua individual. This is the sense in which Claire pays Lilith, and not Angie, a compliment.

§4.3 Final Value and Trust's Meno Problem(s)

We have seen, with the help of Hawley (2012; 2019), that it is plausible that trust's value is not merely instrumental, but also *final*. If this is right, then a straightforward road to solving Trust's Meno Problem(s) suggests itself. Before demonstrating that this road is open to proponents of the DV in particular, let me firstly explain what it looks like exactly.

In order to see why trust's having final value solves the problem, let us return to the two most challenging interpretations of the problem. Firstly, we have Trust's Tertiary Meno Problem, which we found to be worthy of our attention given that, if any view on trust could solve it, they would thereby have also solved Trust's Primary and Secondary Meno Problems. Now, the reason that our findings in §3.2 help with this is that, if trust's value consists - at least partly - in the compliment it embodies, then, at least so long as your view of trust can explain this value, it looks as though your view can differentiate trust from that which falls short in terms of a difference in *kind,* rather than merely degree. An attitude which embodies a compliment about another individual qua individual, I contend, looks like a different kind of attitude from one that does not. If trusting involves paying a compliment, then it is not clear that merely dialling up the degree of your reliance will ever do the same.

Now for Trust's Swamping Problem, which looked to line up particularly well with the challenge to the DV as Keren sets it out. For this, the solution in terms of trust's final value is more straightforward. Trust's Swamping Problem, recall, demands an explanation of the distinctive and indispensable value of trust that is not swamped by the value of having the goods associated with trust already in one's possession. This was particularly pertinent to the DV, recall, given that it seemed as though one may *act as if* one held a trusting belief - and thereby secure the goods associated with trust - without really holding one. However, if it is true that trust has final value, then even if one could secure the goods associated with trust by adopting some attitude that is less-than-trusting, one would have failed to thereby secure the full value of trust. There would, on such an explanation, be value over-and-above trust's instrumental value.

§4.4 Can the doxastic view account for trust's final value?

What has been said so far lends support to the idea that trust has final value. Moreover, nothing so far is theory-laden; it does not lend itself to any particular view of trust. However, while our answer is not yet theory-laden, our question pertains to the DV in particular. The final piece of our puzzle thus requires exploring whether the DV can account for trust's non-instrumental, final value.

Consider the following case, designed to show that it can.

> **SCIENCE LAB** Simon is in charge of a lab of scientists. Due to gender bias, Simon trusts the men in his group far more than any other gender identity. However, Simon recognises that instilling an atmosphere of mutual trust is beneficial for the continued success of the lab, which, in turn, will reflect kindly on his own reputation and career. As a result, though it is never manifested in his behaviour, when he sets a task for the lab, he believes that the men will fulfil it and does not believe that anyone else will.

Here is the thought: irrespective of what Simon believes, he does not pay the characteristic trust compliment when he merely acts as if he believes; he does so only when he really believes.[135] As a result, the relation between Simon and those whom he believes to be trustworthy and those whom he does not is different (and lesser) in evaluative character. This thought can be brought out by considering a scenario in which, for example, the women in his group found out that he did not really believe them to be trustworthy. Suppose they confronted him, questioning why he failed to trust them when he trusted the men on the very same evidence regarding their respective competence. We can expect that they would be decidedly unsatisfied were he to defend himself by pointing out that he never allowed this to manifest itself in his behaviour. Irrespective of his behaviour, the attitude he takes up toward them is different than the one he takes up toward the men.

Moreover, this difference cannot be explained by reference to the fact that the women have now found out, and are thus insulted, or offended. Though this is likely to also be true, the crucial point

[135] In a recent paper, Bieber and Viehoff (2022) briefly address a similar question, namely, why trust-without-belief cannot serve the characteristic function(s) of trust. They argue (2022, 20) that while it makes sense to *say* that someone trusts someone else despite lacking a trusting belief about that individual, this is not the norm, or paradigm case. The reason it makes sense to say this, according to Bieber and Viehoff, is that convincingly acting as if one has a trusting belief can serve (at least some of) trust's functions. However, that this will fulfil trust's function(s) is parasitic on the paradigm case involving bona fide belief. This is so given that trust-without-belief will not be able to fulfil trust's function unless it successfully passes for trust-with-belief, and that its doing so is parasitic on trust-with-belief being the norm, or paradigm, case. If people always acted as if others were trustworthy, without believing, the practice would collapse. While this addresses an important and closely related question, my focus here is slightly different. Firstly, while I have argued here that trust characteristically involves a kind of compliment, I have not argued that this is part of the *function* of trust. Some item P can have value (in particular non-instrumental value) in virtue of some property X without X having anything to do with P's function. Diana's dress has value in virtue of its relation to Diana, but it does not function to manifest such a relation. Of course, it is an interesting and pertinent question how the function of trust relates to its value, but it is not one I have addressed here. Secondly, and relatedly, it is not clear whether, for Bieber and Viehoff, non-paradigmatic trust is nevertheless *bona fide* trust. Presumably, given that they aim to characterise trust along functionalist lines, then simply fulfilling trust's characteristic functions is sufficient for bona fide trust. However, given that I do not share this functionalist ambition, my position is unambiguous on this matter: insofar as trust does not embody the kind of compliment I have explicated here, it is not bona fide trust.

is that they have a legitimate complaint regarding the initial state of affairs. Their complaint regards the attitude he adopted toward them, as is brought out by comparison with that which he adopted toward the men. Furthermore, this difference is also not explained away by the downstream effects of his trust because, we may suppose, there are none. We may stipulate that no concrete difference in outcomes whatsoever results from Simon's trusting only the men. Does this evaporate the difference in the attitude he adopts toward everyone else when compared to that which he adopts toward the men? It doesn't seem that it does. In such cases, the disappointment of the employees is best explained by Simon's lack of trusting *belief*, rather than of trusting *action*. He did successfully act as if he trusted, what he failed to do was believe.

§5 Conclusion

As Hawley put it, we want our friends, family and colleagues to trust us not merely because of the downstream effects, but also because we value their high opinion of us. (2012: 14) Here I have argued that it is in this high opinion - in the form of a kind of compliment - that trust's final value lies, and, moreover, that the DV can give a good account of such value. When we ask our friends, family, and colleagues to trust us, what we want is for them to *believe that we are trustworthy* with respect to the ϕ in question, not merely to act as if they so believe. It seems, then, that what we value essentially involves no less than a trusting belief; a would-be trustor who merely acts *as if* they have a trusting belief fails to pay the sort of compliment associated with full trust. If this much is right, then it looks as though the proponent of the DV can give a satisfying account of the distinctive and indispensable value of trust. To be clear, the purpose of this paper was not to argue that the DV is the only view that can provide such an account, or even that it necessarily provides the best one. Rather, the objective here was more modest: to show that, if there is a reason to abandon the DV, it is not on grounds of Trust's Meno Problem.[136]

[136] Many thanks to Adam Carter, Rory Aird, and Katie Moody, without whose help and support I would not have managed to write this chapter. Thanks also to Jesús Navarro, Tim Kearl, Dario Mortini and Aarón Álvarez-González for helpful discussion. Finally, I am extremely grateful to an anonymous reviewer at *Philosophical Psychology* for their thoughtful and constructive comments.

Chapter VI.

The generality problem and the C-Typing solution

Abstract: A promising new solution to the generality problem for reliabilism, Titus' (2017) C-Typing Solution (CTS), carves out a key role for credences. According to the CTS, the type of process relevant for justification is identified by the features of the scenario to which the agent is sensitive in forming their credence in the belief. In this paper I argue that CTS fails. In §1 I detail the CTS, drawing attention to a particular claim upon which it relies with respect to the way in which credences are formed. This claim - the repeatability claim (RC) - is that credences are formed systematically, in response to repeatable features of our interactions. In §2 and §3 I put pressure on RC with a series of counterexamples in which credences are instead determined by unique, unrepeatable factors of our particular interactions. §2 focuses on cases involving gut feelings and §3 focuses on cases involving implicit bias. The upshot is that the proponent of CTS faces a dilemma: they must either include or exclude such factors from the relevant type-specification. Opting for the former risks rendering the thesis insufficiently reliabilist, while opting for the latter risks rendering the thesis ad hoc. Either way, the solution comes up short.

§1 Introduction

Reliabilists hold that doxastic justification is a matter of the reliability of the process that causally produced, or sustains, the belief in the agent, where the reliability of a process consists in its tendency to produce true rather than false beliefs (Goldman 1979: 10).[137] The generality problem (TGP) for any such view arises once it is noted that the concept of reliability applies only to process-types, not process-tokens. We can get a better grip on the problem by considering a non-epistemic case. Consider a basketballer's jumpshot. One particular shot cannot be either reliable or unreliable; it is simply successful or unsuccessful on this occasion. Only her *type* of shot can be reliable or unreliable. An individual shot, however, might fall under any or all of the following type-descriptions: shot after a long warm-up, shot while sober, shot wearing a vest, and so on. And each of which will, presumably, have a different reliability score.

[137] Alvin Goldman's "What Is Justified Belief?" (1979) is widely considered the seminal work of reliabilism about justification. For further developments, see Goldman (1986; 2012); Kornblith (2002); Lyons (2009). For influential applications of the view to trust and testimony specifically, see Goldberg (2010) and Graham (2015).

Similarly, when it comes to beliefs, there are a myriad of factors that play some minimal causal role in the formation of any particular belief. Any token belief-forming process can thus be characterised in terms of numerous different types. Take Conee and Feldman's example of Smith (1998: 1).[138] Smith has good vision and can tell the difference between different kinds of common species of tree. Having seen one from his window, Smith comes to believe that there is a maple tree near his house. All else equal, Smith's belief is justified, but notice: reliabilism gets this result only if the process that caused the belief is reliable. The process that caused Smith's belief, however, was a particular set of concrete, neural events involving light reflecting off the tree and onto his eyes, Smith's recognising what he sees as a particular species of tree, and so on; in other words, a unique token belief-forming process. But, as has been noted, tokens cannot be (un)reliable, only types can, and we might describe this particular token in terms of numerous different types, e.g., forming a belief based on visual perception, inferring tree species from shape of leaves, and so on. We need to know, then, which description of the type (hereafter, 'type-specification') needs to be reliable in order for this belief to count as justified.

Conee and Feldman lay out three requirements any adequate solution will need to meet:

(1) **The solution must be principled and non-ad hoc**. With multiple variations of belief-forming processes, it would be easy enough for every token to select the most intuitive type on an ad hoc, case-by-case basis (Conee and Feldman 1998: 4). A satisfactory answer, however, must be complemented by a substantive account of what makes the type reliable.

(2) **The solution must be intuitive**. In other words, the solution must make intuitively correct epistemic classifications. We are not searching for just any single type-specification among many, it needs to be the correct one.

(3) **The solution must remain true to the spirit of reliabilism**. The relevant type-specification must at least implement the basic idea that it is indeed the *reliability* of the process that confers the status of justification (*Ibid*.: 4). That is, we require a reliabilist solution to a reliabilist problem.

