Essays on Knowledge and Justice

Martin Miragoli
MSc

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy

Philosophy
School of Humanities
College of Arts
University of Glasgow

November 2024

Abstract

This thesis can be seen as a modest contribution to a growing literature that aims to challenge some core assumptions of a widely shared image of epistemology. This is an image of epistemology as a discipline centred on the individual, whose core assumptions concern the way in which this individual relates to the world around them and to other people. This thesis contributes to a critique of this image in a *non-direct* and *non-unitary* manner. The critique is not direct because (with the exception of chapter one and, to some extent, chapter six) none of the works here collected offers an explicit challenge to this image. The critique is instead *positive*, as it furthers a competing image of epistemology as a deeply social discipline. Finally, this critique is non-unitary because the chapters put forward independent arguments, each attempting to capture a different angle of the multiple ways in which social and political relations structure how we think about core epistemic concepts. The result is a harlequin work describing the branching trajectory of an ongoing research into some fundamental philosophical questions on the nature of our epistemic lives.

Table of Contents

Introduction 7
1. Preamble, or: the Story of S. and the Epistemic Garden of Eden 7
2. The General Picture 11
3. Summary of the chapter 17

Chapter 1

The Environmental Image: The Case of White Ignorance for Epistemic Justice 20
Introduction 21
§1. The Case 22
1.1. Sanctioned White Ignorance 24
1.2. The Epistemic Side 29
§2. The Agential Image 31
2.1. Individualistic Accounts 31
2.2. Socialist Accounts 34
§3. The Environmental Image 36
3.1. Epistemic Environments 38
3.2. Epistemic Justice and White Ignorance 42
Coda 43

Chapter 2

Conformism, Ignorance and Injustice: AI As A Tool Of Epistemic Oppression 45
§1. AI, Justice and the future of research 45
§2. Biassed data and Epistemic Conformism 48
§3. Hermeneutical Lacunae and White Ignorance 53
3.1. Hermeneutical Lacunae and Zetetic Injustice 54
3.2. White Ignorance and Epistemic Spurning 57
Coda 61

Chapter 3

What Is Mansplaining? 64
§1. The Standard View 64
§2. Mansplaining and Haughtiness 69
§3. Mansplaining and Epistemic Injustice 74
Coda 80

Chapter 4

Groups Believe In Many Ways (And They're All Fine) 82

Introduction 82
§1. Monist Accounts of Group Belied 84
1.1. Complementary Accounts: Deflationism and Strong Inflationism 85
1.2. Between Universality and Particularity 89
1.3. Idiosyncratic Monism 90
§2. Radical Functionalism 91
2.1. The Hypothesis of Multiple Realisability 91
2.2. Radical Functionalism about Group Belief 93
§3. Aggregates, Social Groups & Other Concerns 97
3.1. A Functional Analysis of Aggregate Belief 97
3.2. Social Groups 102
3.3. Problems of Inheritance 105
3.4. Over-generation 107
Coda 108

Chapter 5

Race, Gender and Group Disagreement 109
Introduction 109
§1. Gender, Race and Group Peer Disagreement 110
§2. A (Problematically) Narrow Methodological Choice 111
§3. A Functionalist Solution 114
3.1. The Peerhood Constraint: Functionalism About Group Belief 114
3.2. The Normative Constraint: Functionalism About the Epistemology of Disagreement 117

Chapter 6

A Final Word On Hinge Epistemology 119
Introduction 119
§1. The Framework Reading 121
§2. Extending Rationality 126
§3. Surrender To The Angsts 133
Concluding Remarks 142

Bibliography 143

A mamma e papá
ai nonni
a mio fratello, a Rebecca
e ai piccolini
—la mia casa

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work. Chapter Two is an adapted version of the paper "Conformism, Ignorance & Injustice: AI as a Tool of Epistemic Oppression" published in *Episteme*. Chapter Five is an adapted version of the paper "Gender, Race and Group Disagreement" published in the volume *The Epistemology of Group Disagreement*. Both Chapter Three and Five are the product of a research collaboration. In both cases, I wrote the original and final manuscript. The ideas contained in Chapter Three originated from discussions with Daniela Rusu. The main ideas and the structure of Chapter Five come from discussions with Mona Simion. Mona Simion has also contributed to the writing of the final section of Chapter Five, and made substantive changes throughout the paper.

This thesis complies with the regulations of the College of Arts at the University of Glasgow, and it comprises around 80,000 words including the main text, references, and appendices and does not exceed the maximum of 100,000 words.

Introduction

"[...] it is the source of the basic misconception of modern philosophy, that the task of philosophy is to bridge an ontological and epistemological gulf across which the subjective and the objective are supposed to face one another"
McDowell (1995, 889)

1. *Preamble, Or: The Story of S. & The Epistemic Garden of Eden*

Once upon a time there was a man called S. S. was just like any other man, and like any other man he was unlike everyone else. Perhaps not unlike others, he had very common beliefs, beliefs everyone has. He believed that 2 plus 2 is 4, for instance, that cats don't grow on trees and that water boils at 100 degrees Celsius. Naturally, some of the things he believed he also knew. If he knew anything, for instance, he knew that *this* is a hand, as he sometimes liked to put it, raising his hand, and that *this* is another —after all, couldn't he just see it? In this way (that is, just by paying heed to the testimony of his senses) S. had come to know almost everything he knew. And S. knew *a lot*.

S. was a man, for sure, but somehow he wasn't a 'man' in the sense we use the word to distinguish it from, say, 'woman' —not in any gendered sense, that is. Or so he liked to think. Rather, S. liked to think of himself as *the* man, as in the 'man' we think of when we think of *Man*kind. He was this man just like any other. Perhaps because of this, he didn't really have anyone else. Not that he needed them, of course —he was a *man* after all. He was absolutely self-sufficient. He would hunt his own food, chop his own wood, build his own house and laugh at his own jokes. True, S. was a lonely man too. So lonely, in fact, that on wet autumn evenings, by the warmth of a log fire, fingering pensively the waxy flame of a candle, sometimes horrible thoughts would assail S.'s mind: "What am I?" the thoughts would bang "Is this all just a dream?". On days like this, it was as if nothing were real to S. One night, the terror got hold of him so fiercely he almost fell from his armchair. But S. was a strong man, and a healthy one too, and he knew not to let the disease of these thoughts run pathologically unchecked. A man should not let doubt creep in. "A healthy mind is in a healthy body", S. would tell himself in these moments, and spend long hours in his solitary woods, swim naked in cold silvery lakes, listen to the songs of the birds or rest his eyes on the relaxing colours of a sunset. Often, this was enough for S. to feel again as if he was part of it all.

The story of S., unremarkable and caricatural as it is, presents an image of the subject who has been at the centre of, and of inspiration for, much western epistemological theorisation. This image is useful, I think, because it helps tease out some underlying features of the metaphor that was tacitly endorsed by key thinkers from the recent history of epistemology, and which is now laid before us —for better or worse— as the foundation and starting point for thinking about epistemic matters. The story of S. makes salient two features of this image in particular:

the contrast between the inner realm of the Subject and the external world of Objects, and the contrast between the Subject and Others. These two features are important because they have contributed to create a powerful illusion: the illusion of the possibility of a subject like S., someone who thinks, represents and dominates the world, and who can do this without the help of others. The material world, with its social, political and ethical complications, stands passively before S. as if beyond a gulf, open to be known and exploited. If the subject is perhaps not prior to the world of objects and to others, the illusion tells us, nonetheless that's what is given to the subject. They lie on the other side of a gap that the subject must be shown to be able to bridge in order to enjoy communion with them.

This is not a new story, and I am not the first one to tell it or —which is more relevant— the first to criticise it. So much so, in fact, that much contemporary Western anglophone epistemology can be seen, and correctly I believe, as a reaction to it. After all, isn't it true that social epistemologists have finally given 'other people' their pride of place in the discipline? And haven't epistemic externalists, among others, finally split the curtains of the Cartesian theatre and freed the subject from its crippling doubts? Perhaps they did. My worry, however —and perhaps this should sound as a warning— is that contemporary epistemologists, in the hasty attempt to overthrow the metaphor of the story of S., are running the risk of replacing it with another, even more subtle, illusion.

Let me explain. The attack of contemporary epistemologists against the myth embedded in the story of S. occurred on two fronts: one focusing on core (read: traditional) epistemological issues (i.e., the nature of warrant and justification, truth and knowledge), and the other focusing on new problems concerning the role of other people as sources and subjects of a collective epistemic enterprise. On the former front, epistemologists confronted the sceptic on issues concerning the relationship between the subject and the world, and about the validity of our system of rational evaluation. On the latter front, the one battling for the liberation of the subject from its isolation from other subjects, epistemologists have started theorising about the epistemic dependance of S. from other people (for instance, as sources of information), and about the importance of groups and collective entities as epistemic agents in their own rights.

The main outcomes of this radical assault on the story of S. have been, on the one hand, the formulation of 'modest' varieties of classical foundationalist answers to the problem of scepticism, and the flourishing of a tripartite conception of the sociality of knowledge on the other. The breakthrough of modest foundationalism has been to reject the sceptical demand that our knowledge of the world should be built starting from a subject's inchoate mental impressions of the world. According to modest foundationalism, the world of physical objects doesn't hide beyond a veil, accessible only through rational inference from the colourful pictures hanging inside one's minds, like pictures on a wall, but it manifests itself whole in perception.

Having received the world as a gift from the modest foundationalist, the subject of this new picture could finally make headway on another crucial philosophical puzzle, the 'problem of other minds'. More exactly, the main advancement consisted precisely in undermining the problem of other minds *as a puzzle*: like the world had come to be seen as presenting itself 'in the flesh' to the subject in perception, so too the existence of other minds stopped being considered as something fundamentally in need of validation. People, with just as complex a mental life as S. himself, could finally begin to populate his existence —as individuals, as part

of a group, or as members of a broader social network. As a result, three main conceptions of sociality were born from this new image: an individualistic one, which took the epistemic relevance of others to be about the ways in which their presence can influence and determine one's (routes to) knowledge; a collectivistic one, concerning the ways in which groups of people can themselves be considered epistemic agents in their own right, forming beliefs and coming to have knowledge just like individual agents do; and a transactionalist one, focused on the way epistemic agents interact with each other in complex social networks.

Targeting the two main features of the story of S., the two fronts have (I believe) successfully torn down its simplistic conception of our epistemic relationship with the world and with others, and have profoundly reshaped epistemological theorisation. In this new image, the relationship between subject and object, and between subjects, is not considered as a puzzle needing to be figured out from the controlled space of a selfless ego. Instead, this image has allowed philosophers to take seriously the role of other people in making our epistemic life what it is —the rich, complex, layered network of resources and exchanges we participate in in our daily lives.

Despite the significant progress, the relationship it establishes between core epistemic issues on the one hand, and social and other-regarding issues on the other is, I believe, at the root of a fundamental misconception in this new image. For how I see it, and very roughly, the issue is that the modest foundationalist who pulled S. out of his solitary dream and into the world didn't bother to make sure he had company. As if it was possible, somehow, to bring the subject before the world without other people. As if it was possible, that is, to come to a satisfactory conception of core epistemic notions —such as knowledge, truth and justification— without taking in consideration, *right from the start*, the relationship of mutual dependence between epistemic agents and their position in the normative fabric of a socio-political network.

This assumption, embedded in this new image of epistemological theorisation, risks generating a misconception that still carries the unpleasant mark of the story of S.: the thought, that is, that our epistemic relationship with the world is somewhat 'private'. Perhaps not 'private' in the sense of *internal*, where this is understood as a space, opposed to the mind-independent space of physical reality, to which the subject has a somewhat special or privileged access. It is, however, still 'private' as opposed to, say, *public*. For this new image knowledge is, despite its ineliminable social dimension, and fundamentally, an intimate connection between an isolated 'I' and their object. An intimate connection to which social relations are attached only *post facto*, as an important, but tangential, additional consideration.

Here then is what I take to be the illusion that is borne out of the new image: the illusion of an 'Epistemic Garden of Eden', where the subject, like Adam, the first man to be brought into the world, enjoys the (epistemic) delight of a material life. The Epistemic Adam, contrary to S., is not threatened by an *in principle* separation from the world —he has breached the veil of perception and enjoys the favour of the world. His epistemic abilities allow him to come to know the world, and in this world he finds the company of other people. With them he shares, stores and accumulates his epistemic capital, in the form of knowledge and truth, and comes to dominate the world much in the same way in which he would do by himself —although, perhaps, not to the same extent.