Titus (2017) offers a solution - dubbed the *C-Typing Solution* (CTS) - that purports to meet all three requirements. What is distinctive about CTS is that it represents what we might call a *hybrid* solution. A hybrid solution sets out not only to solve TGP for reliabilism, but to do so in a way that bridges an important gap between externalist and internalist theories of justification. Titus is compelled by the internalist's complaint that externalist epistemologies often paint a picture of doxastic justification such that the agent merely has processes inside of her, and if she's lucky, they're reliable and she counts as justified; if she's unlucky, they're unreliable and she doesn't. Titus agrees with the internalist that this is a thin sense in which the agent herself is justified, which is what we initially set out to theorise. (*Ibid.:* 109) For Titus "…justification is a property possessed by people *qua inquirers* in the first instance, and so […] whatever belief-formation types are relevant

[138] The generality problem for reliabilism was originally anticipated by Goldman in his seminal (1979), but was given its most thorough and systematic treatment by Feldman (1985) and Conee and Feldman (1998). I focus here on the latter formulation so as to remain consistent with Titus (2017).

to justification will connect up to the subject as epistemic agent." (Titus 2017: 110, italics in original)

An adequate hybrid solution, then, will respect the tight connection between doxastic justification, on the one hand, and epistemic agency, on the other. For Titus, *credences* are the key theoretical tool the reliabilist needs for such a solution. For any candidate belief, there is always a fact of the matter regarding what it was that caused the agent to form their credence.[139] How an epistemic agent forms her credence is, for Titus, what determines the epistemically relevant process type for that belief. The type-specification relevant for doxastic justification is thus individuated by whatever features of the scenario the agent was, as a matter of fact, sensitive to in assigning their credence to the belief.

In this paper I argue that the CTS faces a troublesome dilemma, which has ominous implications for any hybrid solution to TGP. Here is how I proceed. In §1 I lay out the CTS, drawing attention to a particular claim upon which it relies regarding the way in which credences are ordinarily formed. The claim, which I call the repeatability claim (RC), is that credences are formed systematically, in response to general, repeatable features of our interactions. In §2 I put some initial pressure on RC with discussion of what I call *gut feelings* cases - cases in which our credences are determined not by general, repeatable features, but rather by feelings which are provoked by unique, situational factors. The pressure such cases place on CTS comes from what I call the *specificity problem* for reliabilism. As Goldman put it: "… if a type is specified so narrowly that only one instance of it ever occurs, for example, then it will be either entirely reliable or entirely unreliable." (Goldman 1979: 12) In §3 I put further pressure on CTS through discussion of another situational factor which will affect credence-formation: implicit bias. Implicit biases, much like gut feelings, will inevitably factor into the everyday credences of ordinary epistemic agents, and will do so in ways that are not responsive to general, repeatable features of our interactions. The upshot is that the proponent of CTS faces a troublesome dilemma: type-specifications must either include or exclude such situational factors as gut feelings and implicit biases. Opting for the former risks rendering the thesis insufficiently reliabilist, thus jeopardising the proposal with respect to (3), while opting for the latter risks rendering the thesis ad hoc, thus jeopardising the proposal with respect to (1). In §4 I conclude by discussing the prospects for similar such hybrid solutions to TGP.

§2 The C-Typing Solution

§2.1 The C-Typing Solution

Whenever a subject has a belief, we can always ask what determines their credence in said belief. For every belief, there will be a fact of the matter about which features of their environment or interaction the agent was sensitive to in assigning their credence (Titus 2017: 111). The relevant process-type for any belief, then, or *C-Type*, can be individuated simply by whatever features the agent was in fact sensitive to in assigning their credence to this belief. Here is how Titus defines her view:

[139] Following Titus' use of the term, 'credence' is here intended simply to mean 'degree of confidence'. See Titus (2017: 111).

> **C-Type**: X is the C-type for the token way Y that a subject S forms or retains a belief B(p) in P iff X is the type that is individuated by exactly the features of Y causally relevant to the formation or retention of S's credence C(p)
>
> **C-Typing Thesis**: For any belief B(p) held by any epistemic agent A, the way Y in which B(p) is formed or maintained is a token of a unique C-Type X, as defined above. X is the type of belief-forming process that determines Y's reliability relevant to epistemic justification. (Titus 2017: 112)

It is important to note that the features one is sensitive to when forming credences need not be explicit or conscious considerations – credences are often formed in subtle, automatic ways. (*Ibid.*: 112) The features that determine the C-Type are those that, as a matter of fact, determine the agent's credence, rather than those that are subjectively accessible to her.

§2.2 Repeatability claim

It is crucial for CTS that the features we are sensitive to when forming our credences are not unique to each and every scenario we find ourselves in. (Titus 2017: 112) As Titus puts it:

> "Although we often form degrees of confidence in claims about particulars on the basis of token interactions with those particulars, what credence we form is determined by repeatable features of belief-forming processes that are instantiated on that occasion." (Titus 2017: 112)

Suppose I ask you why you are so confident that there is milk in the fridge, and you reply that there was plenty left this morning and that you are the only one who has been home. Though you cite particular features that are unique to this specific inquiry (when you last checked the fridge, and who has been in your house today), you recognise these as particular instances of more general, repeatable features which indicate how likely it is that a belief formed *in this way* will be true. Your degree of confidence is high because you recognise that believing on the basis of perception combined with some basic induction is a reliable way to form beliefs. Here is Titus again:

> "Even where particular interactions are crucial to credence-formation, it is highly plausible that we systematically form credences on the basis of general features of such interactions, features we take to indicate how likely a belief is to be true." (Titus 2017: 112)

This descriptive claim - about the way in which credences are formed - is crucial for Titus's argument, so it is best to summarise it and give it a name:

> **Repeatability claim (RC)**: Our credences are formed systematically in response to general, repeatable features of our interactions.

RC is crucial for Titus because it is what makes credences useful for identifying type-descriptions. If our credences are formed systematically in response to general, repeatable features of our interactions, then in our sensitivity to such features, we ourselves are implicitly individuating types of belief-forming processes (Titus 2017: 112). If RC is true, then there will be a fact of the matter

about which token belief-forming processes would be such that the agent's credences would be formed in the same way, that is, through sensitivity to the same features of the scenario (Titus 2017: 112). Conceived of in this way, our credence-forming processes identify a unique type-specification for every token belief, which in turn can be assessed for reliability.

If I place low credence in a belief on the basis that it was caused by a process that often goes wrong - a perceptual belief in poor lighting conditions, for example - then just by assigning this low credence I have implicitly placed this token instance into a broader type. I am, in fact, assigning a low credence precisely because operating this type of belief-forming process has often led me to false beliefs. According to the C-Typer, what is happening in this kind of case is that I am recognising (consciously or subconsciously) particular process-tokens as instances of process-types and assigning my credence accordingly (Titus 2017: 112).

Let us now apply the CTS to our previous example in order to see how it works. Recall the example in which Smith, upon seeing one, forms the belief that there is a maple tree outside his house. The problem for the reliabilist consisted in specifying - in a principled and distinctly reliabilist manner - which type-description must be reliable in order that Smith's belief counts as justified. The relevant process-type, we now know, is not the one that caused the belief *per se* - that is, the complex set of neural events - but rather the one that caused Smith to form his credence in the proposition that there is a maple tree outside. If it was stipulated that the reason Smith is able to confidently differentiate between tree species is by recognising the shape of their leaves, then the process-type will be 'forming beliefs about tree species on the basis of the shape of their leaves'. This, presumably, will have a high reliability score and the belief will count as justified. If, on the other hand, it was stipulated that Smith simply took a momentary glance outside and guessed about the species of the tree, then the C-Type is 'forming beliefs based on a wild guess'. This process will have a low reliability score and his belief will not be justified.

As a solution to TGP, then, the CTS looks promising. Recall Conee and Feldman's requirements laid down at the outset, which set the bar in terms of what constitutes a satisfactory solution. At least prima facie, the CTS seems to meet all three. Firstly, it is principled and non-ad hoc. Given that, as Titus well puts it, there will always be a fact of the matter regarding the way in which one formed their credence in any given belief, CTS provides us with the conceptual tools with which we may type all token processes and assess for reliability, rather than doing so on a case-by-case basis. Secondly, it seems clear that it will get intuitive results. The CTS can straightforwardly get the result that Smith's belief about the maple tree will count as justified if predicated on a good process, as discussed, but unjustified if predicated on a bad one. Finally, it remains true to the spirit of reliabilism insofar as 'good' and 'bad' processes are here defined by reference to their reliability in bringing about true beliefs in agents.

§3 Counterexamples to the C-Typing Thesis

§3.1 Testimony

Promising as the initial results may be, the remainder of the paper is dedicated to demonstrating that the CTS fails. The reason, in short, is that I do not think credences behave the way Titus

supposes they do - or indeed needs them to. In order to demonstrate this, I present counterexamples to RC, the claim that credences are formed systematically in response to general, repeatable features of our interactions. As we have noted, this claim is important for Titus because, if it were true, it would show that just in assigning a credence to a belief we are implicitly typing the process that produced it. However, contra-RC, I contend that credences are often formed in distinctly unsystematic ways, in response to seemingly unrepeatable, unique features of our situations and interactions.

Consider the following case:

> **EDINBURGH** Jennifer is heading to Edinburgh for the first time and wants to visit the Scot's Monument. She meets an Edinburgh local, Ed, on the train and asks him for directions. However, upon receiving the directions, Jennifer is overcome with the feeling that she is being deceived. There is no reason for this feeling; Ed is not drunk, nor acting shifty. Moreover, she has never been deceived by someone who looks like Ed, nor deceived in any relevantly similar situation. In fact, on the face of it, she has every reason to think that Ed is a good source of testimony - after all, he's a local and has no reason to lie. Nevertheless, she cannot escape her feeling and assigns extremely low credence to the directions.

What's important to notice is that, in forming her low credence, Jennifer is only really sensitive to this feeling that she has. Perhaps her feeling is best characterized as emotional, intuitional, or simply as a gut feeling. But however we choose to characterize it, there are no general features present in the case indicating that the individual's directions are unlikely to be correct; it is but a feeling. If Ed had been drunk or acting shifty, then Jennifer's low credence would have lined up perfectly with RC; it would have been an instance in which the agent's credence is determined by their recognizing this particular instance as a token of a broader type, (that is, believing on the testimony of a drunk person) and assigning a corresponding credence to this token *qua* member of the broader type. However, this was not the case, and the problem this poses for RC should be plain to see. Plausibly, there is just a feeling we sometimes get, which does not arise in response to general, repeatable features of our interactions, but is such that when it does arise, we cannot help but have our credences guided accordingly. In such cases, our credences are not responsive to general, repeatable features of our interactions; rather, they are responsive to unique features of our particular situation.[140]

§3.2 Beyond testimony

Now consider a further case, designed to show that this point generalises beyond cases involving testimony.

[140] There might be interesting parallels here with the well-known situationist challenge for virtue theories. One reading of what is going on in that literature is that the situationist challenge forces virtue theorists to specify the virtues more finely than they would like. However, the parallel is imperfect, and I lack the space in the current paper to explore whether and to what extent there is overlap. For the situationist challenge as it pertains to virtue ethics in particular, see Doris (1998, 2002); Harman (1999, 2009); Merritt, Doris, and Harman (2010). For the most thorough exploration of the ways in which this challenge extends into epistemology, see Alfano (2012; 2013). See Alfano (2014) in particular for discussion of the challenge with respect to reliabilism. For helpful overview, see Turri, Alfano, and Greco (2021).

DREAM Julie is an average maths student. She understands the basic principles and can figure out the answers to basic sums when given the time. One night, however, Julie has a bad dream in which she sits down to begin her final year maths exam only to realise that all her numeric skills have escaped her. In the dream, basic sums look totally foreign to her and all of her attempts to work them out come up short. As a result, her confidence in her abilities in real life are significantly diminished. She cannot help but significantly lower the credence with which she believes the conclusions of her own mathematical workings.

Note that in neither case does the agent form their credence on the basis of general, repeatable features of the interaction which they take to indicate how likely their belief is to be true. Rather, they form credences based on a feeling which is, in an important sense, unique to each situation. All the general features of each case could have been exactly the same and yet the subject would have formed a different credence if she had not experienced this precise subjective feeling. Moreover, as DREAM demonstrates, this is not merely an idiosyncrasy of testimonial belief. Rather, this is just what our credences sometimes do.

Here is the problem: credences identify a promising solution to the generality problem only insofar as RC is plausible. If credences are assigned in response to general features of interactions consistently whenever these features are present, then these features will indeed specify a C-Type for the token belief-forming process at hand. RC, however, is subject to counterexamples. In particular, there seem to be cases in which credence-assignment is determined by gut feelings, which, in turn, arise in response to unique, unrepeatable features of our interactions. If, contra-RC, credences are in fact sometimes (perhaps oftentimes) formed on the basis of seemingly unique, unrepeatable features of our interactions, then so much the worse for the thesis that relies on them to solve the generality problem.

§3.3 Specificity

At this stage, one might wonder whether there is not a quick-fix solution for the C-Typer. Might we not simply specify C-Types very finely so as to account for the cases in which credences are guided by unique, unrepeatable features of our particular situations? After this minor modification, the process-type in cases like our examples will be *coming to believe on the basis of a gut feeling*. This, of course, can then be assessed for reliability like any other process-type.

However, the problem with this response, which we may call *the specificity problem,* has to do with the fineness of our type-descriptions. To see it, note that it is uncontroversial that reliabilists in general are under distinct pressure not to specify the process-type relevant for justification too finely. This point is nicely illustrated in the following passage from Goldman: "If a type is specified so narrowly that only one instance of it ever occurs, for example, then it will be either entirely reliable or entirely unreliable." (Goldman 1979: 12) Reliabilism's explanatory power, then depends upon the view's ability to provide relatively coarse type-specifications.

Notice, however, that incorporating the sorts of unique features at play in gut feelings cases will be no quick fix - it will not merely be a matter of slightly tweaking our initial specifications. Rather,

the myriad different gut feelings which will inevitably arise in different agents in different situations will have vastly differing effects on the overall reliability of the process. The C-Typer will thus need to make myriad further classifications to reflect the unique features of each and every individual case. Notice the looming tension with condition (3) above: the more narrowly one must specify the C-Type, the more one loses hold of the reliabilist intuition that what confers justification to a given belief is its being the product of a *reliable type of process*.

Notice that if it were the case that such feelings arose fairly systematically in response to what the agent took to be epistemically relevant stimuli in her environment (generic, tell-tale signs of deception, incoherence, and so on, for instance) then this issue would evaporate. But, I contend, this is simply not the way in which such feelings will tend to arise; such feelings will, oftentimes, arise in a way that is unique to each agent and their particular situation. **EDINBURGH** and **DREAM** should suffice to demonstrate this point.

The upshot is that the C-Type specification, at least for such cases, will be either wholly reliable, or wholly unreliable, depending on whether this particular belief is true or false. If it is true, then the C-Type will have a 100% reliability score, given that the C-Type reflects the unique, unrepeatable features of this particular situation. If, on the other hand, the belief is false, then it will have a 0% reliability score. Either way, condition (3) cannot be accommodated: a type of process that only ever occurs once can hardly accommodate the reliabilist intuition that what makes a belief justified is its being the product of a *reliable type of process*.

§3.4 Response

Let us consider a potential response on behalf of the C-Typer before proceeding. I have not yet taken a stand on whether the beliefs in our counterexamples are justified or not. But, the C-Typer might point out, my examples are *bona fide* counterexamples to the view if and only if they are cases in which either (i) the agent is intuitively justified and unreliably produced (thus generating a false negative); or (ii) the agent is intuitively unjustified yet reliably produced (thus generating a false positive). Might the C-Typer not simply avoid the conflict by insisting that my examples are both intuitively unjustified and produced by unreliable processes - just as the reliabilist would have it?

Not so fast. Consider **EDINBURGH** again, but now suppose that Jennifer was right - the doubt she felt regarding Ed's testimony turned out to be well-founded. Though he was indeed a local with no good reason to lie, he did so anyway, perhaps for his own amusement. In such a case, Jennifer's process-type is 100% reliable. It was not an instance of a general type of process, involving her response to a set of epistemically relevant features in the context of her belief-formation. Rather, her belief was the product of a unique, one-off gut feeling in a particular case. In such a case, and given that the belief was true, reliabilism generates a false positive: her belief is intuitively unjustified but reliably produced (indeed, as reliably produced as can be).[141] CTS thus collapses the distinction between justified belief and true belief in cases in which the agent's

[141] It is important to note here that I do not have a dog in the fight regarding the intuitive justifiedness of the belief. My point here is simply to demonstrate that the force of the counterexamples is not dissolved by insisting that the protagonists' beliefs are unjustified.

credence is responsive to unique, situational - as opposed to general, repeatable - features of her interaction.

Notice, moreover, an emerging blueprint for similar such counterexamples. All we need to do is imagine a case in which (i) contra-RC, the agent's credence is responsive to unique features of their situation; (ii) their belief is either intuitively justified yet false, or intuitively unjustified yet true. CTS will be unable to account for either justified false belief or unjustified true belief given that the reliability score of a type of process that only ever occurs once will, of course, be wholly determined by the truth or falsity of the belief it produced the one time it occurred. There is no scope, on such a line, for the C-Typing theorist to explain the justifiedness of a false belief given that the process that produced it must have been unreliable, and vice versa with respect to the unjustifiedness of a true belief. In a nutshell, the problem is that the CTS collapses the distinction between *true* belief and *justified* belief in cases in which an agent's credence is responsive to unique rather than general features of their situation.

§4 Counterexamples from implicit bias

§4.1 Implicit bias

In §2, I argued that gut feelings cases put pressure on the CTS with respect to the specificity problem, which consists in the view's ability to specify types relatively coarsely. In this section, through discussion of implicit bias, I argue that the view faces a dilemma: on the one horn, the specificity problem is further antagonised, on the other, the thesis looks ad hoc.

Broadly speaking, implicit biases are automatic cognitive or affective associations between (the agent's concept of) some group and some property or trait (Holroyd 2012).[142] Such biases have an array of well-known adverse effects on cognition, including the perpetuation of various injustices. Unfortunately, for reasons of space, I am unable here to delve too deep into the rich literature in philosophy and social psychology on implicit bias and the complex ways in which it affects our common attitudes, preferences and judgements.[143] Instead, in what follows, I simply take for granted some fairly uncontentious observations about the phenomenon.

The first of these that will be relevant for present purposes is simply that biases are *pervasive*, i.e., there is an abundance of evidence that we all have them - that they are not possessed merely by a cognitively or morally defective minority. Rather, as Sullivan-Bissett puts it, "…empirical work has shown that implicit biases are held by most people, even those who avow egalitarian positions, or are members of the targeted group." (2015: 549)

But as well as being pervasive, these biases are importantly *idiosyncratic*: the way in which they influence our attitudes, preferences and judgements is not homogeneous and stable. (Brownstein 2019) Multiple studies, for instance, have shown significant variation in individuals' scores on implicit association tests over a matter of days. (Cunningham et al. 2007; Devine et al. 2012; Cooley et al. 2014; Gawronski & Brannon 2017) Their variation and fluctuation, even within the same agent at different times, have led psychologists to characterise them as more 'state-like' than 'trait-

[142] For more on defining implicit bias, see Saul (2013).
[143] See Payne & Gawronski (2010), Brownstein and Saul (2016), and Brownstein (2019) for overviews.

like' (Brownstein 2019). In other words, their effects – and even their presence – are oftentimes at least partly determined by situational factors. What this means is that the possession of implicit bias will have no single, systematic effect on cognition; rather, the effects will vary, sometimes dramatically, depending on multiple further factors. In what follows, I take these observations - that bias is in these ways both pervasive and idiosyncratic - as read, rather than provide any further argument for them. I argue that, so conceived, implicit bias will inevitably factor into our credence-forming processes, and will do so in a way that raises a troublesome dilemma for the C-Typer.

§4.2 Quiz Show

Consider the following example:

> **QUIZ SHOW** Jack is stuck on the philosophy question in a quiz show. He has an option available to him whereby he may consult two friends. He texts his two equally-qualified philosopher friends, Joe, a man, and Jessica, a woman. He knows that they are both experts, but has no reason to believe that one is any better than the other. Now suppose that Jack has an implicit bias against women in philosophy. As a result, if Joe were to reply first, Jack would unhesitatingly form the corresponding belief and provide his answer to the question. If, on the other hand, Jessica were to reply first, Jack would be inclined to wait and double check what Joe has to say before committing. As it happens, Jessica gets back with the correct answer, and Joe never responds. Jack could take his money and leave, but he is sufficiently confident in Jessica's testimony to form the corresponding belief and answers accordingly.

The question is how best to describe such a case - and, more importantly, how to specify the relevant C-Type. Here is the puzzle for the C-Typer: Jack's bias causes him to wrongly assign less credence to Jessica's testimony than Joe's, but he nonetheless places sufficient credence in her testimony to form the outright belief that she is correct. Moreover, at least prima facie, Jack's belief seems like a good candidate for testimonial knowledge. After all, Joe failed to respond, so Jack's belief is ultimately a true belief based on expert testimony. At the very least, his belief is surely justified - to deny this would be to adopt a prohibitively restrictive notion of testimonial justification.[144] The CTS thus gets the right result here only if it gets the result that Jack's belief is produced by a reliable type of process, where type of process is defined by reference to its C-Type.

The dilemma arises, however, once we recall that C-Types are determined by reference to exactly the features of the situation causally relevant to the formation or retention of one's credence in their belief[145] and, in Jack's case, this recipe for identifying process types is ambiguous between two possible candidates: one which includes the bias, and one which omits it. In what follows, I explore each option in turn and argue that the C-Typer faces trouble either way. If implicit biases

[144] Note that even reductionism, the more stringent view in the epistemology of testimony, would count this belief as justified. Reductionists (Hume 1740; Fricker 1994; Lyons 1997) demand that a hearer be in possession of independent, non-testimonial reasons to trust what a speaker says in order that their resulting belief is justified - belief on the basis of mere say-so, a la anti-reductionism (Reid 1764; Burge 1993, 1997; Graham 2006, 2010; Simion and Kelp 2020) is never permitted. But notice that Jack *has* such positive reasons for Jessica's testimony; indeed, the very reason he asks her is that he knows her to be an expert in the domain of philosophy. Even on a conservative view of testimonial justification, then, Jack's belief is justified.

[145] See Titus (2017: 112).

are included, this further antagonises the specificity problem from §2, and risks rendering their thesis insufficiently reliabilist. If, instead, implicit biases are omitted, they risk rendering their thesis *ad hoc*.

§4.3 The first horn: including the bias

Let us first explore the C-Typer's first option, namely, specifying the C-Type so as to include the bias. The relevant C-Type would, on such a line, be something like: 'belief on the basis of expert testimony where implicit bias is involved.' Given Titus' explicitly stated ambitions of finding a solution that adequately reflects the features the agent herself treats as important "in going about her epistemic business" (2017: 111), the rationale for going this route might be as follows: what matters for justification is what matters to the epistemic agent, and the relevant features are thus simply those features that play a causal role in an agent's inquiry. Given their salient role in the formation of the belief under consideration, implicit biases are one such factor, and so they must be incorporated.

Recall, however, that in order to satisfy condition (3), the C-Typing solution must at least implement the basic reliabilist conviction that what explains justification is a belief's being the product of a *reliable type of process*. Accordingly, and as has been noted, the C-Typer faces a specificity problem. This problem loomed large for a C-Typer hoping to incorporate gut feelings into type specifications, but looms larger still for one who hopes to incorporate implicit biases as well.