My thesis, quite simply, is that we should reject the metaphor of the Epistemic Garden of Eden if we want to make justice to the possibility of a *thicker* notion of epistemic sociality. We should reject, that is, the separation between core epistemological issues and other-regarding considerations, and we should oppose the reduction of the sociality of knowledge to the tripartite conception borne out of the myth. In other words, we should reject the idea that epistemology is solely (or primarily) about agents —their psychology (as per the story of S.) or their social conduct (as per the myth of the Epistemic Garden of Eden). Instead, my proposal is that we shift the centre of epistemic theorisation from agents to *environments*. In an agent-centric image of epistemology, knowledge is, at its core, a matter of a private relationship between a subject and the world. For this picture, 'knowing' is a form of acquisition, or possession, with respect to which the role of other people, however central, can only be instrumental. In an environment-centric image, on the other hand, the relationships between epistemic subjects, and between them and the world around them, is what constitutes the socio-political fabric of the normative epistemic framework in which they participate. Knowledge, in this picture, and very roughly, can be thought of a privileged position enjoyed by the relationship between the nodes (like, say, a subject and their object) that constitute the socio-epistemic matrix of an epistemic environment. If that's true, the revolution envisaged by an environment-centred image of epistemology consists in a subversion of the standard order of explanation: instead of taking people to be instrumental for the acquisition and the handling of knowledge, according to this image it is knowledge *itself* (as well as other core epistemic notions) that should be understood as a *function* of the complex structure that organises epistemic subjects in the environment.

…or so, at least, I would like to argue. For this idea —the idea, that is, of an 'environment-centred epistemology'— has not yet come to its full maturity, and it doesn't find complete expression in this thesis. What I am suggesting is that the works collected in this thesis should be seen more as expressions of the journey that has brought me to the formulation of this thought rather than as attempts to defend it. As a result, the narrative structure of this thesis is, as it were, 'turned on its head': the clarification of a thicker notion of sociality, and the role it plays in determining fundamental epistemic concepts, remains the 'end goal' of this work or, better perhaps, their 'vanishing point'. In clarifying this end goal, then, my hope is that this introduction will be able to provide the reader with a map to help them navigate the work here collected, and identify the place they occupy in this larger trajectory.

To do so, in the next section (§1.2) I will retell the dialectic movement between the two metaphors I have introduced in very general terms in this preamble (i.e., the Story of S. and The Epistemic Garden of Eden) in slightly greater detail. If all goes well, this should help delineate my thesis more precisely —not though 'positively', by offering arguments in its favour but, as it were, 'contrastively', by making more vivid the view(s) I take my proposal to stand up against. Finally, I conclude in section §1.3 with a brief summary of the content of each chapter.

2. *The General Picture*

"[...] only in an emancipated society, whose members' autonomy and responsibility had been realized, would communication have developed into the non-authoritarian and universally practiced dialogue from which both our model of reciprocally constituted ego identity and our idea of true consensus are always implicitly derived. To this extent the truth of statements is based on anticipating the realization of the good life."

(Habermas 1968, 314)

We make mistakes. Sometimes we make egregious mistakes, and other times we make just small ones. We do it more or less intentionally, with a more or less guilty consciousness, but we do it often. We get things wrong all the time. I personally have not yet learned to drive from point A to point B without getting lost at least once. It's just that things aren't always what they seem to be. In my case, no matter how much it really seems to me that *that* is the motorway exit my navigator tells me to take. Most likely, I have come to learn, it simply isn't. Alas, our senses deceive us. What our senses, our reason, our friends (and, I'm sure, sometimes also our navigator) tell us doesn't always match up with how things really are.

And yet we seemingly manage to carry on with our lives just fine. We get things wrong, and we do that often, but not all the time. By and large, we do get things right. But the point is: how can we tell? How can we make out the appearances that do match up with reality and those that don't? One obvious way to answer this question is to say that, well —we can just see it. Although it might have seemed to me that the motorway exit I took was the right one, it doesn't take me too long to realise I was mistaken. I can *see* it now. I can see that the next sign is for Cumbernauld Town Centre (of all places) and not for Perth. To say this, is to say that we have a procedure that guides us: perception is one of the things we can rely on when it comes to distinguishing good and bad judgements, for instance.

This seems reasonable. And yet: how do we know that perception is up to this task? Isn't it perception itself that leads us astray in the first place? Why think we can rely on the same instrument that deceived us? Again, one could say we can tell that perception is a good procedure because we know that, for the most part, its deliverances are in fact accurate. But clearly this is not a satisfactory answer —for if we could use our knowledge to measure the reliability of our method, then what would be the use of this method in the first place? And so we are caught in a circle: we started off trying to understand how to tell whether we get things right or wrong, and we said that perception is one of the criteria we use to do that. But then, when asked to justify our choice of this method, we ended up replicating Baron Munchausen's heroic gesture, and used our knowledge of the very things we had set out to justify to tragically lift ourselves out of our theoretical swamp.

We started out with a simple question, and now we already find ourselves at a dead end. We take ourselves to have knowledge of all sorts of things, despite our fallibility. And yet, as soon as we start looking more carefully at the things we know and the things we only thought we knew, the difference between them seems to vanish. We can't say that we know what we think we know unless we have a valid method, and we can't be sure that our method is valid unless we have some piece of secure knowledge to verify it. It is as though we were stuck in between two demands that seem reasonable, each of which uncomfortably depend on the other for its satisfaction. And so we are forced to iterate an infinite circular movement from

one to the other, until our rational grip on reality loosens and, caught in this wheel, we start questioning our understanding of the most basic facts of everyday life. How is this possible? What went wrong? And how can we get off this wheel?

The 'problem of the criterion' or the 'diallele' (this is the name with which Roderick Chisholm has popularised this old philosophical puzzle) hinges on two separate, although seemingly related questions, regarding the extent (1) and criteria of validity (2) of our knowledge:

1. What do we know?
2. What is the criterion in virtue of which we know what we know?

The problem consists precisely in the thought that in order to answer one of the two questions, one must have an answer to the other, and vice versa. In order to be in a position to tell what we know, we need to provide a valid criterion; and in order to guarantee the validity of this criterion, we need to rely on what we know. If we accept the validity of the puzzle, our only way out of this circle, according to Chisholm, is to pick our favourite point of departure; once an answer to one of the two questions is established, the answer to the other can simply be derived from it. Chisholm proposes to call 'methodists' those who start by committing to an answer to (2), and 'particularists' those who start from (1). For the methodist, it is the method, the criterion, what helps us distinguish appearance from reality. The particularist, instead, prefers to start from particular items of knowledge —appearances whose veridicality we can be sure of even if we can't cite a secure method by which we have obtained them ("Here's a hand, and here's another").

So far so good. The confusion we felt when we were caught in the grip of the problem of the criterion was generated by our inability to make out the way something appears to us from the way it really is. Methodism and particularism now propose two strategies we can use to leave the impasse behind and begin to walk the distance between appearance and reality —the firm guidelines of a criterion, or the secure ground of items of knowledge we hardly ever doubt. The point, now, is to try to understand what kind of solution they offer to the problem of the diallele —or, better, what kind of move they propose within the broader dialectic of this argument.

Let's start from methodism. The methodist strategy is characterised by a confidence in the possession of a (reliable) method for discerning true from false judgements. Simple empirical observations seem to reassure the methodist: after all, when we mistake our coat hanger for an unexpected guest, it is still thanks to our perception that we are able to resolve the illusion. Indeed, it may seem that the wise thing to do, even though it sometimes falters, may just be to trust our senses. But how legitimate is this confidence? Suppose we asked the methodist to motivate their confidence in their method —what reasons would they give us? One option we can safely exclude: they wouldn't want to ground the validity of the method on its results. The pieces of knowledge we obtain thanks to the application of the criterion cannot be used as a warrant for its validity, on pain of abandoning methodism for particularism, and setting the wheel back in motion. The methodist who wants to offer a solution to the problem of the criterion, then, will naturally keep away from this option.

Alternatively, the methodist could ground the validity of their method on another method —say, testimony, or inference. This, they would say then, is the real criterion. The question though would arise again about the validity of this further criterion. Unless something other than another method could be found to answer that question, it is clear that the threat of an infinite regress should discourage the methodist from pursuing this option as well. But what else could the methodist offer as a guarantee for the validity of their method? If they cannot rely on the knowledge they obtained thanks to the criterion (on pain of circularity), nor on any other method (on pain of infinite regress), the only option left for the methodist may just be to dogmatically assume that their method is a good method. Since it rests on no secure ground, however, this assumption would be arbitrary, and endorsing it would ultimately be tantamount to committing to the fundamental irrationality of the choice of the criterion. And that's not a good solution.

Perhaps then the methodist's solution to the problem may not be as straightforward as we might have initially thought. But then why not opting for particularism? For recall that, unlike methodists, particularists don't think we need a valid rational criterion for figuring out whether we know something. If I know anything at all, the particularist would say, I know that this is a hand, for instance: and to know this I need not be concerned by whether I do in fact possess a (valid) criterion for this judgement. Or so the thought goes.

But suppose you were dreaming, we could ask the particularist. In this dream, everything looks just as it normally looks to you in your conscious daily life. You are coming back home, wearing your usual clothes, worried about the next day's meeting, wondering what you'll have for dinner. Every detail in this dream looks exactly as it would look in your waking life. Now, suppose even you have a moment of doubt —something doesn't feel right. You look down, and you have the exact same experience you would have while looking at your real hands. "If I know anything" you confidently utter to yourself, "I know that this is a hand". The thought reassures you, and you continue living your dream experience confident you are a real person walking down a real street.

If in the dream scenario and our waking life everything looks just the same to us, there is no way to tell whether we really are in either. But then how can one claim to have knowledge of even the most ordinary empirical claim, like that 'this is a hand'? The problem here is a familiar one, and not one the particularist can hope to offer any straightforward solution to.

We started off from the consideration of a very simple fact: that we make mistakes. This fact, we noted, invites us to distinguish between true and false judgements, or, more generally, between the way something appears to us from the way it really is. With this distinction in place, however, we have found ourselves stuck in a whirlpool generated by two contrasting but mutually dependant demands: on the one hand, that we ought to know what is true or false in order to find a method to discriminate truth from falsehood; and on the other, that we ought to have a method for discriminating between them in order to know which is which. To overcome the impasse and free ourselves from the whirlpool, we considered two strategies: methodism and particularism. What we discovered, however, is that devastating sceptical worries threaten any attempt at crossing the fine line between appearance and reality: following methodism, we discovered that there is no valid criterion we can use to move safely

from the way things appear to us to the way they are. Following particularism, we were led to doubt the veracity of our knowledge of the most ordinary things.

Between the disconcerting whirl of the diallele and the chilling gorge of sceptical collapse, then, all we are left with is a lonely subject, incapable of giving substance to the deliveries of his senses, to his trust in others, or to any of his beliefs. Stripped of any conviction in the validity of his rational capacities, this subject is consoled only, perhaps, by the reassuring company of his thoughts, where the only fact he can hold on to with certainty (by the warmth of his log-fire) is this: that he feels lost, alone, and puzzled.

What now then? One way out of this impasse could be to say that the demand imposed by the sceptic against methodism and particularism is not legitimate. And it is not legitimate, the thought would go, because it understands methodists and particularists as offering a rational guarantee for stepping out of the wheel. Methodism, one could argue, is not rightly understood as posing the criterion as something that we can have good or bad reasons to adopt —that is, something that can itself be evaluated rationally. Nor is particularism offering the secure ground of self-evident pieces of knowledge as something that is compatible with radical error, as the sceptic suggests. Instead, the thought goes, methodism and particularism should be understood as proposing a much bolder response to the problem of the diallele: they propose to individuate a point (the criterion, or this or that piece of empirical knowledge) beyond which we can't meaningfully bring ourselves to exercise rational doubt. This, one could say, is the place where theorisation begins.