To see how this is so, notice that incorporating the effects of implicit bias into our C-Types will not simply be a matter of positing one additional C-Type, that is, one further type of belief-forming process which includes implicit bias and which can be assessed for reliability. Such a result might be palatable. However, it is not available to the C-Typer. Rather, further, more fine-grained specifications will need to be made to accurately reflect all of the different ways in which implicit bias impacts the cognition of agents in their respective situations. Biases will not affect credences in one particular way whenever a certain set of general features are present. Rather, much like gut feelings, different implicit biases in different agents and different situations will have different effects on the reliability of the process. Some will doubtless have the effect of producing wildly irrational, false beliefs in their agent. But others might cause their agent to form true beliefs they otherwise would never have formed. Others still might for the most part remain dormant - largely inefficacious with respect to their credences. What seems clear, though, is that the nature of implicit bias is simply too heterogeneous to figure in an informative specification of a C-Type.

Fricker's (2007) oft-discussed notion of testimonial injustice (TI) will be instructive here. The dominant view since Fricker's seminal 2007 work was that TIs are unwarranted credibility deficits assigned to speakers as a result of prejudicial identity stereotypes. However, as Davis (2016) and Lackey (2018) have convincingly argued, TIs can be committed not only in cases of unwarranted credibility *deficit*, but also unwarranted credibility *excess*. Consider, for example, Davis' example of a group of American high-school students who, when stuck on a maths question, seek the help of their Asian-American classmate because they've heard that Asian-Americans are good at maths. What's important for our present purposes is that this is further evidence of the differential effects of bias upon credences and credibility assignment. Though the TI discourse is carried out in terms

of *prejudice* and *credibility assignment*, the road to *bias* and *credences* is straightforward. Identity prejudices will, of course, lead to biases of the sort we have discussed, and credibility assessments will determine credences (at least in straightforward cases of testimony). What Davis and Lackey's development of the notion of TI further demonstrate is that the effects of such biases on our credences, then, will not be simple, homogeneous, and consistent. Sometimes they will result in a lower credence - as in **QUIZ SHOW** - but sometimes they will result in a higher one - as in Davis' example of the high-school students. Clearly, then, one cannot treat cases of bias-inflected credence as one, unified, homogenous process-type.

The upshot of all of this for the C-Typer is that the first route out of the dilemma is closed to them. What looked like a plausible quick-fix, whereby they need only stipulate some slightly more finely-specified C-Types, is in fact far more complicated. Rather, C-Types will need to be specified sufficiently finely so as to capture the effects of this particular implicit bias I in this particular agent A in this particular situation S. Of course, once this is done, the specificity problem returns, and with far more bite than in its previous iteration.

§4.4 The second horn: excluding the bias

Given this, perhaps a more plausible route for the C-Typer would be to opt for the second horn of the dilemma, and specify the C-Type so as to *exclude* bias. The relevant belief-forming process will simply be 'forming a belief on the basis of expert testimony'. In opting for this route, however, the C-Typer undertakes the task of providing a principled explanation for omitting implicit bias from not *just* this C-Type, but *all* C-Types. After all, in order to avoid providing a blatantly *ad hoc* solution, and in order to satisfy (1), types cannot be specified on a case-by-case basis so as to get the right result.

The problem with this route is that it is exceedingly difficult to see what such a principled explanation might look like, in particular given the C-Typer's broader commitments. Recall that Titus is sensitive to the internalist's complaint that externalists often implausibly detach the agent herself from the question about justification. Titus wants the C-Typing thesis to capture the way in which the agent really does relate to the world in her inquiry, what she really is sensitive to in going about her epistemic business. But what Jack really is sensitive to in going about his epistemic business is, at least partly, features that trigger his bias. This is evidenced by Jack's notably lower credence in Jessica's testimony when compared to Joe's; implicit bias causes Jack to form a lower credence than he otherwise would. In omitting biases from C-Types, you omit a feature of the causal story of the belief that played a significant role in the agent's own cognition with respect to said belief. This was precisely what the view intended to capture, and so it is exceedingly difficult to imagine what the C-Typer could point to as the non-ad hoc reason for omitting such features. It seems, then, that satisfying (3) involves excluding biases from C-Types, but that doing so means failing to satisfy (1).

§4.5 Negligible factors

Perhaps this is too quick. Perhaps there is a route to excluding bias in cases such as **QUIZ SHOW** that is not ad hoc. After all, while we may all agree that Jack's biased process is in some important respect *bad*, the central question for the C-Typer is whether the process that produced Jack's belief

is reliable. If the particular way in which Jack's process is bad is in fact negligible with respect to the (un)reliability of his process, then we have on our hands a principled, non-ad hoc reason for omitting it from the C-Type. There are a myriad of minimally causally efficacious factors for any candidate belief, the C-Typer might retort, and so omitting all of those factors which play a merely negligible role is quite consistent with a theory whose ambitions are to capture those features which agents themselves treat as epistemically significant in the context of inquiry. It is just that negligible factors do not play an epistemically significant role.

Further support for this line of defence comes from the fact that Jack's bias in **QUIZ SHOW** does indeed seem to be one such negligible factor. To see this, note that if it was stipulated that Jack's bias impacted his belief-forming process in a way that, in fact, *did* render his process unreliable, then we would straightforwardly lose the intuition that Jack gains knowledge, or justified belief. Accordingly, it remains open for the C-Typer to escape the problem by including bias in C-Types only when they are causally efficacious with respect to the agent's credence, and excluding them when they are not so.

Unfortunately, I do not think that this helps much. Consider a slight tweak on our original case, call it **QUIZ SHOW***. In **QUIZ SHOW***, all the details are as is in **QUIZ SHOW**; owing to gender bias, Jack would have placed higher credence in Joe's testimony than Jessica's, had Joe responded. However, as well as his gender bias, Jack also has a racial bias which leads him to assign higher credence to the philosophical workings of white philosophers, when compared to non-white philosophers. As a result, in the tweaked scenario, Jack would have assigned even lower credence to a non-white woman than he did to Jessica, a white woman. In fact, we may suppose, if Jessica had been non-white, Jack would have withheld belief altogether.

In such a case, Jack's bias is causally efficacious, and thus cannot be ruled out as irrelevant. Notice, moreover, that the particular way in which it is efficacious - and thus to be distinguished from the non-tweaked case - is entirely determined by unique features of Jack's new situation. As a result, positing that the relevant C-Type in **QUIZ SHOW*** is something like 'belief on the basis of expert testimony where implicit bias is involved' is still not specific enough, and the specificity problem from the previous section resurfaces once again. What really explains his precise degree of credence is more precise, and situationally-dependent, than brute implicit bias; it is that in this case Jack's bias in favour of white philosophers cancelled out his bias against women philosophers such that his credence was still sufficiently high for outright belief. Moreover, this indeed seems to be the way that bias will impinge upon our agency; its effects will not be neat, stable, and systematic, nor should we expect them to be. Rather, given what we know about implicit bias, its pervasiveness and idiosyncrasy, it will be wholly at the mercy of peculiar, situational factors. Even within the same agent, biases will be interweaving and intersectional, as in our cases, crossing and clashing across class, gender, race, not to mention other, less ethically charged domains.

Given all this, while this route may allow the C-Typer to account for some cases (cases in which implicit bias, though involved, is causally negligible), it still admits of problem cases - cases in which bias needs to be included, and cannot be accounted for in broad strokes. In order to reflect the particular workings of bias in such cases, and its particular effects on credences, the C-Typer will be forced to posit a type-specification that is either wholly reliable or wholly unreliable. As we

have seen, such a route, ultimately, gives up condition (3), that of implementing the basic reliabilist intuition that what confers justification is a belief's being produced by a reliable *type* of process.

§5 Conclusion

In sum, we see that the CTS runs into a troublesome dilemma. On the one horn, they give up criterion (1); on the other, they give up (3). On closer inspection, then, as a solution to TGP, whilst CTS might satisfy one or the other of Conee and Feldman's requirements, it cannot simultaneously satisfy both.

Whilst I have focussed here on CTS in particular, the implications of our discussion span farther and wider. As noted early on, what is distinctive - and especially interesting - about CTS is its *hybrid* ambitions. The hope was that the C-Typing thesis could serve two masters: it could capture both the internalist insistence on the importance of epistemic agency, whilst ultimately offering an externalist, reliabilist account of what makes a belief justified. I have here demonstrated that CTS falls short of carving out a tenable hybrid position, but my suspicion is that the way in which it does so reveals not merely a frailty in this particular solution, but rather an underlying structural problem for similar such solutions.

The reason for this, put simply, is that adequately capturing epistemic agency means capturing what epistemic agency is *really* like. And, as we have seen, epistemic agency is impinged upon by a complex and heterogeneous collection of, among other things, feelings and biases. Such phenomena are by nature idiosyncratic. Adequately reflecting epistemic agency thus will just inevitably involve very finely specified process-types, the ilk of which simply do not sit comfortably within reliabilist, externalist theories of justification. Right at the very core of reliabilist theories is the highly intuitive thought that what confers justification to a token belief is its being the product of a reliably *type* of process, but the more narrowly you fill in the details of the 'type' here, the more this strong intuitive appeal starts to wane. Perhaps, then, in setting this hybrid position as one's goal, one creates opposing pressures that turn out, ultimately, to be irreconcilable.

References

Abreu Zavaleta, M. (2021). Communication and Indifference. *Mind and Language* 36. 81-107.

Adler, J. E. (1994). Testimony, Trust, Knowing. *Journal of Philosophy*, 91(5): 264–275.

Ahlstrom-Vij, K. (2015). The Social Virtue Of Blind Deference. *Philosophy and Phenomenological Research* 91 (3): 545–82.

Ahlstrom-Vij, K. (2016). Is There a Problem With Cognitive Outsourcing? *Philosophical Issues* 26 (1): 7–24.

Alfano, M. (2012). Expanding the Situationist Challenge to Responsibilist Virtue Epistemology. *Philosophical Quarterly*, 62(247): 223–249.

Alfano, M. (2013). *Character as Moral Fiction*, Cambridge: Cambridge University Press.

Alfano, M. (2014). Expanding the Situationist Challenge to Reliabilism About Inference. In *Fairweather*: 103–122.

Alston, W. (1985). Concepts of epistemic justification. *The Monist*, 68(1), 57–89.

Anderson, E. (1995). Knowledge, Human Interests, and Objectivity in Feminist Epistemology. *Philosophical Topics*, 23(2), 27-58.

Archer, A. (2021). The Aim of Inquiry. *Disputatio* 13 (61):95-119.

Audi, R. (2006). Testimony, Credulity and Veracity, In Lackey and Sosa (Eds.).

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon.

Baier, A. (1986). Trust and Antitrust. *Ethics*, 96(2), 231–260. https://doi.org/10.1086/292745

Baker, J. (1987). Trust and rationality. Pacific Philosophical Quarterly, 68(1), 1–13. https://doi.org/10.1111/j.1468-0114.1987.tb00280.x

Becker, L. C. (1996). Trust as Noncognitive Security about Motives. *Ethics*, 107(1): 43–61.

Beddor, B. (2021). Reasons for Reliabilism. In *Brown and Simion* 2021, pp. 146–176.

Beddor, B. (2023). Inquiry beyond knowledge. *Philosophy and Phenomenological Research* 109 (1):330-356.

Benton, M. A. (2016). Gricean Quality. *Noûs* 50 (4): 689–703.

Bergmann, M. (2006). *Justification Without Awareness*. Oxford: Oxford University Press.

Berker, S. (2013). 'Epistemic Teleology and the Separateness of Propositions'. *The Philosophical Review* 122 (3): 337–93.