If this is true, there is something very important that scepticism gets right: the sceptic is right in thinking that the demand it advances cannot be satisfied. It is true, according to this line of thought, that there is no assurance we can find —by reason alone— that things really are what they seem. To attempt to do so is to commit a very simple mistake —namely, the mistake of expecting a system to be able to provide the rationale for its own validity. The sceptic is right to think that if we are too confident about what we can obtain through the unaided resources of our reason, all we are left with is nothing but the comfort of a lonely doubting self. The problem with scepticism, however, is that it takes the satisfaction of this demand to be necessary for us to be able to walk the distance between appearance and reality. But —so the thought would go— we already know before we can say why or how. Even if we don't know that we know, we do already know.

If this is right, then, particularism and methodism should be seen as proposing to reject the problem of the criterion, rather than as offering a solution to it. Theirs is a stance one is invited to assume when reflection on the problem of the diallele and the sceptical collapse it threatens brings to light the limit of our rational grasp over the world. To bring doubt to bear on the nature and limits of our knowledge is to illegitimately extend the 'claim of reason' beyond its scope. The methodist displays a confidence in the validity of their method, and the particularist is sure about their knowledge. This confidence and this certainty, however, have no ground to stand on:

> We learn and teach words in certain contexts, and then we are expected, and expect others, to be able to project them into further contexts. Nothing insures that this projection will take place (in particular, not the grasping of universals nor the grasping of books of rules), just as nothing insures that we will make, and

> understand, the same projections. That on the whole we do is a matter of our sharing routes of interest and feeling, modes of response, senses of humour and of significance and of fulfilment, of what is outrageous, of what is similar to what else, what a rebuke, what forgiveness, of when an utterance is an assertion, when an appeal, when an explanation—all the whirl of organism Wittgenstein calls "forms of life". Human speech and activity, sanity and community, rest upon nothing more, but nothing less, than this. It is a vision as simple as it is difficult, and as difficult as it is (and because it is) terrifying. (Cavell 1976, 52)

Scepticism is just a natural possibility of our condition —i.e., the condition of creatures who entertain a relationship with the world that can never be fully rationally clarified (that is not 'one of knowing'). In this condition, according to Cavell, we cannot avoid the terrifying vision of the gap that opens between what is given to us in appearance and the world of objective reality. Between these two extremes, we can rely on no assurance of the safety of our path. The sceptical pretence that this assurance is needed to bridge this gap is the natural response of fear in the face of this terror.

When we lean into this fear, and let its demands rule the standards of our conduct, scepticism becomes a disease, the pathologization of a healthy 'cognitive immune system', where the mechanism "designed to protect our conception of the world from harmful errors turns destructively on that conception itself" (Williamson 2005, 1). Against the onset of this disease, and to soothe our terror, the ailment offered by methodists and particularists is simply to resign ourselves to the limits of our rational claim over the world, and accept that, if we want to learn to walk the distance between appearance and reality, we cannot do that by our own unaided resources. We cannot do that, that is, without needing the world to 'do us a favour'.

We started off with what looked like a simple and legitimate request: to find a criterion to tell apart truth from falsehood. This request, we have found, is a request to extend the claim of our reason beyond its scope, and is thus anything but simple or legitimate. So we rejected it. If this is right, all we can say is that sometimes, if everything goes well, what we thought about the world turns out to be true, and we have knowledge. Sometimes, when we are less fortunate, appearances mislead us. That's when we make mistakes. Asking for more, so this picture suggests, is asking too much.

Here's the picture then: our condition —the one that makes scepticism a 'natural possibility' for us, like a vulnerability to bacterial infections, something that needs to be kept in check to prevent the spread of the disease— is the 'terrifying' condition of a lonely subject, stuck at the edge of a precipice, beyond which lays the material world. Nothing we can resort to from our position can help us cross the distance over the gulf and into the world. The best we can do (and this is the key move proposed now, as a way of understanding the methodist and particularist response to the problem of the criterion) is just to rely on the world itself to carry us safely to the other side. This, I take it, is the most natural way of interpreting the idea that we need the world to 'do us a favour': we need to acknowledge the fact that we depend on the world to cover the distance between appearance and reality.

At this level of generality, the form of this relationship of dependence is still open to very different interpretations. Very roughly, these interpretations come in two main varieties,

depending on their 'direction of fit': from the subject to the world, or inside-out, and from the world to the subject, or outside-in. One way of understanding these interpretations is as two different 'kinds of favour' one can ask the world: those who prefer to start with the way things appear to us will posit that appearances do generally speak of the world. This is the 'inside-out' direction. Those who prefer to start from external reality, on the other hand, will posit that it is indeed the world itself that provides us with our appearances. This is the 'outside in' direction.

So as I have presented them, these two positions are specular and, from a dialectical point of view, equally legitimate. (Naturally, this is not to say that there is a perfect symmetry between them, or that they are vulnerable to the same problems, and naturally it is not to deny the possibility of hybrid positions combining elements of the two). For instance, for inside-outers, what matters the most is what is available to us, the way things appear, and how we make use of them. Then, when the way things look to us really are about the world, we are compensated for our good conduct and we get knowledge. Because they give primary importance to our rational conduct, views that belong to this family are often well positioned to vindicate the normative aspect of our epistemic relationship with the world —i.e., the sense in which knowledge claims position themselves in a space where they can be evaluated, where reasons can be demanded, beliefs permitted or prohibited, and where one can be held responsible, and be blamed, for their beliefs.

Naturally, the main challenge for proponents of this view is to explain how, or why, one should take the way things look to them, and the way they have used them in reasoning, as a good indicator for how things really are. Couldn't someone have conducted themselves in perfect accord with the norms of rationality and be radically mistaken? The worry, in other words, is that proponents of this view risk making the contribution of the world to the rationality of our beliefs somewhat mysterious —something that adds on to our conduct, but that doesn't seem to have any intuitive connection with it. Call this the 'external-world problem'.

'Outside-in' type views, on the other hand, take this connection as their starting point. This makes it easy to explain the world's contribution to our knowledge. For proponents of this view, we enjoy a felicitous attunement with the world, and it is this objective fact that gives our beliefs the epistemic status they have. Beliefs need only walk down the right path from the world to the subject in order to achieve the status of knowledge. The advantage they gain on this side, though, they appear to lose on the other. For if what ultimately matters for a belief to be justified (and become knowledge) is the aetiology of its journey, where this is understood as a natural fact described by the sciences, the status it obtains, be it positive or negative, is not any more a standing in the normative space of reason than it is the temperature reading on a thermometer, or the ring of an alarm clock. Call this the 'normativity problem'.

'Inside-outers' and 'outside-inners' constitute the two main families of views about epistemic justification in contemporary Western analytical philosophy, broadly overlapping with what are typically referred to as internalist and externalist accounts of epistemic justification and knowledge. Part of the goal of this thesis will be to put pressure on the two faces of this general response to the problem of the criterion. Throughout the thesis, I do this in two different ways. Chapters 2, 3, 4 and 5, which make up the bulk of this work, have a more constructive role: here is where I look at specific cases —like the implementation of machine

learning-based technologies in our societies, the role of gender and race in cases of group disagreement, the phenomenon of mansplaining and instances of group-based beliefs—where facts about an agent's social positioning, their relationship with other people, or about the political organisation of their environment, seem to offer precious insights into the way we theorise about fundamental epistemic concepts. The aim of these chapters, then, in a nutshell, is to bring to light the importance of socio-political factors for epistemological theorisation.

The role of chapter 1 and 6, on the other hand, is more critical, at least partly. Chapter 6, for instance, looks at one of the most promising attempts to tackle the 'external world problem' for 'inside-out' views, and offers a sustained criticism of their proposed solutions. Chapter 1, on the other hand, confronts 'outside-in' views that propose to accommodate the 'normativity problem' by making important concessions to the sociality of knowledge. More importantly, this chapter also begins to sketch a more positive proposal, which represents perhaps the most explicit attempt at formulating the main thesis of this work. The idea, simply put, is that of a 'paradigm shift' from an agent-centred normativity, to an environment-centred one. In the attempt to reframe epistemic normativity along these lines, I see the wider ambition of this thesis being brought to light —namely, to make space for an image of the epistemological domain as fundamentally political.

3. Summary of the Chapters

This thesis comprises six chapters, each addressing a specific topic in contemporary epistemology. Although I see them as united by a common interest in challenging some core assumptions of individualist epistemology, the works here collected offer stand-alone arguments, and can be read independently from each other. In what follows, I offer a brief summary of the debates in which the chapters position themselves and their main contributions to those debates.

Chapter one connects the literature on white ignorance (Mills 1997, Spivak 1999 and Frye 1983), and some recent advances in the debate on epistemic normativity, especially as discussed by Goldberg (2017, 2018), Chrisman (2020, 2022), Lackey (2016, 2021) and Simion (2024). The importance of building a bridge between these debates is both theoretical and political. Its theoretical importance stems from the fundamental insights into the nature of epistemic normativity I take to be offered by what came to be known as 'the epistemologies of ignorance' (Sullivan and Tuana 2007). In particular, starting from a case of structural white ignorance discussed by Martín (2021), I have attempted to draw inspiration from the structural-level normative judgement that are commonplace in political theory (in particular, looking at Rawls 1971, Anderson 2012 and Young 2011) to theorise about the nature of epistemic normativity more in general. In particular, in this paper I offer a first attempt at modelling epistemic normativity on the normativity in the political sphere. Within the broader scope of this thesis, one way of interpreting this move would be as offering key tools to address what I have called the 'normativity problem' faced by outside-in views. In addition to its theoretical aim, I take the contribution of this paper to be also, and perhaps most importantly, political. White ignorance is a lively discussed topic in critical race theory, but it still lies at the margins of mainstream epistemological theorisation. By focusing on white ignorance, then, the goal of this paper is to reclaim the centrality of the feminist and decolonial project of bringing to light structures of systemic epistemic oppression for the development of epistemology *tout court*. I am convinced that mainstream epistemology and,

more generally, the philosophical tradition that calls itself 'analytical' suffers, as a whole, from the same problem I want to draw attention to here —that of systematically ignoring the role of structural-level considerations and of the plurality of epistemic perspectives in shaping our epistemic lives. By calling the attention on the phenomenon of white ignorance in particular, then, my aim is to criticise the forceful universalisation of normative theories that, in virtue of their ignorance of the particular, colonial, cultural and historical contexts in which they originate, are harmfully applied to all contexts.

If the first chapter addressed what I named the 'normativity problem' for 'outside-in' type views, the concluding chapter instead addresses the specular problem for 'inside-out' type views. More exactly, this chapter focuses on one of the most promising strategies to solve the 'external-world problem'. This strategy, known as hinge epistemology, is characterised by commitment to a set of sui generis propositions ('hinge propositions') which constitute the limits and the conditions of validity of our epistemic practices. According to some popular formulations of this view, it is our pre-rational commitment to these propositions as the rules, or hinges, of our system of rational evaluation that guarantees the appropriateness of our movement from the way things appear to us to the way things really are. This paper offers a sustained criticism of two main articulations of this view. To the extent that hinge epistemology represents one of the most relevant options available to internalists to avoid sceptical collapse, the results of this discussion contribute to cast a grim light on the chances of a successful defence of internalistic minded notions of epistemic justification. Although the aim is purely critical, then, if correct, the argument I offer in this chapter has a profound impact on current epistemological theorisation more broadly.

The central essays in this thesis (chapter two to five), represent a heterogeneous, but unitary, contribution to the existing body of literature in social epistemology that challenges the individualist trend of the epistemological tradition. In particular, chapters two and three start each from actual and timely phenomena to make the case for important extensions to the notion of epistemic injustice and advance feminist theorisation. Chapter two in particular focuses on the widespread and often unchecked implementation of machine learning-based technologies in our everyday lives. The main contribution of this paper is on two fronts: first, it identifies a fundamental epistemic fault (what I call 'epistemic conformism') in the way modern machine learning-based AIs are designed and function. Second, it draws a direct connection between the systemic flaws of these technologies and the structural epistemic harms that they contribute to generate and sustain. These harms are both old (like forms of testimonial and hermeneutical injustices) and new (either new kinds of epistemic injustice, like epistemic spurning, or new categories of epistemic harms altogether, like forms of zetetic injustice). With regard to the former front, this essay addresses an important gap in the literature on the philosophy of AI, making available crucial tools and indications for the development of a more just AI. With regard to the latter, by drawing connections between technological advancement and epistemic oppression, it tells an important cautionary tale about the future of research in AI and its impact on our society.