Berstler, S. (2025). The Grice Is Right: Grice's Non-Cooperation Problem and the Structure of Conversation. *Philosophical Perspectives* 38 (1):26-40.

Bieber, F., and Viehoff, J. (2022). A paradigm-based explanation of trust. *Synthese*, 201(1), 2. https://doi.org/10.1007/s11229-022-03993-4.

Bok, S. (1978). *Lying: Moral Choice in Public and Private Life*, New York: Pantheon Books.

Boult, C. (2021). There is a Distinctively Epistemic Kind of Blame. *Philosophy and Phenomenological Research*, 103(3):518–534.

Boult, C. (2024). *Epistemic Blame: The Nature and Norms of Epistemic Relationships*. Oxford: Oxford University Press.

Boult, C. (forthcoming). The Significance of Epistemic Blame. *Erkenntnis*.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M. E. (2014). *Shared Agency: a Planning Theory of Acting Together*. New York: Oxford University Press.

Bratman, M. E. (2022). A Planning Theory of Acting Together. *Journal of the American Philosophical Association*, 8(3), 391–398. https://doi.org/10.1017/apa.2021.17
Brown, J. (2010). Knowledge and Assertion. *Philosophy and Phenomenological Research*.
Brownstein, M. and Saul, J. (2016). *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford University Press.
Brownstein, M. (2019). Implicit Bias. *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/implicit-bias/>.
Buchanan, R. (2010). A puzzle about meaning and communication. *Noûs* 44 (2):340-371.
Burge, T. (1993). Content Preservation. *The Philosophical Review*, *102*: 457–488.
Burge, T. (1997). Interlocution, Perception, and Memory. *Philosophical Studies*, *86*: 21–47.
Camp, E. (2006). Metaphor and that certain 'je ne sais quoi'. *Philosophical Studies* 129 (1):1-25.
Camp, E. (2017). Perspectives in Imaginative Engagement with Fiction. *Philosophical Perspectives*, 31: 73-102.
Camp, E. (2017b). Why metaphors make good insults: Perspectives, presupposition, and pragmatics. *Philosophical Studies*, 174(1), 47–64.
Camp, E. (2018). Insinuation, Common Ground, and the Conversational Record. In Daniel Fogal, Daniel W. Harris & Matt Moss, *New Work on Speech Acts*. Oxford, GB: Oxford University Press. pp. 40–66.
Camp, E. (2019). Perspectives and Frames in Pursuit of Ultimate Understanding. In *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*, edited by Stephen R. Grimm, 0. Oxford University Press.
Camp, E. (2020). Imaginative Frames for Scientific Inquiry: Metaphors, Telling Facts, and Just-So Stories. In *The Scientific Imagination*, by Elisabeth Camp, 304–36. Oxford University Press.
Capellen, H. and Dever, J. (2021). On the Uselessness of the Distinction between Ideal and Non-Ideal Theory (at least in the Philosophy of Language). In *Routledge Companion to Social and Political Philosophy of Language*. Edited by Justin Khoo and Rachel Stern. 91-105. New York: Routledge.
Carr, J. (2021). Why Ideal Epistemology? *Mind* 131 (524):1131-1162.
Carruthers, P. (2024). 'The Contents and Causes of Curiosity'. *British Journal for the Philosophy of Science* 75 (4): 871–95.
Carston, R. (2002). *Thoughts and Utterances: The pragmatics of explicit communication*. Oxford: Blackwell.
Carter, J. A. and Simion, M. (2020). The Ethics and Epistemology of Trust. In the *Internet Encyclopaedia of Philosophy*, https://iep.utm.edu/trust/, <accessed 6th November, 2022>.
Carter, J. A. (2016). Robust Virtue Epistemology As Anti-Luck Epistemology: A New Solution. *Pacific Philosophical Quarterly* 97 (1):140-155.
Carter, J. A. (2022). Trust as performance. *Philosophical Issues* 32 (1):120-147.
Carter, J. A. (2023). Trust and trustworthiness. *Philosophy and Phenomenological Research* 107 (2):377-394.
Carter, J. A. (2024). Knowledge Norms and Conversation. In Waldomiro J. Silva-Filho, *Epistemology of Conversation: First essays.* Cham: Springer Nature Switzerland. pp. 127-147.
Carter, J. A. (2024a). *A Telic Theory of Trust.* Oxford: Oxford University Press.
Carter, J. A. (2024b). Therapeutic trust. *Philosophical Psychology* 37 (1):38-61.
Carter, J. A. and Willard-Kyle, C. (2025). Virtue Epistemology for the Zetetic Turn. *Mind* 134 (536):943-966.

Carter, J. A., Jarvis. B. W., and Rubin. K. (2016). Belief without Credence. *Synthese* 193 (8): 2323–51.
Carter, J. A. and Nickel, P. J. (2014). On testimony and transmission. *Episteme* 11 (2):145-155.
Coady, C. A. J. (1973). Testimony and Observation. *American Philosophical Quarterly* 10 (2):149-155.
Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
Coffman, E. J. (2014). Lenient Accounts of Warranted Assertability. In *Epistemic Norms: New Essays on Action, Belief, and Assertion*, edited by Clayton Littlejohn and John Turri. Oxford University Press.
Coliva, A. (2025). Hinge trust. *Philosophy and Phenomenological Research* 110 (3):939-958.
Comesaña, J. (2006). A Well-Founded Solution to the Generality Problem. *Philosophical Studies*, 129 (1): 27–47.
Conee and Feldman. (1998). The Generality Problem for Reliabilism. *Philosophical Studies* 89 (1): 1–29.
Cooley, E., Payne, K., & Phillips, J. (2014). Implicit bias and the illusion of conscious ill will. *Social Psychological and Personality Science* 5 (4):500–7.
Craig, E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. New York: Oxford University Press.
Cunningham, W., Zelazo, P., Packer, D. & Van Bavel, J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, 25(5): 736–760.
D'Cruz, J. (2019). Humble Trust. *Philosophical Studies*, 176(4): 933–953.
D'Cruz, J. (2020). Trust and Distrust. In J. Simon (2020): 41–51. *The Routledge Handbook of Trust and Philosophy*. Routledge.
Darwall, S. (2014). Agreement Matters: Critical Notice of Derek Parfit, On What Matters. *Philosophical Review*, 123(1):79–105.
Dasgupta, P. (1988). Trust as a Commodity, in ed. D. Gambetta, *Trust: Making and Breaking Cooperative Relations*, New York: Basil Blackwell, pp. 49–72.
David, M. (2001). Truth as the epistemic goal. In M. Steup (Ed.), Knowledge, truth and duty: Essays on epistemic justification, responsibility and virtue (pp. 151–169). Oxford: Oxford University Press.
Davis, E. (2016). Typecasts, Tokens, and Spokespersons: A Case for Credibility Excess as Testimonial Injustice. *Hypatia* 31 (3):485-501.
Devine, P., Forscher, P., Austin, A., & Cox, W. (2012). Long-term reduction in implicit race bias; a prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6): 1267–1278.
Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
Dewey, J. (1933). *How We Think*. Chicago: Henry Regnery.
Dickie, I. (2020). Understanding Singular Terms. *Aristotelian Society Supplementary Volume* 94 (1):19-55.
Dimock, S. (2020). Trust and Cooperation. In J. Simon (2020): 160-174. *The Routledge Handbook of Trust and Philosophy*. Routledge.
Doris, J. (1998). Persons, Situations and Virtue Ethics. *Noûs*, 32 (4): 504–30.
Doris, J. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.
Douven, I. (2008). The lottery paradox and our epistemic goal. *Pacific Philosophical Quarterly*, 89:204–25.
Dover, D. (2024). Two Kinds of Curiosity. *Philosophy and Phenomenological Research* 108 (3): 811–32.

Egan, K. (1993). Narrative and learning: A voyage of implications. *Linguistics and Education*, *5*(2), 119 126. https://doi.org/10.1016/0898-5898(93)90028-9
Egan, K. (1997). *The educated mind: How cognitive tools shape our understanding*. University of Chicago Press.
El Shazly, A. (forthcoming). Communicating Understanding. *Australasian Journal of Philosophy*.
Elgin, C. Z. (1999a). Epistemology's Ends, Pedagogy's Prospects. *Facta Philosophica* 1 (1):39-54.
Elgin, C. Z. (1999b). Education and the Advancement of Understanding, In *The Proceedings of the Twentieth World Congress of Philosophy* vol. 3, pp. 131-140).
Elgin, C. Z. (2017). *True Enough*. Cambridge: MIT Press.
Ellefsen, G. (2021). Conversational cooperation revisited. *The Southern Journal of Philosophy, 59*(4), 545–571.
Falbo, A. (2021). Inquiry and confirmation. Analysis, 81(4), 622–631. https://doi.org/10.1093/analys/anab037
Falbo, A. (2023). Inquiring Minds Want to Improve. *Australasian Journal of Philosophy* 101 (2):298-312.
Falbo, A. (2023). Should epistemology take the zetetic turn? *Philosophical Studies*, 180(10-11), 2977–3002.
Falbo, A. (2025). The Zetetic. In Kurt Sylvan, Ernest Sosa, Jonathan Dancy & Matthias Steup, *The Blackwell Companion to Epistemology*, 3rd edition. Wiley-Blackwell.
Faulkner, P. (2007a). On Telling and Trusting. *Mind* 116 (464): 875–902.
Faulkner, P. (2007b). A Genealogy of Trust. *Episteme,* 4(3), 305–321. https://doi.org/10.3366/E174236000700010X
Faulkner, P. (2010). Norms of Trust. In Adrian Haddock, Alan Millar & Duncan Pritchard (eds.) *Social Epistemology*. Oxford University Press.
Faulkner, P. (2011). *Knowledge on Trust*. New York: Oxford University Press.
Faulkner, P. (2025). *Trust and Cooperation*. Oxford: Oxford University Press. https://doi.org/10.1093/9780191999284.003.0001
Feldman, R. (1985). Reliability and Justification. *The Monist*, 68(2): 159–174.
Firth, R. (1998) [1978]. The Schneck Lectures, Lecture 1: Epistemic Utility. In *Firth* 1998, 25–49. First delivered as a lecture at the University of Arizona, 1978.
Flamini, L. (2025). Inquiry and conversation: Gricean zetetic norms and virtues. *Asian Journal of Philosophy* 4 (1): 1-30.
Fleisher, W. (2024). Epistemic Practices: A Unified Account of Epistemic and Zetetic Normativity. *Noûs*.
Flores, C. (2021). Epistemic Styles. *Philosophical Topics* 49 (2):35-55.
Flores, C. (2023). Epistemic style in OCD. *Philosophy, Psychiatry, and Psychology* 30 (2):147-150.
Flores, C. (forthcoming). Delusions and epistemic style: A neurodiversity approach to reasoning in schizophrenia. *Synthese*.
Flores, C. and Woodard, E. (2023). Epistemic norms on evidence-gathering. *Philosophical Studies* 180 (9):2547-2571.
Foley, R. (1987). *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.
Fraser, R. (2021). Narrative Testimony. *Philosophical Studies* 178, no. 12: 4025–52.
Freire, P. (2000). *Pedagogy of the oppressed* (30th anniversary ed.). Continuum.
Fricker, M. (2007) *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Fricker, E. (1994). Against Gullibility. In B. Matilal and A. Chakrabarti (eds), *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, pp. 125–62. Dordrecht: Kluwer.
Fricker, E. (1995). Critical Notice: Telling and Trusting: Reductionism and Anti-reductionism in the Epistemology of Testimony. *Mind*, *104*: 393–411. doi:10.1093/mind/104.414.393
Fricker, E. (2006a). Varieties of Anti-Reductionism About Testimony–A Reply to Goldberg and Henderson. *Philosophy and Phenomenological Research* 72 (3): 618–28.
Fricker, E. (2006b). Second-Hand Knowledge. *Philosophy and Phenomenological Research* 73: 592–618.
Fricker, E. (2012). Stating and Insinuating. *Aristotelian Society Supplementary Volume*, 86(1), 61–94.
Fricker, E. (2017). Norms, Constitutive and Social, and Assertion. *American Philosophical Quarterly* 54 (4): 397–418. https://doi.org/10.2307/45128590.
Friedman, D. (2025a). Zetetic Rights and Wrong(ing)s. *Philosophical Quarterly* 75 (4):1321–1343.
Friedman, D. C. (2025b). Intentions and Inquiry. *Mind* 134 (533):85-106.
Friedman, D. C. (forthcoming). Cooperation and Shared Inquiry. *Australasian Journal of Philosophy*.
Friedman, J. (2019a). Teleological Epistemology. *Philosophical Studies* 176 (3): 673–91.
Friedman, J. (2019b). Inquiry and Belief. *Noûs* 53 (2):296-315.
Friedman, J. (2019c). Checking again. Philosophical Issues 29 (1):84-96.
Friedman, J. (2020). 'The Epistemic and the Zetetic'. *The Philosophical Review* 129, no. 4: 501–36.
Frost-Arnold, K. (2014). The Cognitive Attitude of Rational Trust. *Synthese*, 191(9). https://doi.org/10.1007/s11229-012-0151-6
García-Carpintero, M. (2020). On the nature of presupposition: A normative speech act account. *Erkenntnis*.
Gardiner, G. (2022). Attunement: On the Cognitive Virtues of Attention. In Mark Alfano, Jeroen De Ridder & Colin Klein, *Social Virtue Epistemology*. Routledge.
Gaszczyk, G. (2024). The Informativeness Norm of Assertion. *Review of Philosophy and Psychology* 15 (4): 1157–78.
Gaszczyk, G. and Grodniewicz, J. P. (forthcoming). The practice of explaining. *Philosophical Quarterly*.
Gawronski, B. & S. Brannon (2017). Attitudes and the Implicit-Explicit Dualism. In *The Handbook of Attitudes, Volume 1: Basic Principles (2nd Edition)*, D. Albarracín & B.T. Johnson (eds.), New York: Routledge, pp. 158–196.
Gerken, M. (2015). The epistemic norms of intra-scientific testimony. *Philosophy of the Social Sciences*, 45(6), 568–595.
Gigerenzer, G. & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 1–12). The MIT Press.
Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies, 144*, 167–187.
Goldberg, S. (2007). *Anti-Individualism: Mind and Language, Knowledge and Justification*. New York: Cambridge University Press.
Goldberg, S. (2010). *Relying On Others*. Oxford: Oxford University Press.
Goldberg, S. (2013). Epistemic Dependence in Testimonial Belief. In the Classroom and Beyond. *Journal of Philosophy of Education* 47(2): 168–186, https://doi.org/10.1111/1467-9752.12019
Goldberg, S. (2015a). Should Have Known. *Synthese* 194 (8): 2863–94.
Goldberg, S. (2015b). *Assertion: On the Philosophical Significance of a Speech Act*. Oxford: Oxford University Press.