Chapter three (coauthored with Daniela Rusu) looks at the phenomenon of mansplaining. The aim in this paper is to offer an analysis of this concept that meets two important criteria concerning the extension of the phenomenon and its role in upholding and reinforcing existing systems of gender-based oppression. This is done by proposing to understand mansplaining as occurring at the intersection between assertorial violence and

epistemic oppression. On the linguistic side, we take mansplaining to consist in a violation of a norm of cooperative conversation. On the epistemic side, we take mansplaining as perpetrating a form of epistemic injustice. More precisely, the thought is that mansplaining consists in a violation of the answerability norm of conversation by excess of explanation, where this results in a speaker (the mansplainer) treating their interlocutor as their epistemic inferior, thereby degrading their epistemic status in the community to which both interlocutors participate.

In the final two chapters, I develop a new account of group belief (chapter four) and implement it in a particular case of group disagreement (chapter five). The central insight of the view I defend in chapter four is to extend to groups a functionalist analysis commonly adopted to account for beliefs at the individual level. More precisely, I argue for a weak inflationist version of functionalism about group belief that, I argue, makes available an attractive middle ground between the two main opposing grounds in the literature on the topic: the one between monism and pluralism, and, in the monist camp, between inflationists and inflationists accounts of group belief. In chapter five (coauthored with Mona Simion), I apply this model to group disagreement, and show how it can help deal with special instances of the phenomenon that are particularly problematic for more traditional individualistic accounts. We start by considering cases of group disagreement that constitute epistemic injustice. These cases, we argue, generate problems for extant internalist accounts of group disagreement, which motivate two desiderata: one concerning what it takes for the disagreeing party to be considered as peers, and the other about the normative evaluation of prejudiced beliefs. We conclude by arguing that by adopting a functionalist account of group belief (as defended in chapter four) and of group justification (as defended by Simion 2019) both desiderata can be met.

Chapter One

The Environmental Image: The Case of White Ignorance for Epistemic Justice

ABSTRACT

Some epistemologists believe that epistemic agents sometimes ought to be sensitive to evidence they do not possess —or, as it is often put, that there are things that they should have known. If that's true, an interesting consequence is that, by being ignorant of a fact, one can sometimes be in violation of an epistemic obligation. More generally, the relationship between ignorance and epistemic obligations has recently offered fertile ground for attempts at redrawing the boundaries of epistemic normativity. So far, however, the literature has centred on a very narrow conception of ignorance. In particular, attention has been dedicated almost exclusively to instances of agential ignorance —that is, ignorance that is due to the shortcomings of the ignorant agent, whether at the level of their cognition or their will. In this paper, I argue that consideration of cases sanctioned white ignorance (Spivak 1999, Martín 2021) suggests a new picture of epistemic normativity that takes epistemic environments as their centre. In shifting the perspective from agents to environments, the main goal of this paper is to outline a normative framework in which, drawing inspiration from the political sphere, epistemic goodness can be understood in terms of epistemic justice. If, following Rawls (1971), we take justice to be 'the first virtue' of political institutions, in the attempt to reframe epistemic normativity in this way a wider ambition of this paper also is brought to light —namely, to make space for an image of the epistemological domain as modelled on the political domain.

We ignore a lot. For instance, there are things we ignore just because it so happens. After all, we don't know everything. In fact, we just *can't* know everything, and so it's only natural that there are things we ignore. I don't know how many leaves the pine tree in front of me has. I don't know how many times I've uttered 'by the way' in the last hour, or what the exact distance between the Sun and the Moon is right now. Not that these facts don't matter —it's just that they don't matter to me now. In this sense, my ignorance of these facts is just *contingent*.

Then there are things that I *really* don't want to know, even if I *really* should. It doesn't matter how urgent the call, how favourable the conditions, or how abundant the evidence: I just won't be persuaded to believe them. This ignorance is not just a rational misstep in my otherwise commendable cognitive stride, something that can happen *to* anyone. It is something that I am *doing*.

That's one way in which ignorance can be *active* —i.e., when it is *wilful*. We often hide from more or less uncomfortable facts, more or less voluntarily. But then there are also things that are hidden from us, and that we fail to know even if they matter to us, and even if we wished we knew them. That's *another* sense in which ignorance can be active —although perhaps in a loose sense of the word. That is, when it is not the result of my doing (or at least, not wilfully), but because it is enforced onto me by the morphology of my environment. Ignorance here is active in the sense that it is brought about by the institutionalised practices that shape the epistemic environment. Following Spivak (1999), I shall call this *sanctioned ignorance*.

Cases of active ignorance (whether sanctioned or willful) seem to invite us to think normatively about ignorance. For instance, if I fail to believe the testimony of a female scientist on climate change *because* she is a woman, it is natural to think that I am ignoring something that I *should* have known. Philosophers are divided as to whether the normativity displayed in this type of case is distinctively epistemic. Sceptics (like Wrenn 2007, Nelson 2010 and Nottelmann 2021) think that the parallelism between the epistemic and the ethical domain breaks down precisely when it comes to individuating positive epistemic obligations of this sort. They think that there is nothing we ever ought to believe, on purely epistemic grounds.

A growing number of epistemologists, on the other hand, take cases of active ignorance to offer new and interesting insights about the extension and the nature of epistemic normativity. Some of them (call them *individualists*), for instance, claim that facts about individual epistemic agents (e.g., their position in the environment, their cognitive abilities, and so on) determine what kind of things they ought to be responsive to, epistemically (Simion 2023, Lackey 2016). Others (call these *socialists*), instead, attempt to stretch the boundaries of epistemic normativity beyond individuals, and give central space to the role of social expectations and interpersonal relations as sources of norms of epistemic conduct (Goldberg 2016, 2018, Lackey 2021, Simion and Kelp forthcoming, Kelp 2023).

In line with this social turn, this paper introduces a new parallelism between political and epistemic normativity. With socialist accounts, I agree that a full appreciation of the richness of the epistemic normative domain requires us to look *beyond individuals*. Contra

socialists accounts, on the other hand, I take this to suggest that we should move *beyond agents themselves*. More exactly, by analogy to the normativity of the political sphere, this paper argues that epistemic normativity should be understood as regulating not only the conduct of individual and collective epistemic agents, but also, and perhaps most fundamentally, the distribution, management and access to epistemic resources in the environment.

The consequences of this move are large-scale, and involve a fundamental redrawing of the boundaries of the epistemic domain. The new image of epistemic normativity that emerges, what I call the *environmental image*, opens the way for a new family of epistemic norms. Drawing inspiration from a broadly Rawlsian view of political normativity (from Anderson 2012 and Young 1990, 2011), I take these norms to be centred on the notion of *epistemic justice*, where this amounts to the goal of promoting the epistemic betterment of the epistemic community as a whole.

The plan is as follows. In the first section, I start with a paradigm example of a case that calls for the kind of expansion of epistemic normativity I want to argue for. Here I construct an analogy between the political and the epistemic, and begin to make the case for a kind of structural normative judgements that are distinctively epistemic. In §2 I briefly survey attempts at making sense of this new dimension of epistemic assessment by individualist and socialist accounts of epistemic normativity. §3 concludes the argument by outlining the environmental image, and sketching its implications for epistemic normativity at large.

§1. THE CASE

Consider this case:

> DOCTOR Doctor D believes *q*: that drug X is effective to treat their Indigenous Latina patient, P. This belief is the result of very meticulous research. Doctor D has done all they could to come to this conclusion, which was unequivocally supported by all the evidence available to them (e.g., books, articles, medical practice). However, drug X has been trialled only on White Europeans and, as it turns out, since it acts on portions of the genome that present slight differences between Europeans and Indigenous people, it is ineffective for the latter group. Consequently, the doctor ignores the fact *p*: drug X is not effective on Indigenous people[1].

The doctor should have known that the drug isn't effective, and the fact that they don't is *bad*. Morally, for sure, because it leads to the discrimination of a group of people on the basis of their ethnic background. This is how bad it is, morally: it contributes to furthering racial discrimination and oppression. But the doctor's ignorance is also bad *epistemically* —or so at least I shall argue, on the grounds that the doctor does not have access to evidence that *ought to* have been made available to them.

[1] This case is adapted from Martín (2021)

The problem, of course, is to decide if there is any distinctively epistemic way of making sense of this 'ought'. Traditionally, epistemologists have been convinced that, epistemically speaking, one should guard themselves only against *things* (read: evidence, reasons) that are psychologically available to them. In a slogan: I can't be said to be epistemically in the wrong if I ignore evidence I don't have or that is inaccessible to me. In this particular case, the evidence is not only unavailable (to the doctor or to anyone else), but it doesn't even *exist* yet (i.e., it hasn't yet been produced by the relevant research). If this is true, it seems rather mysterious how evidence that hasn't yet been discovered could create epistemic trouble —or so the thought goes.

Notice though that, at least in the moral sphere, there is hardly anything mysterious about the idea that normative pressure can come from facts external to the individual agent. The fact that they don't have access to the evidence doesn't seem to undermine the intuition that the doctor, in some morally relevant sense of the word, *should* have known. On the contrary, it is precisely *because* the doctor doesn't have access to the evidence that we think there is something morally problematic —the doctor *should have known*!

But we are treading on ambiguity here, so let's proceed with care. When we say that the doctor 'should have known', perhaps the most immediate way of reading this is as a command issued *at the doctor*: 'you, doctor, should have known, and by failing to do so, you have fallen short of fulfilling your *duty* as a doctor!' This is a normative judgement that centres on human conduct (that is, the doctor's), and it is concerned with what one does, or with the standards one ought to meet when one acts. This is obviously not the sense of 'should' that applies in this case though. Not just because we have stipulated that the doctor has done everything a doctor should do in this case but, even more strikingly, because the doctor's ignorance lies beyond what anyone could reasonably be expected to know in their circumstances. After all, the evidence has not been discovered yet.

In fact, *this* is precisely what we find problematic —that is (to borrow an expression from Arendt 1987) our concern is not with the *doctor*, but with the *world*. What we find problematic is the way in which medical research is conducted in that environment: for instance, the way in which evidence and information is organised and made available in the medical world, the way in which medical professionals set their research priorities, maybe the way in which funding, authority and power are allocated —a complex set of facts about the organisation of medical research that has caused some group of people in that community to be (at the very least, physically) worse off. Our concern, that is, is not only strictly speaking moral, but *political* —it is not primarily about the transactions of individual or group agents amongst each other ('What have they done?', 'Did they do it properly?'), but about the global state of a complex system; it includes the state of institutions and their norms, the allocation of power and resources, the access to and the amount of knowledge and information in the system and their fair distribution.

To distinguish it from the *transactional* normative judgement of personal level morality, I propose (following Rawls 1971) to call the kind of judgement we appeal to when we say, in this case, that the doctor 'should have known', a *structural normative judgement* —that is, a normative judgement that is concerned not just with the interactions between moral actors, but with the global state of a complex political (in this case, also medical) system.

The claim I want to put forward in this paper is that it is possible to make a similar distinction also in the epistemic sphere. Drawing an analogy with the political sphere, I

want to say that despite the preponderance of personal-level normative considerations in contemporary analytic epistemology, there is a structural-level, political normative pull that we are subject to as participants in a shared epistemic environment. To do so, I'll begin by saying something more about the specific moral/political problem raised by this case. This will offer the inspiration for the epistemic framework I will sketch in the final section.

I.1 Sanctioned White Ignorance

Once the focus on the personal-level decisions and the choices that are available to the doctor is blurred out, a new picture emerges where what matters, for assessing the failure in the DOCTOR case, is the complex interplay between personal- and institutional-level norms, the availability of relevant resources and their distribution. Consider for instance the long journey that brought drug X into existence, starting from, say, writing of a proposal motivating research on that particular disease, liaising with labs across the world to test the most effective chemical structures for the drug, getting involved in Genome-Wide Association Studies (GWAS) to identify the portion of the genome normally affected by the disease, publishing and disseminating the initial results, seeking partnerships with pharmacological firms to obtain funding and begin the long period of testing and trials —and so on.

When we bring our attention to this process, we can better understand how the doctor's knowledge, or their lack thereof, regarding some particular medical fact is, in some important respect, also a function of macroscopic considerations that only tangentially have to do with their skill *as a doctor*, and that instead concern the *quality of the medical environment* at large —including things like: the institutional and social pressure to produce new results, the incentive to conduct high-risk high-gain research, the availability of information and its ease of access, the inclusivity of research practice and its norms of justification and so on.

Here's then the sense in which we can understand the doctor's ignorance as a case of *sanctioned* ignorance. Sanctioned ignorance is different from mere 'passive' ignorance because *it is not accidental*: it doesn't just so happen that the doctor fails to know *p*. Every other doctor also fails to know that *p* in that community, and for a precise reason: because the drug has been tested only on White Europeans. It doesn't matter how hard doctor D or anyone else tries —there simply *isn't* information about *p* available in that environment.

So D's ignorance is not accidental, but systematic. One way in which something can be said to occur systematically is when it springs from a sufficiently stable feature of one's character —like, say, a strong conviction, a prejudice, a cognitive bias. Clearly, this is *not* the sense of the term at play in DOCTOR: the doctor's ignorance is systematic, but it doesn't have to do with anything the doctor has done, or the way they are. Instead, it seems more plausible to think of the doctor's ignorance as a 'trickle-down effect', the cumulative outcome of a series of transactions that nevertheless end up shaping the environment in a way that contingently but systematically blocks access to particular resources.

In fact, we can even suppose that, in the long chain of individual and institutional decisions that have taken place between the study of its chemical composition and its delivery to hospitals worldwide, little or no transactional wrong has been committed. Drug X has been tested by multiple medical firms across the world, by multiple research teams. Members of these research teams, we can suppose, were under pressures imposed by their

funders, for instance, or the demands of the market, or simply by prestige, or by other academic obligations, to produce certain kinds of results within a certain time limit. We can suppose, finally, that, considering their limits, they did the best they could, and that everyone involved not only acted morally, but that they did so in a genuinely human way, respecting more nuanced social and professional obligations, and with a genuine interest to offer a treatment to this rare disease.

In this sense, then, the doctor's ignorance speaks of a failure at a broader, structural level —that is, at the level of the organisation and management of the resources in that particular community. That's why it is *sanctioned*: because although it is brought about by the actions of individuals, it is not up to the individuals themselves, but to the system of more or less explicit rules, the power dynamics, the official norms and the habits that sanction this ignorance.

That's one important feature revealed by the doctor's ignorance. Another crucial feature is its normative dimension. The doctor's ignorance, sanctioned by the environment-specific practices of research, leads to some individuals being more likely to be offered worse medical treatment in virtue of belonging to a particular racialised group. In other words, this ignorance *harms individuals on the basis of their ethnic background*. For this reason —i.e., because it furthers racial oppression— I take the doctor's ignorance to be a form of sanctioned *white ignorance*.

Some may be sceptical about the use of the label 'white ignorance' in this case. On a narrow interpretation of the notion, it may indeed seem to be unfairly liberal: the term 'white ignorance', coined by Charles Mill (1997), refers to epistemic agents *actively resisting* the acquisition of a piece of knowledge. Of course, Mills didn't intend the use of this notion to apply 'in the vacuum' —what he had in mind was a very particular kind of resistance: the resistance of dominant groups to acknowledge facts concerning the hegemonic structure of their social reality and their position within it. Besides, Mills thought of white ignorance as extending to those forms of 'latent', or 'implicit' resistance, to include subjects whose ignorance is actively, albeit involuntarily, cultivated. Ultimately, however, Mills conceived of white ignorance in *agential terms*, as consisting in a dysfunction of one's cognitive system. Not that he neglected the (important) role played by more structural or systemic considerations (quite the opposite in fact[2]), but he nonetheless imputed the failure involved in cases of white ignorance *to the agent*.

Similar considerations also apply to the account of white ignorance proposed by Pohlhaus (2012) and Medina (2013), who expanded on Mill's to include, other than corrupt intellectual habits, also ignorance stemming from 'failures of the will'[3]. According to Medina, white ignorance is a form of *willful ignorance* that is ultimately down to agents' (heinous) desires —say, the desire to resist to acknowledge facts that would jeopardise one's image of oneself (as, say, non-racist).

The accounts of white ignorance proposed by Pohlhaus and Medina on the one hand, and Mills on the other, despite their differences, underwrite a similar view of white

[2] As Martín points out (2021, 872), Mills sometimes also refers to white ignorance as a 'structural group-based miscognition' (2007, 13). Similar observations are also put forward in Alcoff (2007).
[3] Variations of this view are explicitly defended or suggested, other than by Medina, also in Bailey (2017), Spelman (2007) and Woomer (2019).

ignorance as an agent-centred phenomenon —namely, concerning the way agents think or feel about the world and their place in it. This is an important dimension of the wrong involved in these cases. Oftentimes, it is the actions of biassed, ill-willed individuals that cause harm, or create harmful epistemic gaps. These agential considerations, however, do not apply to the DOCTOR case. The doctor's ignorance is not up to them —neither at the level of their will nor (we can suppose) of their cognition. Should we then infer that this is not a case of white ignorance, and that there is nothing morally problematic about it?

In her (2021) paper, Annette Martín argues that both cognitive and wilful ignorance accounts should be seen as concerning two different manifestations of a broader, and more structural, phenomenon that the notion of white ignorance refers to. Martín proposes to understand this underlying structural phenomenon in functional terms, where white ignorance is seen at one time as a *consequence* and a (non-contingent) *enabling factor* of systems of racial oppression. In her words:

> [W]hite ignorance refers to ignorance that [...] systematically arises as part of some social structural process(es) that systematically gives rise to racial injustice. [Moreover,] the ignorance that arises through these processes is not an incidental by-product of these processes, but is rather an active player in them. That is, it must be, at least in part, through their systematic epistemic effects that these social structural processes systematically contribute to and help sustain white racial domination. (Martín 2021, 875)

Sometimes, these social structural processes are rightly understood as taking the shape of the cognitive and wilful barriers erected by individual agents against racial equality and justice. Sometimes, they manifest in the social, architectonical, technological, political, educational, cultural, financial or, more generally *morphological* barriers imposed at the level of the ignorant agent's environment itself. To help us think of the different ways in which white ignorance sets up defence mechanisms to protect itself against evidence and knowledge, Martín offers an illuminating analogy of a stronghold under siege:

> On one level, individual inhabitants can wield personal weapons to defend themselves in close combat. On another level, there are soldiers that enact coordinated manoeuvres that help prevent situations in which individual inhabitants need to draw their swords. On yet another level, the inhabitants are protected by key structural features of the castle, such as the moat, the drawbridge, and the thick castle walls; even the geography of the land plays a role in defending the castle. [...] [T]here are multiple kinds of defence mechanisms for active ignorance, some of which involve action on behalf of the white ignorant agent, and *others which act upstream of the individual*. (Martín 2021, 879) (my emphasis).

This analogy brings structural mechanisms of white ignorance into sharper focus by distinguishing them from agential ones. On the first two levels, the activeness of white ignorance comes from the defences put up by agents (individual or social) who, wilfully or owing to their (more or less culpable) cognitive shortcomings, fail to take up the relevant evidence, form the relevant belief or simply do the right thing. On this level, the activeness

of ignorance is due to *agential resistance*. On the uppermost level, on the other hand, resistance to evidence results from the way in which an epistemic community is organised, the kind of evidence that is (made) available, its quality and the patterns of its distribution, as well as the quality and distribution of the tools and resources that are employed to obtain it. On this level, according to Martín, the badness of white ignorance is independent of agential efforts, but takes place at the level of structures. Ignorance, like shadows, is cast by the 'morphological' features of the environment onto those who live within it.

If this is right, Martín's account can help us understand something important about the root of the injustice at play in DOCTOR —that is, the fact that the doctor's ignorance is part of a broader system of oppression that has generated their ignorance and that profits from it. On the one hand, the doctor's ignorance is the result of a social system that doesn't have the interests of its minorities in mind (when, for instance, it comes to taking decisions about the scope and objectives of medical research); on the other hand, because (among other things) it causes doctors to systematically offer worse medical treatment to non-Whites, it non-contingently contributes to sustaining systems of racial oppression.

Thinking of the doctor's ignorance as a case of *sanctioned*, *white* ignorance helps us bring out two important aspects implicit in the intuition that 'the doctor should have known'. First, the identification of the wrong caused by the doctor's ignorance. This is the sense in which, in saying that the doctor should have known, we are calling out the perpetration of an injustice. Second, the identification of the kind of failure that led to that injustice —that is, a failure that does not concern the doctor themself, but rather the system of which the doctor is part. That's why we think that the doctor should have known that *p* even if, by that, we don't mean that it is *the doctor* who has done something wrong. What we mean is rather that there are ways in which medical research *ought* to be organised, and that one way in which it surely ought *not* to be organised is so that it furthers racial oppression.

Normative judgements of this sort —i.e., centred not on the agents, but on the broader structures that 'set the stage' for individual action— are commonplace. When we say that a playground ought to be safe, for instance, or that rents ought to be affordable, we are employing some such judgements[4]. What we are doing, that is, is assessing a particular environment with respect to a norm (a good or an ideal) it should conform to. If a kid falls and gets injured, or if a single working mother finds herself vulnerable to homelessness, it is also norms of this sort we sometimes appeal to when we say that *this shouldn't happen*. Similarly, it is norms of this sort we refer to in DOCTOR, when we say that the doctor *should have known* —that is, we are judging the way in which the medical environment ought (or perhaps ought *not*) to be organised. A medical system offering medical care that

[4] Naturally, there is broad continuity between so-called 'environmental' and 'agential' norms —often, environmental oughts will imply that someone acts in a particular way to bring about the desired state of affairs, in the same way in which agential oughts will effect a change in the distribution and arrangement of resources in the environment. But this isn't enough to think that there isn't a distinction to be made between them. At a minimum, in fact, so long as it makes sense to think that there is a regulative norm (a 'good', or an 'ideal') that a particular structure (a practice, an environment) should respect, which is not reducible to the norms of any one (individual or collective) agent, then speaking of environmental 'oughts' should not be problematic —even if only in a thin normative sense. A similar point between the relationship between environmental and agential norms is also suggested in Chrisman (2008) and (2022), in terms of the relationship between state norms (roughly, what I call environmental norms) and norms of action (what I call agential norms), and based on an earlier distinction made by Sellars (1969) between 'rules of action' and 'rules of criticism'.

systematically discriminates against some group of people (e.g., by sanctioning specific information lacunae amongst the medical practitioners) is not a *good* medical system.

Following Rawls, I have proposed to call these normative judgements *structural*, to distinguish them from *transactional* normative judgements. Structural and transactional judgements are at the centre of two parallel frameworks for normative assessment —a broadly Rawlsian inflationary framework (from Rawls 1971) and a Nozickian deflationary one (Nozick 1974). The main difference between these frameworks can be understood in terms of what they take to be at their centre —whether it is individuals and their actions, according to the Nozickian picture, or the fundamental macrostructure of society as a whole, as per the Rawlsian one.

These two frameworks serve different, often non-overlapping purposes. Deflationary judgements about individual transactions, for instance, aim to measure the harms, responsibility and virtues of individual actors in their interactions with the world and other agents. For instance, in her (2012), Elizabeth Anderson offers an interpretation of Miranda Fricker's celebrated notion of epistemic injustice as an instance of a normative epistemic framework that broadly follows this schema. In identifying the wrong suffered in cases of testimonial and hermeneutical injustice at the level of the individual's inability to make themselves heard or to understand their own experience, for instance, Anderson notes how Fricker's attention focuses on the microscopic dynamics between individual agents —i.e., the credibility that a hearer owed to their interlocutor, or the hermeneutical damage suffered by someone who is deprived of fundamental interpretative tropes.

Despite its merits, this deflationary Nozickian picture has been widely criticised for its inability to keep track of the background conditions that are often required to offer large scope analyses of more complex situations. One criticism in particular takes issue with the assessment of the cumulative effects of countless individual transactions on the global state of a system. The idea is that the way in which the actions of a multitude of individual (and collective) agents connect and interact with each other can give rise to unjust outcomes even if no injustice is present in any individual transaction (Anderson 2012, Young 2011).