Goldberg, S. (2017). The Asymmetry Thesis and the Doctrine of Normative Defeat. *American Philosophical Quarterly* 54 (4): 339–51.
Goldberg, S. (2018). *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity*. Oxford University Press.
Goldberg, S. (2024). On the Contours of a Conversation. In Waldomiro J. Silva-Filho, *Epistemology of Conversation: First essays.* Cham: Springer Nature Switzerland. pp. 29-47.
Goldberg, S. and Henderson, D. (2006). Monitoring and Anti-Reductionism in the Epistemology of Testimony. *Philosophy and Phenomenological Research*, *72*: 600–617.
Goldman, A. (1979). What Is Justified Belief? In *Justification and Knowledge*, edited by G. Pappas. Dordrecht: Reidel.
Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
Goldman, A. (1991). *Knowledge in a social world*. New York: Oxford University Press.
Goldman, A. (2012). *Reliabilism and Contemporary Epistemology: Essays*. Oxford University Press, New York.
Goldman, A. and Beddor, B. (2016). Reliabilist Epistemology. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/reliabilism/>.
Gordon, E. C. (2022). When Monitoring Facilitates Trust. *Ethical Theory and Moral Practice*. https://doi.org/10.1007/s10677-022-10286-9
Graham, P. J. (2006a). Liberal Fundamentalism and Its Rivals. In J. Lackey and E. Sosa (Eds.), *The Epistemology of Testimony*. Oxford University Press.
Graham, P. J. (2006b). Testimonial Justification: Inferential or Non-Inferential? *Philosophical Quarterly*, 56(222): 84–95.
Graham, P. J. (2010). Testimonial Entitlement and the Function of Comprehension. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Social Epistemology*. Oxford: Oxford University Press.
Graham, P. J. (2014). Warrant, Functions, History. In Abrol Fairweather & Owen Flanagan, *Naturalizing Epistemic Virtue*. New York: Cambridge University Press. pp. 15-35.
Graham, P. J. (2012). Epistemic Entitlement. *Noûs*, 46(3): 449–483.
Graham, P. J. (2015). Epistemic Normativity and Social Norms. In D. K. Henderson & J. Greco (Eds), *Epistemic Evaluation: Purposeful Epistemology*. Oxford University Press.
Graham, P. J. (2018a). Social Knowledge and Social Norms. In Markos Valaris & Stephen Hetherington (eds.), *Knowledge in Contemporary Philosophy*. London, UK: Bloomsbury Publishing. pp. 111-138.
Graham, P. J. (2018b). Formulating Reductionism about Testimonial Warrant and the Challenge from Childhood Testimony. *Synthese* 195 (7): 3013–33. https://doi.org/10.1007/s11229-016-1140-y.
Graham, P. J. (Forthcoming). Deception Detection Research: Some Lessons for Epistemology. In Waldomiro Silva Filho (ed.), *Epistemology of Conversation: First essays*. Cham: Springer.
Greco, D. (2023). *Idealization in epistemology: A modest modeling approach*. Oxford University Press.
Greco, J. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge: Cambridge University Press.
Greco, J. (2020). *The Transmission of Knowledge*. Cambridge: Cambridge University Press.
Green, A. (2017). An Epistemic Norm for Implicature. *Journal of Philosophy*, 114(7), 381–91.
Green, A. (2019). The Transmission of Understanding. *Res Philosophica* 96(1), pp. 43-61.

Grice, H. P. (1975). Logic and conversation. In Cole, P., editor, *Syntax and Semantics 3: Pragmatics*. Academic Press, New York.
Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.
Grimm, S. (2020). Transmitting Understanding and Know-How. In *What the Ancients Offer to Contemporary Epistemology*, edited by Stephen Hetherington and Nicholas D. Smith. New York, USA: Routledge.
Grimm, S. (2008). Epistemic Goals and Epistemic Values. *Philosophy and Phenomenological Research*, 77: 725-44.
Guilielmo, B. (2024). Inquiry, Questions, and Actions. *Dialogue* 63 (2):339-356.
Hardin, R. (2002). *Trust and Trustworthiness*. Russell Sage Foundations Series on Trust. New York, NY: Russell Sage Foundation.
Harding, M. (2011). Responding to Trust. *Ratio Juris* 24 (1):75-87.
Harman, G. (1973). *Thought*. Princeton: Princeton University Press.
Harman, G. (1999). Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society* (New Series). 119: 316–31.
Harman, G. (2009). Scepticism About Character Traits. *Journal of Ethics*, 13: 235–42.
Harnish, R. M. (1991) [1976]. Logical Form and Implicature. In *Pragmatics: a Reader*. Reprint. Edited by Stephen Davies. 316-364. New York: Oxford University Press, 1991 [1976].
Hawley, K. (2010). Testimony and Knowing How. *Studies in History and Philosophy of Science Part A*, Explanation, inference, testimony, and truth: Essays dedicated to the memory of Peter Lipton, 41 (4): 397–404.
Hawley, K. (2012). *Trust: A Very Short Introduction*. Oxford: Oxford University Press.
Hawley, K. (2014). Trust, Distrust and Commitment. *Noûs*, 48(1), 1–20. https://doi.org/10.1111/nous.12000
Hawley, K. (2019). *How To Be Trustworthy*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198843900.003.0001
Hawthorne, J, (2012). Some Comments on Fricker's 'Stating and Insinuating'. *Aristotelian Society Supplementary Volume*, 86(1), 95–108.
Haziza, E. (2022). Assertion, Implicature, and Iterated Knowledge. *Ergo an Open Access Journal of Philosophy* 8 (0).
Haziza, E. (2023a). Inquiring and Making Sure. *Philosophical Topics* 51 (2):185-195.
Haziza, E. (2023b). Norms of Inquiry. *Philosophy Compass* 18 (12):e12952.
Hazlett, A. (2025). Understanding and testimony. In J. Lackey & A. McGlynn (Eds.), *Oxford Handbook of social epistemology*. (pp. 600–619). Oxford University Press.
Hazlett, A. (forthcoming) Testimony, understanding, and art criticism. In *Philosophy and Art: New Essays at the Intersection*, edited by C. Mag Uidhir, Oxford: Oxford University Press.
Hieronymi, P. (2005). The Wrong Kind of Reason. *Journal of Philosophy*, 102(9), 437–457. https://doi.org/10.5840/jphil2005102933
Hieronymi, P. (2006). Controlling Attitudes. *Pacific Philosophical Quarterly*, 87(1), 45–74. https://doi.org/10.1111/j.1468-0114.2006.00247.x
Hieronymi, P. (2008). The reasons of trust. *Australasian Journal of Philosophy*, 86(2), 213–236. https://doi.org/10.1080/00048400801886496
Hieronymi, P. (2009). Two kinds of agency. In L. O'Brien & M. Soteriou (Eds.), *Mental actions*. Oxford: OUP.
Hills, A. (2016). Understanding Why. *Noûs* 50 (4): 661–88.