In order to be able to offer an assessment of the global state of a system, it is necessary to adopt a macroscopic perspective, fixed on the background conditions of its fundamental structure. This is the kind of normative evaluation that is made available by the Rawlsian inflationary picture. Following Rawls (1971, 1993), we can take structural normative judgements to concern precisely the ordering of a complex system according to the principle of *justice*. In the same way in which moral goodness is the ordering principle of individual transactions, Rawls thinks of justice as 'the first virtue' of *political institutions* —i.e., a regulative ideal that offers guiding principles to the determination of rights and duties, and that regulates the distribution of resources amongst the participants in a society.

Epistemology has traditionally been dominated by a focus on the kind of transactional normative judgements typical of a deflationary Nozickian image of normativity. Knowledge, after all, has traditionally been thought of first and foremost as a (more or less private) relationship between a Subject and *his* Object. Even the more recent attention to the sociality of knowledge (from testimony to epistemic injustice), which has brought attention to the role of other people in shaping our epistemic lives, has not challenged the centrality of the individual agent(s) in this picture.

It is only when we come to theorise about cases like DOCTOR, however, that the explanatory limits of this framework become evident. If all we had were 'local' criteria of normative assessment (i.e., pertaining to individual transactions), we would *have to* infer, from the (assumed) rule-abiding behaviour of the individual agents involved, the *necessary goodness* of their combined effect —that is, the necessary goodness of the racial discrimination of Indigenous Latino people by the medical system.

Rawlsian normative structural judgements, on the other hand, allow us to "bring under normative evaluation the aggregate consequences of a combination of many individual actions" (Young 2011, p.67), and so to understand and assess 'trickle down' cases of political wrongs, like in DOCTOR, where the injustice perpetrated is not attributable to the individuals involved as such. It is only within an inflationary Rawlsian framework, then, that we can identify the distinctively political dimension of the wrong perpetrated in this case, and recognise, in the way in which the information economy is structured, its systemic roots.

II.2 The Epistemic Side, or: Leaving the Psychological Image Behind

The information economy in the doctor's medical environment is organised in such a way that it tends to make some group of people in that community *physically* worse off. This is *morally* (or, perhaps more exactly, *politically*) bad. At least on the face of it, however, to the extent that it does so by sanctioning a specific gap in the environment's *epistemic* resources, it seems plausible to think that the information economy is spoiled also in a distinctively *epistemic* sense —after all, it is an epistemic resource that the doctor is missing out on, and it is the lack of availability of this piece of evidence which sanctions the doctor's ignorance. Of course, I don't mean to say that this is *trivially* so. At the very least, something would need to be said about the general shape of the normative epistemic landscape in which a lacuna of this sort would counts as an *epistemic* problem —and possibly, on pain of making epistemic failure ubiquitous, in a way that doesn't make of every case of ignorance a foul. This being said, though, there still seems to be a *prima facie* case to be made for the epistemic analogue of the moral and political wrong.

Or is there? For it is precisely this initial plausibility that seems to be incompatible with a common image of epistemic normativity. According to this picture, what matters epistemically ultimately concerns the ways in which one responds to evidence that is *psychologically available* to one[5]. Call this the *Psychological Image* of epistemic normativity (or simply PI). According to the PI, actively ignoring evidence that we have at our disposal when it is relevant for a belief we are forming is among the things that plausibly we *ought not* to do[6]. More generally, the idea is that when one has psychological access to a *defeater* for *p* (roughly, a reason against believing it), one (epistemically) ought *not* to believe *p*[7] —whence the daunting Cliffordian adage that "it is wrong always, everywhere, and for anyone to believe anything on insufficient evidence" (1877 [1886], 5).

[5] Most importantly, Conee and Feldman (1985), Pollock (1986) although see also Moss (2015) about a possible reading of the root cause of this stance in contemporary epistemology. Externalist (reliabilist) variations of this view have been defended also by (Goldman 1979), Beddor (2015) Graham and Lyons (2021) and others.

[6] Although see Siegel (2012) for a criticism based on considerations of cognitive penetrability, and also Kelp (2023) for an argument from defeater defeaters.

[7] This is how Lackey (2021), for instance, puts it. For a more complete discussion of epistemic defeat, see Kelp (2023), Simion (forthcoming).

The problem though is that this psychological image seems to lead to very implausible results. A crucial worry that is often raised, for instance, concerns its extensional adequacy. Consider these cases:

> SEXIST SCIENTIST: Hem regularly dismisses testimony from expert women scientists because he thinks that women are not good at science. Indeed, he is so entrenched in his conviction that when women talk about scientific matters he just zones out. One day his colleague Jem, a renowned expert on the relevant topic, tells Hem that the experiment he and his colleagues ran to prove that *p* was seriously flawed, and she presents him conclusive evidence to this effect. As per usual, Hem doesn't pay any attention to what Jem says, and continues believing that *p*.[8]

> RACIST HERMIT Len carefully curates his information intake so as to preserve and nourish his racist belief that *p*. He does what he can to suppress the doubts that from time to time emerge about his convictions, avoids investigation into questions that could shake them, makes sure to spend time with like-minded people, reads news outlets that confirm his beliefs and cherry picks information that reinforces them. His convictions, no matter how sincere, are not earned by honest and patient inquiry, but 'stolen by listening to the voice of prejudice and passion' (Clifford 1886 [1877], 3).

In both these case, the problem is clearly that, although the evidence is not psychologically available to them, Hem's sexist belief and Len's racist belief that *p* are not justified, and it seems right to think that they should abandon them on the basis that *there is* evidence showing that they're wrong[9]. To anticipate this and other worries[10], a growing number of epistemologists now recognises that normative epistemic pressure often comes not only from evidence that is available to one psychologically, but also from evidence *external* to the agent's ken[11]. Since the sexist scientist and the racist hermit's ignorance concerns a piece of information that is not psychologically available to them, the epistemic 'ought' that they are subject to, we can say, imposes an obligation that comes from something external to the agent's psychology. Although the two cases are otherwise substantively different, something analogous seems to be going on also in the DOCTOR case —i.e., in this case too it seems wrong to trace the origin of the obligation to anything *internal* to the agent.

If that's the case, the existence of normative epistemic judgements of a structural variety should not be thought to be implausible, or so one may think. It is true, however, that it remains to be clarified what exactly the external facts that exercise this normative epistemic pressure are, why they exert it, and on whom.

---

[8] Case readapted from Simion (forthcoming).

[9] Proponents of the psychological image could agree with this verdict on the grounds that both Hem and Len have higher-order defeaters —that is, they could argue that they have evidence that there is evidence that should have been gathered. Although sometimes this will surely be the case, however, it seems easy to come up with situations where one does not have a clue as to the evidence they are missing out on. Arguably, this is the case with Hem, since he literally zones out whenever a woman talks about science. And this could also be the case with Len, if we imagine a point in which his life choices have brought him to a level of isolation that prevents him from coming to know that evidence contrary to his conviction even exists.

[10] See Simion (forthcoming) for an argument against the sufficiency condition of psychological accounts.

[11] Most notably Lackey (2014, 2016), but also see Kelp (2022, 2023) and Simion (forthcoming) for a useful summary of the debate surrounding defeat.

These are not easy questions to answer. A natural source of doubt, for instance, seems to come from the thought that there are simply too many things out there for one to be sensibly required to be responsive to them all. To make the case more pressing, consider the following trivial fact: at any given time, there is an unlimited amount of evidence we do not possess, as opposed to the restricted set of evidence we do possess. For instance, there is evidence we do not have access to at time t but could access at a later time t_n (like, for instance, concerning the state of affairs in the other room, or at my friend's house). Then, there is evidence we could never have access to (say, for psychological limitations, or for limits of time). And there is evidence that we could access, but that would require some digging to render available. These are only some of the demands that one would be required to satisfy, and they already go way beyond what it seems reasonable to require (rationally or otherwise) of one. If so, though, how can we make sense of epistemic obligations grounded on facts *external* to the individual's psychology?

This is an important challenge. Far from being a fatal objection, however, I take this to offer a stimulating point of departure for theorising about a new image of epistemic normativity, alternative to that offered by the PI. In the next section, I look at two main attempts that have been made in this direction by what I call Individualist and Socialist accounts of epistemic normativity. In a nutshell, I take Individualist Accounts to centre epistemic normativity on facts about the individual agent *beyond its psychology*, and Socialist Accounts on facts about the *individual's position in the social matrix*. Both accounts, I show, offer plausible strategies for carving the normative environment in a way that clearly identifies which external facts matter to us epistemically. On the other hand, because they are incapable of accounting for the DOCTOR case, I find that the image of epistemic normativity they commit to, vulnerable as it is to the same problem of extensional accuracy levelled against the psychological image, is still incomplete.

§II. THE AGENTIAL IMAGE

II.1 Individualist Accounts

According to Simion (2023), a subject has an epistemic obligation to form a belief that p if sufficient and undefeated evidence supporting p is easily available to the subject. The obligation is grounded on two main assumptions: that epistemic agents function properly when they generate knowledge (call this the *Proper Functionalism Commitment*, or PFC), and that evidence indicates knowledge (call it the *Evidence as Knowledge Indicator Thesis*, or EKIT). PFC and EKIT make it a condition of our functioning properly, as epistemic agents, that we don't fail to pick up *easily available* evidence. Epistemic agents who fail to do that —who resist evidence that is easily available to them— are malfunctioning epistemic agents: they are failing to take up information that is likely to generate the good (knowledge) whose production is constitutive of their proper functioning. In normative terms, then, whenever we have knowledge-indicating evidence around us, in our capacity as creatures whose

epistemic proper functioning depends on generating knowledge, we *ought to* pick it up[12]. Call this the *Functionalist Account*, or FA.

When is evidence easily available to a subject? According to Simion, these availability conditions are taken to track a psychological 'can' for an average cogniser, where the 'can' describes the capacity of the agent's cognitive process, at a time *t*, to pick up information that is *qualitatively* (of the kind that the epistemic agent can process), *quantitatively* (of the amount that they can process) and *environmentally* (within immediate reach of the agent) 'at hand' at that time. Simion calls evidence that satisfies these criteria evidence the agent is in a *position to know*. Although these guidelines are fairly permissive, it is clear that the restriction Simion has in mind is rather severe, since she thinks that if, in order to access a piece of evidence, "I need to open my eyes, or turn around, or go to the other room, or give you a call, then I am [...] not in a position to know it" (Simion 2023, 7).

A different individualist account, one that offers a more relaxed modal profile, could be proposed by taking epistemic norms to consist of *procedural obligations*[13]. Call this the Procedural Account (or PA). Epistemic 'oughts', according to the PA, are understood as obligations to "perform procedures, which could, under favourable circumstances, expand a subject's evidence base or improve the reliability of her conclusions" (Nottelmann 2021, 1187). In a similar spirit to Simion's teleological account, this account also takes that we have epistemic obligations to perform certain actions that can make us epistemically better off (or prevent us from being epistemically worse off). Unlike the FA, however, the PA is more flexible with respect to the kind of actions one can take to achieve their epistemic ends, and their target.

With respect to the kind of action, a non strictly-speaking epistemic activity (like keeping your workspace tidy) for instance, or training your open-mindedness and curiosity, can be taken to be epistemically beneficial for a proponent of the PA —if, for instance, it makes it easier to identify relevant evidence. With respect to the target, on the other hand, since, unlike the FA, this view doesn't tie the source of the obligation to the proper functioning of the individual cognitive system, this view is compatible with the idea that obligations to perform certain procedures may not only concern the (epistemic) benefit of those who perform the actions themselves, but also the benefit of others[14]. In this sense, we can say that the PA is a distinctively altruistic view of positive epistemic 'oughts'[15].

Putting these observations together, then, we get that, according to the PA, one has an epistemic obligation to perform a procedure (φ) if φ-ing requires very little effort and, by φ-ing, one can normally make someone (either themselves or others) epistemically better

[12] A neighbouring idea, not grounded on proper functioning, is proposed by Ichikawa (2022). According to Ichikawa, you ought to believe p if, roughly, (a) you are in a position to know that p and (b) you are considering the question whether p. Because Ichikawa doesn't offer an account of the normative pressure determining (b), I don't take his account to offer a solution to the specific problem I am interested in here.

[13] The view sketched here is an individualistic interpretation of a position suggested by Lackey in her (2014, 2016 and 2021), and framed in similar terms by Nottleman (2021).

[14] The suggestion that intrapersonal duties should be accompanied also by interpersonal ones is explicitly made in Lackey (2021), on the grounds of what she calls the *parity thesis*, according to which "if it is an epistemic duty to promote an aim in myself, then a duty that is identical except that it regards others is also epistemic." (2021, 285).

[15] I suspect that there is a variation of the FA that would be compatible with the idea that sometimes we have duties to the epistemic betterment of others. However, since these duties would be ultimately grounded on the proper functioning of the epistemic agent, it wouldn't be properly an *altruist* view.

off, or prevent one from being epistemically worse off[16]. How much effort is 'very little effort'? Since it is motivated by over-demandingness worries[17], this modal constraint should be understood as designed to only ward off the possibility that one is required to go at great length to satisfy one's own and other people's epistemic needs all the time, and so it should then be taken as fairly relaxed —especially if compared to the FA's modal constraint. Indeed, we can expect that most of the time, actions like opening one's eyes and going to another room, as well as sorting through the letters on your desk or the emails in you junk folder[18], will be, according to the PA, just what you ought to do in order to epistemically improve yourself and others[19].

Individualist accounts can give a unitary explanation of a wide variety of cases of epistemic wrongdoings. They can explain what's wrong in standard cases of testimonial injustice, for example, of perceptual non-responsiveness, wishful thinking and more (Simion 2023). But what about cases of white ignorance? Consider a slight variation of the DOCTOR case. Like in the original case, dr. D ignores the effects of drug X on their patient P. In this variant, however, (call it DOCTOR*) Genome Wide Association Studies (=GWAS) have been very recently conducted also on non-White European populations. So here the evidence *is* available, although not that easily —say, it can only be found in specialistic journals. Like in the original case, dr. D's ignorance is problematic: their lack of knowledge consists in a failure to obtain a relevant piece of knowledge. The verdict, also, stays the same: dr. D *should have known* the effects of drug X on their patient.

Do these views give us the right verdict in DOCTOR*? In its current formulation[20], it is not clear that the FA can give a satisfactory explanation of this case. Information available in a specialistic journal, although it respects the qualitative and quantitative dimension of FA's modal constraint, is *not* immediately *environmentally* 'at hand' in the relevant sense. It would surely require a much greater effort than, say, opening one's eyes, turning around, or going to the other room —conditions that the FA identifies as already beyond the scope of what one ought to do epistemically. What about the procedural view then? Initially it may seem that the PA stands better chances here. Consider the case of Mary, who has misleading evidence indicating that her friend Norman is in Rome. Strong conclusive evidence that in fact Norman is in San Francisco, and not in Rome, is contained in a letter on Mary's desk, stuck in a pile but easily within her reach. Had Mary opened that letter, she

[16] This definition integrates the 'procedural account' that, following Rosen (2004), Nottleman (2021) attributes to Lackey (2014, 2016), and the following definition, due to Lackey, of interpersonal epistemic duties: "If it is in our power to prevent something epistemically bad from happening through very little effort on our part, we ought, epistemically, to do it." (2021, 287), which applies to the epistemological domain the notorious ethical principle proposed by Singer (1972, 231).

[17] In fact, the modal aspect of Lackey's definition is taken from Singer's master argument, which was proposed in order to avoid over-demandingness worries. Lackey herself acknowledges this explicitly, quoting Singer: "This is to satisfy the over-demandingness constraint—just as there is said to be a 'limit to how great a sacrifice morality … can legitimately demand of agents'"

[18] And, perhaps more problematically, also giving someone a sandwich, cognitively enhancing drugs, perhaps even reading them the Critique of Pure Reasons (no matter whether they want to listen or not).

[19] Obligations of this sort are pretty easy to come by. Naturally, then, we should take them to be fairly lightweight. Other kinds of obligations (e.g., moral, prudential) may often override them, and make it all-things-considered permissible to fail to comply with them.

[20] The suggestion I offer at the end of this paragraph regarding the procedural account also applies to the FA. Notice however that, since both views offer only sufficient conditions for epistemic duties, they are only shown to be incomplete by the case discussed.

would have known. Similarly, one may think, had dr. D read that article, they would have known.

What's not clear, though, is whether this view captures the sense in which the obligation to read the article in a specialised journal, although it pertains to the doctor, doesn't extend to other agents as well. For notice that, while opening a letter on our desk is an action that anyone can be reasonably expected to have an obligation to perform (provided it is sufficiently easy for them to do so), the same is not true of reading a specialised scientific journal. A view requiring that one stayed up-to-speed with all medical knowledge produced would indeed be over-demanding. Yet, it is intuitive that a doctor has an obligation of this kind. Perhaps it would be possible to narrow down this view by including conditions tracking how, say, social facts about the individual agents (i.e., their being a doctor) influence the kind of actions one is under the epistemic obligation to perform[21]. Accepting this, however, would be tantamount to accepting that facts beyond the individual play an essential role in shaping the domain of the things that matter to us epistemically.

In the next section, I move on to consider views that take this intuition as their starting point. More in particular, I consider two ways in which the accounts discussed have been narrowed down to fit this consideration: one suggested by Sandy Goldberg (2017, 2018), and the other by Simion and Kelp (Simion and Kelp forthcoming, Kelp 2022 and 2023).

II.2 Socialist Accounts

Let's start from this seemingly obvious consideration: it is part of a doctor's obligations to keep themselves informed on advances in their field. Reading the latest issues of medicine journals, for instance, is one of the things that a doctor should do. It is as clear that this is an obligation for a medical doctor as it is that this is *not* an obligation for, say, a doctor in philosophy. The difference in epistemic standards we set for medical doctors and philosophers is hardly surprising —it is because one is *a medical doctor* (rather than, say, an electrician) that they have specific obligations —or so the thought goes[22]. But what is it exactly about the fact that one is a doctor that gives rise to these obligations? One way we could go about this is by thinking of the doctor's role as a place in a web of social relations[23]. But how do these social relations give rise to the obligations?

One influential way to answer this question is offered by Sandford Golberg (2017, 2018). For Goldberg, given a particular type of knowledge-generating process, a belief is justified (epistemically 'proper') *vis-a-vis* that process if it respects both evidence one has and that one should have had. The latter, normative, evidence, thinks Goldberg, is evidence one is *expected to have*. What Goldberg has in mind here is a particular kind of normative expectation (of the form 'you should X!') members of a particular community are entitled to have of one based on the social position they occupy in that community. So to the extent

[21] Lackey suggests something along these lines in her (2014). I consider one such expansion of this view in the next section, when I look at the ways in which Goldberg, Simion and Kelp could be interpreted to have elaborated Lackey's suggestion.

[22] I take this idea to have been first explicitly brought up by Lackey (2016) in her discussion of the (epistemic) duties befalling on someone in virtue of their group membership. The views I discuss in this section can be taken as offering different formalisations of Lackey's central intuition.

[23] I take this idea to have been first explicitly brought up by Lackey (2016) in her discussion of the (epistemic) duties befalling on someone in virtue of their group membership. The views I discuss in this section can be taken as offering different formalisations of Lackey's central intuition.

that others are entitled to expect that one φs, then this person will have an obligation to φ. Call this the *Social Expectations Account* (or SEA).

For Goldberg, we are entitled to have these normative expectations because participating and engaging in social practices is ultimately *practically* good for us. But social obligation theories need not be committed to reducing epistemic 'oughts' to socio-practical considerations. In a series of recent papers, for instance, Chris Kelp and Mona Simion (Kelp 2022 and 2023, and Kelp & Simion manuscript, Simion forthcoming) defend a fully epistemic view that, like the SEA, takes social position to play a key role in determining the set of evidence one ought to be responsive to. Instead of cashing it in terms of the *social expectations* that social roles give rise to, however, they appeal to their *function*. On this view, we are asked to think of social positions in terms of the function they play in the system of which they are part: being a doctor is to occupy the function of treating ill people in a particular community, as well as being an electrician is to occupy the function of 'treating' issues concerning electrical systems.

If this is so, we should be able to extract epistemic obligations from social positions in the same way in which Simion's FA proposed to do with individual agents[24]. Plausibly, and very roughly, in order to treat ill people, the doctor will have to satisfy some desiderata that are essentially *epistemic* —such as, for instance, reading medical books and, plausibly, keeping up to date with the medical literature— in such a way that satisfying them is constitutive of the role itself: ideally, a doctor would not be a good doctor if they failed to keep up with advancements in the field, since this would lead them to consistently fail to do their job properly. Importantly, then, on this view, since fulfilment of social roles mandates compliance to genuine epistemic norms, although they are grounded on social facts, unlike in the SEA, they don't reduce to social ones.

It should be clear by now that both the SEA and the SFA can easily account for DOCTOR* cases. Recall this variation of our case: dr. D doesn't know whether p (= drug X is safely metabolised by their patient P), despite the fact that a study showing that p is the case has recently been published in a prestigious medical journal. SEA gives us the right verdict here: since we are entitled to expect of a doctor that they keep up to speed with the relevant literature, the doctor *should have known* that p was the case. The SFA also gives the same verdict, although it doesn't rely on what we can reasonably expect of a doctor, but on the epistemic standards that the doctor is naturally subject to in virtue of the function they occupy in a society —that of treating ill people.

What about the original case though? The difference between DOCTOR and DOCTOR* is that in the former, unlike the latter, evidence regarding the effects of drug X on non-White European patients are not available to dr. D. Indeed, the relevant research hasn't been conducted yet —GWAS, Genome Wide Association Studies, have only been conducted on White European populations. Can either SEA or SFA give us the right verdict for this case of white ignorance? I think the answer here must be negative. Start with SEA. According to SEA, you should have known X if it is reasonable to expect that you knew X given the social position you occupy. Clearly, however, it would *not* be reasonable *at all* to expect of dr. D that they knew whether p, given that no study to establish p has yet been conducted. Indeed, it would obviously be over-demanding to expect *anyone* to be sensitive to *all* the

[24] In fact, this view can be seen as an integration to Simion's individualist view considered earlier.

evidence relevant to their social position, especially in cases, like DOCTOR, where such evidence has yet to be discovered, and where it is clear that there is nothing the subject could do that would put them in a position to obtain that knowledge. And obviously, since it would be equally implausible to take this to be an epistemic requirement for the proper fulfilment of any social position whatsoever, the same applies also to SFA. The proper functioning of the doctor-role in a community, although it may very well depend on whether their knowledge is up to date with the relevant literature, doesn't have anything to do with whether they themselves have conducted this or that relevant study on their own.

Perhaps their verdict is ultimately right —that is, they may be right to think that there is no epistemic failure that is imputable to the *doctor*. Even so, however, this is still compatible with recognising a sense in which the doctor's ignorance poses a distinctive *epistemic* worry. In what follows, I suggest a way in which proponents of these views could modify their account in order to accommodate cases of sanctioned white ignorance like DOCTOR.

§III. THE ENVIRONMENTAL IMAGE

A traditional image of the source of the normativity of the epistemic understands obligations as deriving from facts about the evidence that is available to us psychologically. For this reason, I proposed to call this the *psychological image* of epistemic normativity. The psychological image vindicates an extremely intuitive idea: that there is no way that a subject can epistemically be in the right by actively ignoring evidence they possess and that is relevant for their beliefs. That this is *all* that matters for one to be epistemically in the right, however, seems much less obviously true. This is the point I take to have been strongly suggested by both individualist and socialist views of epistemic normativity —i.e., the point that one can fail epistemically not only by failing to be appropriately responsive to the available evidence, but also by failing to respond adequately to nearby evidence, or by failing to stand up to the standards imposed by one's social role. Somewhat ironically, a gesture towards the shortcoming of the psychological image was first made by William Clifford himself when, with customary gravity, he writes that:

> If a man, holding a belief which he was taught in childhood or persuaded of afterwards, keeps down and pushes away any doubts which arise about it in his mind, purposely avoids the reading of books and the company of men that call into question or discuss it, and regards as impious those questions which cannot easily be asked without disturbing it—*the life of that man is one long sin against mankind*. (1877, 5). (my emphasis).