Hills, A. (2020). Moral Testimony: Transmission Versus Propagation. *Philosophy and Phenomenological Research* 101 (2): 399–414.
Hinchman, E. (2013). Assertion, Sincerity, and Knowledge. *Noûs* 47 (4): 613–46.
Hinchman, E. (2020). Trust and Will. In J. Simon (2020). *The Routledge Handbook of Trust and Philosophy*. Routledge.
Hollis, M. (1998). *Trust Within Reason*. Cambridge: Cambridge University Press.
Holroyd, J. (2012). Responsibility for Implicit Bias. *Journal of Social Philosophy* 43 (3): 274–306.
Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, 72(1), 63–76. https://doi.org/10.1080/00048409412345881
Hookway, C. (2006). *Epistemology and Inquiry: The Primacy of Practice (in Epistemology Futures)*. Oxford, United Kingdom: Oxford University Press, Incorporated.
Hookway, C. (2010). Some Varieties of Epistemic Injustice: Reflections on Fricker. *Episteme*, *7*(2), 151–163.
Horsburgh, H. J. N. (1960). The Ethics of Trust. *The Philosophical Quarterly*, 10(41): 343–354.
Hume, D. (1978) [1740]. *A Treatise of Human Nature*, London: John Noon. Reprinted in L. A. Selby-Bigge (ed.), second edition, revised P. H. Nidditch (ed.), Oxford: Oxford University Press.
Inglehart, R. (1999). Trust, Well-Being and Democracy. In Warren 1999: 88–120.
James, W. (1896) [1979]. *The Will to Believe and Other Essays in Popular Philosophy*, Cambridge, MA and London: Harvard University Press.
Jones, K. (1996). Trust as an Affective Attitude. *Ethics*, 107(1), 4–25. https://doi.org/10.1086/233694
Jones, W. E. (1997). "Why Do We Value Knowledge?" *American Philosophical Quarterly*, 34(4): 423–440.
Jönsson, M. (2013). A Reliabilism Built On Cognitive Convergence: An Empirically Grounded Solution To The Generality Problem. *Episteme* 10 (3): 241–68.
Jope, M. (2021). The symmetry problem for testimonial conservatism. *Synthese* 199 (3-4):6149-6167.
Kauppinen, A. (2018). Epistemic Norms and Epistemic Accountability. *Philosophers' Imprint* 18.
Kavka, G. S. (1983). The Toxin Puzzle. Analysis, 43, 33–36.
Kearl, T. (2023). Epistemic control without voluntarism. *Philosophical Issues* 33 (1):95-109.
Keiser, J. (2022). *Non-Ideal Foundations of Language*. New York: Routledge.
Keller, P. (2024). Testimony by Presupposition. *Erkenntnis*, 89 (6):2149-2167.
Kelly, T. (2003). Epistemic Rationality as Instrumental Rationality: A Critique. *Philosophy and Phenomenological Research* 66 (3): 612–40.
Kelp, C. (2014). No Justification for Lottery Losers. *Pacific Philosophical Quarterly* 95 (2):205-217.
Kelp, C. (2018a). Assertion: A Function First Account. *Noûs*, *52*(2), 411–442.
Kelp, C. (2018b). Inquiry, Knowledge and Understanding. *Synthese*.
Kelp, C. (2021a) *Inquiry, Knowledge, and Understanding*. Oxford: Oxford University Press.
Kelp, C. (2021b). Theory of inquiry. *Philosophy and Phenomenological Research* 103 (2):359-384.
Kelp, C. (2022). Defeat and proficiencies. *Philosophical Issues*. 32,82–103
Kelp, C. (2023). *The Nature and Normativity of Defeat*. Cambridge: Cambridge University Press.
Kelp, C. and Simion, M. (2023). What is trustworthiness? *Noûs* 57 (3):667-683.
Kelp, C. and Simion, M. (2021). *Sharing knowledge: A functionalist account of assertion*. Cambridge University Press.

Kelp, C., Boult, C., Broncano-Berrocal, F., Dimmock, P., Ghijsen, H., and Simion. M. (2020). Hoops and Barns: A New Dilemma for Sosa. *Synthese, 197*: 5187–5202.
Kenyon, T. (2013). The Informational Richness of Testimonial Contexts. *The Philosophical Quarterly*, 63(250): 58–80. doi:10.1111/1467-9213.12000
Keren, A. (2014). Trust and belief: A preemptive reasons account. *Synthese,* 191(12), 2593–2615. https://doi.org/10.1007/s11229-014-0416-3
Keren, A. (2020). Trust and Belief. In *The Routledge Handbook of Trust and Philosophy*. Routledge.
Khoo, J. & Sterken, R. K. (Eds.). (2021). *The Routledge Handbook of Social and Political Philosophy of Language* (1st ed.). Routledge.
Kirton, A. (2020). Matters of Trust as Matters of Attachment Security. *International Journal of Philosophical Studies* 28 (5):583-602.
Kolodny, N. (2005). Why Be Rational? *Mind*, 114(455), 509–563. https://doi.org/10.1093/mind/fzi509
Kornblith, H. (2002). *Knowledge and its Place in Nature*. Oxford University Press.
Kornblith, H. (2007). Is Philosophical Knowledge Possible? In Christoph Jäger & Winfried Löffler (eds.), Epistemology: Contexts, Values, Disagreement. Papers of the 34th International Ludwig Wittgenstein-Symposium in Kirchberg, 2011. pp. 285-304.
Kvanvig, J. L. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.
Lackey, J. (1999). Testimonial Knowledge and Transmission. *Philosophical Quarterly* 49 (197): 471–90.
Lackey, J. (2007) Norms of Assertion. *Noûs* 41: 594-626.
Lackey, J. (2008). *Learning From Words*. Oxford: Oxford University Press.
Lackey, J. (2011). Assertion and Isolated Second-Hand Knowledge. In *Assertion: New Philosophical Essays*. Oxford University Press.
Lackey, J. (2018). Credibility and the Distribution of Epistemic Goods. In *Believing in Accordance with the Evidence*, edited by McCain Kevin. Springer Verlag.
Lasonen-Aarnio, M. (2024). Perspectives and good dispositions. *Philosophy and Phenomenological Research* 109 (3): 774-798.
Lasonen-Aarnio, M. (2025). Knowledge-Conducive Dispositions: How to Do Consequentialist Epistemology. *Synthese*, *205*(3).
Leonard, N. (2023). Epistemological Problems of Testimony. *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2023/entries/testimony-episprob/>.
Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford: Oxford University Press.
Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
Litzky, B. E., Eddleston, K. A, & Kidder, D. L. (2006). The Good, the Bad, and the Misguided: How Managers Inadvertently Encourage Deviant Behaviors. *Academy of management perspectives* 20(1): 91-103.
Lyons, J. (1997). Testimony, Induction and Folk Psychology. *Australasian Journal of Philosophy*. 75(2): 163–178.
Lyons, J. (2009). *Perception and Basic Beliefs: Zombies, Modules and the Problem of the External World*. Oxford: Oxford University Press.
MacFarlane, J. (2020a). Indeterminacy as Indecision, Lecture 2: Seeing Through the Clouds. *Journal of Philosophy* 117 (11/12). 617-642.
MacFarlane, J. (2020b). Indeterminacy as Indecision, Lecture 3: Indeterminacy as Indecision. *Journal of Philosophy* 117 (11/12). 643-667.
Maguire, B. (2018). There Are No Reasons for Affective Attitudes. *Mind*, 127(507), 779–805.
Maguire, B. & Woods, J. (2020). The game of belief. *The Philosophical Review*, 129, 211–249.

Malfatti, F. (2019). Can Testimony Generate Understanding? *Social Epistemology* 33 (6):477-490.
Malfatti, F. (2021). On Understanding and Testimony. *Erkenntnis* 86 (6): 1345–65.
Malfatti, F. (2025a). Understanding and Transmission. In Kurt Sylvan, Ernest Sosa, Jonathan Dancy & Matthias Steup, *The Blackwell Companion to Epistemology,* 3rd edition.
Malfatti, F. (2025b). ChatGPT, Education, and Understanding. *Social Epistemology*.
Mallory, F. (2026). Inquisitive injustice. *Philosophical Studies*, *183*(1), 143–163.
Malmgren, A. (2006). Is There A Priori Knowledge by Testimony? *The Philosophical Review*, 115(2): 199–241. doi:10.1215/00318108-2005-015
McGeer, V. (2008). Trust, hope and empowerment. *Australasian Journal of Philosophy*, 86(2), 237–254. https://doi.org/10.1080/00048400801886413
Mchugh, C. (2012). The Truth Norm of Belief. *Pacific Philosophical Quarterly*, 93(1), 8–30. https://doi.org/10.1111/j.1468-0114.2011.01413.x
McHugh, C. (2017). Attitudinal control. *Synthese*, 194(8), 2745–2762. https://doi.org/10.1007/s11229-014-0643-7
McHugh, C., and Way, J. (2018). What Is Reasoning? *Mind*, 127(505), 167–196.
McKinnon, R. (2015). *The Norms of Assertion: Truth, Lies, and Warrant*, London: Palgrave Macmillan UK.
McLeod, C. (2002). *Self-Trust and Reproductive Autonomy*, Cambridge, MA: MIT Press.
McMyler, B. (2011). *Testimony, Trust, and Authority*, Oxford: Oxford University Press.
Merritt, M, W., Doris, J. M. and Harman, G. (2010). Character. In John M. Doris and The Moral Psychology Research Group (eds.), *The Moral Psychology Handbook*, Oxford: Oxford University Press, ch. 11.
Michaelian, K. (2010). In defence of gullibility: The epistemology of testimony and the psychology of deception detection. *Synthese* 176 (3):399-427.
Titus, L. (2017). Epistemic Agency and the Generality Problem. *Philosophical Topics* 45, (1): 107–20.
Montgomery, B. (2014). In Defense of Assertion. *Philosophical Studies* 171 (2): 313–26.
Moran, R. (1994). The Expression of Feeling in Imagination. *Philosophical Review* 103:1, 75106.
Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton: Princeton UP.
Munton, J. (2021). Prejudice as the Misattribution of Salience. *Analytic Philosophy* 1–19.
Munton, J. (2022). Answering Machines: How to (Epistemically) Evaluate a Search Engine. *Inquiry* 1–29.
Nguyen, C. T. (2022). Trust as an unquestioning attitude. *Oxford Studies in Epistemology* 7:214-244.
Noddings, N. (1995). *Philosophy of education*. Boulder, Colo.: Westview Press.
Nottelmann, N. (2021). Against normative defeat. *Mind* 130 (520):1183-1204.
O'Neil, C. (2012). Lying, Trust, and Gratitude: Lying, Trust, and Gratitude. *Philosophy & Public Affairs*, 40(4): 301–333
O'Neill, O. (2002). *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press.
Olsson, E. (2016). A Naturalistic Approach to the Generality Problem. In *McLaughlin and Kornblith*: 178–199
Pagin, P. and Marsili, M. (2025). Assertion. *The Stanford Encyclopedia of Philosophy* (Fall 2025 Edition), Edward N. Zalta & Uri Nodelman (eds.), forthcoming URL = <https://plato.stanford.edu/archives/fall2025/entries/assertion/>.
Pasnau, R. (2013). Epistemology idealized. *Mind* 122:987–1021
Patrizio, R. F. (2024a). Testimonial liberalism and the balance of epistemic goals. *Philosophical Studies* 181 (11):2997-3015.

Patrizio, R. F. (2024b). Trust's Meno problem: Can the doxastic view account for the value of trust? *Philosophical Psychology* 37 (1):18-37.
Paulson, S. (manuscript). Open by default.
Payne, B. & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In *Handbook of implicit social cognition: Measurement, theory, and applications*, B. Gawronski, & B. Payne (eds.), New York, NY: Guilford Press, pp. 1–17.
Peet, A. (2015a). Testimony, Pragmatics, and Plausible Deniability. *Episteme* 12 (1): 29–51.
Peet, A. (2015b). Testimony, context, and miscommunication. PhD thesis. University of St Andrews.
Peet, A. (2016). Testimony and the epistemic uncertainty of interpretation. *Philosophical Studies, 173*(2), 395–416.
Peet, A. (2021). Defective contexts. In R. K. Sterken, & J. Khoo (Eds.), *Routledge handbook of social and political philosophy of Language*. Routledge.
Peet, A. (2023). Collective Communicative Intentions in Context. *Ergo: An Open Access Journal of Philosophy* 10:211-236.
Pelling, C. (2013). Assertion and The Provision of Knowledge. *The Philosophical Quarterly*, *63*(251), 293–312.
Pelling, C. (2014). Assertion, Telling, and Epistemic Norms. *Australasian Journal of Philosophy* 92 (2): 335–48.
Perrine, T. (2014). In defense of non-reductionism in the epistemology of testimony. *Synthese* 191 (14): 3227-3237.
Peters, U. (2019). Implicit Bias, Ideological Bias, and Epistemic Risks in Philosophy. *Mind & Language* 34 (3): 393–419.
Pettigrew, R. (2022). *Epistemic Risk and the Demands of Rationality*. Oxford: Oxford University Press.
Pettit, P. (1995). The Cunning of Trust. *Philosophy & Public Affairs*, 24(3), 202–225. https://doi.org/10.1111/j.1088-4963.1995.tb00029.x
Plantinga, A. (1993). Other Persons and Testimony. In *Warrant and proper function*. New York: Oxford University Press.
Pollock, J. (1986). *Contemporary Theories of Knowledge*. Lanham: Rowman and Littlefield.
Pollock, J. (2021). Linguistic Understanding and Testimonial Warrant. *Erkenntnis* 88 (2):457-477.
Pollock, J. (2024). Content internalism and testimonial knowledge. *Inquiry: An Interdisciplinary Journal of Philosophy* 67 (6):1947-1968.
Pollock, J. (2023). Testimonial knowledge and content preservation. *Philosophical Studies* 180 (10):3073-3097.
Pritchard, D. (2009). The Value of Knowledge. *The Harvard Review of Philosophy* 16 (1): 86–103.
Pritchard, D. (2010). The Value Problem for Knowledge. In *The Nature and Value of Knowledge: Three Investigations*, edited by Duncan Pritchard, Alan Millar, and Adrian Haddock, Oxford University Press.
Pritchard, D. (2011). What Is the Swamping Problem? In *Reasons for Belief*, edited by Andrew Reisner and Asbjørn Steglich-Petersen, 1st ed., 244–59. Cambridge University Press.
Rabinowicz, W. & Rønnow-Rasmussen, T. (2000). A Distinction in Value: Intrinsic and for Its Own Sake. *Proceedings of the Aristotelian Society* 100 (1): 33–51.
Reid, T. (1764). *An Inquiry into the Human Mind on the Principles of Common Sense*. Excerpts in 1975 edition, Indianapolis: Hackett Publishing Company.
Riggs, W. D. (2003). Balancing Our Epistemic Goals. *Nous* 37 (2): 342–52.

Riggs, W. D. (2002). Reliability and the Value of Knowledge. *Philosophy and Phenomenological Research*, 64(1): 79–96.
Rimoldi, F. (2024). Norms of Inquiry Conversations: An Essay. In Waldomiro J. Silva-Filho, *Epistemology of Conversation: First essays*. Cham: Springer Nature Switzerland. pp. 149-171.
Roberts, C. (1996). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics* 5:1-69.
Roberts, C. (2018). Speech acts in discourse context. In D. Fogal, D. Harris, & M. Moss (Eds.), New work on speech acts (pp. 317-359). Oxford: Oxford University Press.
Saint-Croix, C. (2022). Rumination and Wronging: The Role of Attention in Epistemic Morality. *Episteme*, *19*(4), 491–514.
Saul, J. (2013). Scepticism and Implicit Bias. *Disputatio* 5 (37): 243–63.
Saul, J. (2018). Dogwhistles, political manipulation, and philosophy of language. In D. Fogal, D. Harris, & M. Moss (Eds.), *New work on speech acts* (pp. 360–384). Oxford University Press.
Schmitt, F. F. (1987). Justification, sociality, and autonomy. *Synthese* 73 (1):43 - 85.
Schultz, C. (2024). The Deliberative Constraint on Reasons. *Philosophy Compass* 19 (7):e13010.
Schultz, C. (2025). Deliberative Control and Eliminativism about Reasons for Emotions. *Australasian Journal of Philosophy* 103 (1):72-87.
Shah, N., & Velleman, J. D. (2005). Doxastic Deliberation. *The Philosophical Review*, 114(4), 497–534.
Shieber, J. (2012). Against Credibility. *Australasian Journal of Philosophy* 90 (1):1 - 18.
Siegel, S. (2022). Salience Principles for Democracy. In Sophie Archer (3 ed.), *Salience: A Philosophical Inquiry* (235-266). Routledge.
Simion, M. (2018). No Epistemic Norm for Action. *American Philosophical Quarterly* 55 (3): 231–38.
Simion, M. (2019). Knowledge-first functionalism. *Philosophical Issues* 29 (1): 254-267.
Simion, M. (2021). Testimonial contractarianism: A knowledge-first social epistemology. *Noûs* 55 (4):891-916.
Simion, M. (2023). Resistance to Evidence and the Duty to Believe. *Philosophy and Phenomenological Research* 108 (1): 203-216.
Simion, M. (2024). *Resistance to Evidence*. Cambridge University Press.
Simion, M. and Kelp, C. (2020). How to Be an Anti-Reductionist. *Synthese, 197*: 2849–66.
Simion, M. and Kelp, C. (2021). *Sharing Knowledge: A Functionalist Account of Assertion*. Cambridge University Press.
Simion, M. and Kelp, C. (2017). Criticism and Blame in Action and Assertion. *Journal of Philosophy* 114 (2): 76-93.
Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics* 69:99–118.
Simon, J. (eds.) (2020) *The Routledge handbook of trust and philosophy*. Routledge, New York, NY
Simpson, T. W. (2023). *Trust: A Philosophical Study*. Oxford: Oxford University Press.
Simpson, T. W. (2012) What Is Trust? *Pacific Philosophical Quarterly*, 93(4): 550–569.
Singer, D. J. (2019). Permissible Epistemic Trade-Offs. *Australasian Journal of Philosophy*, *97*(2), 281–293.
Singer, D. J. and Aronowitz, S. (2022). What Epistemic Reasons Are For: Against the Belief-Sandwich Distinction. In Dunaway, B. and Plunkett, D. (eds), Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard. Maize Books.

Sliwa, P. (2023). Making sense of things: Moral inquiry as hermeneutical inquiry. *Philosophy and Phenomenological Research* 109 (1):117-137.
Sliwa, P. (2025). Changing minds and hearts: Moral testimony and hermeneutical advice. In R. Shafer-Landau & R. Ross (Eds.), Oxford studies in metaethics (Vol. 20). Oxford University Press.
Smith, L., and Archer, A. (2020). Epistemic Injustice and the Attention Economy. *Ethical Theory and Moral Practice*, *23*(5), 777–795.
Smith, N. (2020). Simply Finding Answers, or the Entirety of Inquiry While Standing on One Foot. *Disputatio* 12 (57): 181–98.
Sosa, E. (1991). *Knowledge in Perspective*. Cambridge: Cambridge University Press.
Sperber, D. and Wilson, D. (1986). *Relevance*. Cambridge: Harvard University Press.
Stalnaker, R. C. (1978). Assertion. In *Syntax and Semantics 9: Pragmatics*, Peter Cole (ed.), New York: Academic Press, 315–332.
Steglich-Petersen, A. (2009). Weighing the Aim of Belief. *Philosophical Studies*, 145, 395–405.
Stokke, A. (2016). Truthfulness and Gricean Cooperation. *Grazer Philosophische Studien* 93: 489-510.
Stotts, M. H. (2024). Conversational Maxims as Social Norms. *Inquiry* 67 (9): 3087–109.
Sullivan-Bissett, E. (2015). Implicit Bias, Confabulation, and Epistemic Innocence. *Consciousness and Cognition* 33: 548–60.
Swinburne, R. (1999). *Providence and the Problem of Evil*. Oxford: Oxford University Press.
Swinburne, R. (2000). *Epistemic Justification*. Oxford: Oxford University Press.
Szabó, Z. G. (2020). The Goal of Conversation. *Aristotelian Society Supplementary Volume* 94 (1):57-86.
Thorstad, D. (2022). There Are No Epistemic Norms of Inquiry. *Synthese* 200 (5): 410.
Thorstad, D. (2021). Inquiry and the Epistemic. *Philosophical Studies* 178 (9): 2913–28.
Thorstad, D. (2024). Why bounded rationality (in epistemology)? *Philosophy and Phenomenological Research* 108 (2):396-413.
Turri, J., Alfano, M. and Greco, J. (2021). Virtue Epistemology. *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/>.
Van Cleve, J. (2006). Reid on the Credit of Human Testimony. In J. Lackey and E. Sosa (eds), *The Epistemology of Testimony,* pp. 50–74. Oxford: Oxford University Press.
Velleman, J. D. (2000). *The Possibility of Practical Reason*. Oxford UP.
Vermaire, M. (2025). Judgment's Aimless Heart. *Noûs* 59 (1):3-21.
Wanderer, J., & Townsend, L. (2013). Is it Rational to Trust? *Philosophy Compass*, 8(1), 1–14. https://doi.org/10.1111/j.1747-9991.2012.00533.x
Watzl, S. (2022). The Ethics of Attention: An Argument and a Framework. In *Salience*, by Sophie Archer, 1st ed., 89–112. London: Routledge.
Way, J. (2012). Transmission and the Wrong Kind of Reason. *Ethics*, 122(3), 489–515. https://doi.org/10.1086/664749
Weiner, M. (2003). Accepting Testimony. *The Philosophical Quarterly* 53 (211): 256–64.
Weiner, M. (2005). Must We Know What We Say? *The Philosophical Review*, 114(2): 227–251. doi:10.1215/00318108-114-2-227
Whitcomb, D. (2017). One kind of asking. *Philosophical Quarterly* 67: 148–68.
Whitcomb, D. (2023). Bullshit Questions. *Analysis* 83 (2): 299–304.
Whitcomb, D. and Millson, J. (2024). Inquiring Attitudes and Erotetic Logic: Norms of Restriction and Expansion. *Journal of the American Philosophical Association* 10 (3): 444–66.

Whiteley, E. (2023). "A Woman First and a Philosopher Second": Relative Attentional Surplus on the Wrong Property. *Ethics* 133 (4): 497–528.
Whiteley, E. (2024). Order-Based Salience Patterns in Language: What They Are and Why They Matter. *Ergo: An Open Access Journal of Philosophy* 11: 689–715.
Whiteley, E. (2022). Harmful Salience Perspectives. In *Salience*, by Sophie Archer, 1st ed., 193–212. London: Routledge.
Whiting, D. (2013). Stick to the Facts: On the Norms of Assertion. *Erkenntnis*, 78(4): 847–867. doi:10.1007/s10670-012-9383-6
Willard-Kyle, C. (2021). "P, but You Don't Know That P". *Synthese* 199 (5–6): 14667–90.
Willard-Kyle, C. (2023). The Knowledge Norm for Inquiry. *Journal of Philosophy* 120 (11):615-640.
Willard-Kyle, C. (2024). Ignorance, Soundness, and Norms of Inquiry. *Philosophical Studies*.
Willard-Kyle, C., Millson, J. and Whitcomb, D. (2025). 'Evoked Questions and Inquiring Attitudes'. *Philosophical Quarterly*.
Williams, B. (1973). Deciding to Believe. In *Problems of the Self* (pp. 136–151). Cambridge UP.
Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.
Williamson, T. (1996). Knowing and asserting. *Philosophical Review* 105 (4):489-523.
Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
Woodard, E. (2024). Why Double-Check? *Episteme* 21 (2):644-667.
Wonderly, M. (2016). On being attached. *Philosophical Studies* 173 (1):223-242.
Wu, W. (2024). On Attention and Norms: An Opinionated Review of Recent Work. *Analysis* 84 (1):173-201.
Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.
Zagzebski, L. (2003). The Search for the Source of Epistemic Good. *Metaphilosophy*, 34(1–2): 12–28; and reprinted in Brady and Pritchard 2003.
Zagzebski, L. (2008). *On Epistemology*. Wadsworth.
Zagzebski, L. (2012), *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford University Press.
Zollman, K. J. S. (2015). Modeling the Social Consequences of Testimonial Norms. *Philosophical Studies* 172 (9): 2371–83. https://doi.org/10.1007/s11098-014-0416-7.