To the extent that it recognises a natural sense in which we are, as epistemic agents, under normative pressure generated not only by evidence we *possesses*, but also the one we actively *ignore*, this thought proposes a different image of epistemic normativity —one whereby it is also the knowledge we do *not* have, the evidence we have *not* gathered, or the things we have *not* done —i.e., things that are external to the subject's psychology— that sometimes matters to us *epistemically*. Call this the *agential image* of epistemic normativity. At the centre of the agential image it is not just a subject with their beliefs, but a (socially) situated epistemic agent whose obligations stem not only from the evidence they possess, but also

from the evidence scattered around them, or possessed by other people, from the expectations they have of them or from the role they occupy in their community.

Because it proposes a significant expansion of the domain of what matters to us epistemically, this image has much greater explanatory power than the psychological one[25]. For instance, I have argued that it can account for cases like DOCTOR*, in that it allows to explain the doctor's ignorance of a fact that is available to them in terms of their failure to comply with an obligation to be responsive to or enquire into matters that are relevant to them given their social role. And it can account for at least some paradigmatic cases of testimonial injustice, where the wrong is cashed out in terms of one's opposing resistance to easily available and reliable evidence (e.g., one's testimony) that they ought to have picked up.

But what about cases where it is not the individual themself who "keeps down and pushes away" evidence, but it is instead the social and institutional barriers that make it unavailable? Recall Martín's siege analogy and the morphological level at which knowledge barriers sometimes arise —think for instance of our DOCTOR case, where dr. D's ignorance is not imputable to anything to do with the cognitive abilities of the doctor or their desires. In fact, it is easy to imagine circumstances where ignorance is encouraged by (more or less) tacit features of the environment that prevent the cultivation and dissemination of knowledge.

Think for instance of a philosophical tradition under a strict regime imposing harsh censure on ideas, publications and events that don't respect the spirit of what they call their 'synthetic philosophy'. We can imagine that, here and there, some members of this community will come into contact with philosophical work from outside their community, or have conversations with philosophers from different traditions. Still, and perhaps owing to their indoctrination, some will not have acquired the conceptual tools necessary to even understand those foreign voices. Among those who could understand, many will just disregard these alien thoughts and seemingly unbelievable ideas. In addition to these agential forms of resistance, however, we should expect stronger and more capillary obstacles coming from the 'synthetic regime' itself, and which includes the barriers that the regime imposes over the circulation of heterodox work in the first place.

In fact, to the extent that it is responsible to set the norm of what counts as proper philosophical thinking, it is clear that the regime plays a central role in the resistance to the epistemic integration of philosophical traditions. Sometimes, these norms will be imposed through official sanction (e.g., by not allowing publications of non-synthetic work in some of their journals, by not teaching the work of authors from different traditions of thought, by discouraging students to write about them and ridiculing those who mention them, by discouraging international travel, or through restrictions on funding allocations); other times they will be the result of less official measures, taking shape in the attitudes and behaviours that normalise ignorance and encourage forms of resistance.

The ignorance in which the doctor and the synthetic philosophers live is not always, nor most relevantly, I believe, wilful. For the most part, their ignorance is rather projected onto them by the barriers erected by the systems within which they exercise their profession. In

[25] This view is receiving growing support in recent years. In addition to the views reviewed in the section above, see for instance Srinivasan (2015), Flores & Woodard (2023) and Huges (2023).

other words, it is a consequence of the corrupt organisation of a system of practices[26]. For this reason, with respect to cases of this sort, the agential image has simply no purchase. And crucially, it doesn't by necessity. For the problem here has nothing to do with the agent (or at least not primarily). The agent's ignorance —we can assume— may not be up to them. Yet, the fact that it is not up to them doesn't make their ignorance any less problematic. In some cases it is *morally* problematic, for sure, when for instance lack of knowledge about how to treat a group of patients leads doctors to systematically offer worse medical treatment to them, like in our case. *Epistemically* too, possibly, insofar as their ignorance is seen as reflecting a structural imbalance in the distribution and availability of evidence in the environment.

But in what sense does the distribution and availability of evidence pose a distinctively *epistemic* problem? Clearly, these factors make no epistemic difference according to the normative requirement imposed by either the psychological or the agential image. The psychological image takes only psychological factors to be relevant to epistemic normativity, and the organisation of evidence in one's epistemic environment has nothing to do with one's psychology. For proponents of the agential image, on the other hand, these facts have some epistemic relevance, but only insofar as they concern their relation with the epistemic agent. The fact that some evidence is thus-and-so arranged is only relevant to proponents of the agential image to the extent that it makes certain obligations salient for an epistemic agent thus-and-so placed with respect to it. For example, on the assumption that nearby evidence is evidence one normally ought to pick up, the fact that there is a table in front of me makes my obligation to form the relevant belief salient.

But this says nothing about the fact that, say, it is evidence about the *chair*, and not the *table*, that ought to be available to me. That is, nothing about the distribution of evidence *per se* matters epistemically according to the agential image. The fact that it is *this* piece of evidence that I ought to pick up now and not *that* one, is of no normative epistemic significance. However, it is precisely this kind of normative evaluation that structural cases of white ignorance seem to elicit —i.e., the fact that there is evidence that is not available to one but that should be available to them. Now, how can we make sense of a normative requirement of this sort? My suggestion is that we endorse another image of epistemic normativity. Since it has to make sense of normative facts concerning the organisation of epistemic goods in a particular environment, the centre of this new image will be occupied not only by agents and their beliefs, but by their epistemic environment itself. For this reason, I call this the *Environmental Image* of epistemic normativity. The rest of this paper will be dedicated to providing an outline of this image by clarifying the notion of epistemic environment and the source and structure of a normative domain that takes epistemic environments as its centre.

III.1 Epistemic Environments and The Environmental Image

I take an epistemic environment to be the *network* of epistemic *interactions* that obtains in a particular *community*. By 'community' I mean a collection of epistemic agents (individuals and groups), resources (e.g., material, like artefacts and technologies, as well as 'schemas',

[26] What I have in mind here is very similar to what Haslanger refers to as *cultural techne*. In fact, this example, not unlike DOCTOR, can be seen as consisting in a case of ideology —i.e., a case where a community's social practices are fundamentally corrupt.

like cultural and social norms, concepts, customs, interpretative tropes and so on[27]) as well as the physical surroundings (both natural and artificial). Together, agents and resources occupy the positions, or *nodes*, of the epistemic network[28]. The epistemic exchanges between nodes and the norms and rules that regulate such exchanges, finally, together constitute what I refer to as *interactions*[29]. For instance, perception and testimony would count as paradigmatic forms of interactions, one involving an agent obtaining an epistemic resource (say, belief or knowledge) from their (physical) surroundings, and the other involving agents exchanging an epistemic resource (the testimony). But I take epistemic interactions to include all sorts of exchanges between (or *inter*-)nodes, as well as some *intra*-nodes exchanges[30].

Examples of interactions of the former kind include exchanges between different sorts of agents (i.e., between groups and individuals, for instance, or between groups) between agents and artefacts (i.e., consulting a thermometer would be some such an epistemic exchange, as well as searching on Google or writing on a notebook) as well as between artefacts. By intranodes exchanges I mean interactions within a node —introspection, for instance, or collective remembering, can be considered an epistemic interaction in this sense[31]. In general, then, we can take the notion of epistemic interaction to be designed to keep track of the zetetic and epistemic practices of gathering, fabricating, sorting, acquiring, transmitting and storing epistemic resources. In a slogan, epistemic interactions can be seen as a measure of the dynamism of an epistemic environment. Finally, nodes and interactions together constitute the epistemic *network* —that is, the very fabric of an epistemic environment.

Having clarified the notion of epistemic environment, the question we should ask now is how switching our focus from agents to environments changes our image of epistemic normativity. To begin with, notice that the epistemic normativity of agent-centred images is essentially *doxastic*, in the sense that it typically concerns the evaluative or prescriptive norms a belief (or the believer) must comply with in order to count as knowledge (or some sort of epistemic success). Put differently: knowledge is the result a belief (or the believer) will obtain if it respects the epistemic norms. Since an epistemic environment thus conceived is not itself a knower[32] (or an agent, for that matter), and there is no specific

[27] See Haslanger (2007)

[28] Note that the nodes of the network are not identical to the members of the community. Members of a community who occupy particular epistemic roles act in their capacity as nodes of an epistemic environment. See Haslanger (2016) for an idea of epistemic roles as nodes in a structure.

[29] The notion of interaction I use here is inspired from Longino (2021).

[30] It may be noted that this way of characterising an epistemic environment doesn't set precise boundaries for what counts as an environment. Is the kind of environment I have in mind the loose, general, environment comprising, say, an entire society, or is it the smaller, structured or unstructured group that participates in that society? This distinction seems to become particularly salient when it comes to assessing the quality of an environment —for, say, couldn't a corrupt system be hosting epistemically 'healthy' pockets of resistance? Although I think that a complete account of the nature of an epistemic environment would need to say something to clarify this, I don't believe this constitutes a serious concern for a defender of this view. The reason, quite simply, is that epistemic environments serve here as anchors to ground a new kind of epistemic normative framework —i.e., one that is structural in nature. Once the very general nature of the normative framework is clarified, it is natural to expect its application to particular *loci* to vary from case to case —in exactly the same way in which the traditional individual-based normative framework is applied piecemeal to assess the epistemic status of specific individuals in specific cases.

[31] I take it that whether this will count as inter- or intra-node exchange will depend on the group, and on one's conception of group.

[32] Nelson (1995) offers a notion of epistemic community not too dissimilar to my epistemic environment as a proper subject of epistemic attributions. See also Calvert-Minor (2011) for a criticism of Nelson's proposal that suggests a picture more similar to the one I sketch here.

belief or set of beliefs it can be said to possess, the norms regulating its epistemic functioning, in that their primary aim won't be to guide an agent's belief towards epistemic success, will not be doxastic —or at least not in the same sense.

What is the kind of epistemic success that an environment should aspire to then? My suggestion is that in a similar way in which Rawls proposes to think of justice as 'the first virtue of social institutions', we can think of a sister notion —say, that of *epistemic justice*— as the norm of epistemic environments, where this amounts to promoting the epistemic betterment of the epistemic community as a whole. To wit, I take the promotion of the epistemic betterment of a community to involve at least two things: *(a)* that the environment ought to be knowledge-conducive, and (following Anderson (1999, 2012) and Young (1990)) *(b)* it ought to do so in a way that respects its participants' epistemic self-determination.

Regarding the latter, we can think of an environment as respecting the collective self-determination of its participants when it respects the development and exercise of individual epistemic virtues and capacities, the fulfilment of the epistemic aspects of one's role(s) in the community, and if it promotes collective coordination, communication and cooperation. This may include things like: making good epistemic tools available, equipping the environment with systems for the monitoring of the goodness of information in circulation, removing barriers to its circulation, and so on.

Regarding the former, there are a number of factors that are relevant to the knowledge-conduciveness (=K-conduciveness) of an environment. For instance, the availability of a particular resource in a particular environment could be measured by determining the frequency with which it appears and how well it is distributed. In this way we can evaluate the epistemic justice of an environment with respect to a particular metric (the resource at hand —say, awareness of gender inequality) by looking at whether the environment has developed certain concepts and interpretative tropes and how frequently they appear, and (since K-conduciveness measures how conducive to knowledge interactions are for the members of a community) also depending on how well distributed among community members these concepts are.

At a minimum, then, we can say that an environment is K-conducive when its interactions meet certain qualitative and quantitative standards. Along the qualitative dimension, a good environment is one where interactions involve the exchange of *positive* and *varied* epistemic resources in a way that respects the norms for the particular kind of exchange. For instance, an environment where lying is widespread will be qualitatively worse than one where it isn't, at least on the ground that there are more true beliefs/knowledge circulating in the latter than in the former. The relative amount of true beliefs should be measured also with respect to their variety, as an environment with more resources will not necessarily be qualitatively better than one where there is a much greater variety of them. In other words, an environment may involve the exchange of positive epistemic resources and still be of poor quality if, say, the resource exchanged is only one —i.e., the true fact that, say, today is Tuesday.

For the same reason I believe that the *frequency* and *distribution* of epistemic resources should also play an important role in determining the K-conduciveness of an epistemic environment. Considerations about frequency are important because they make it possible to measure the goodness of an environment also as a function of the sheer number of

interactions, and thus to assess the difference in K-conduciveness between environments that score well on the qualitative dimension but that differ regarding the amount of interactions. Distribution too is a key unit for measuring the relevance of the community for the goodness of the environment, since it makes available an evaluation of an environment with abundant interactions that are good and diverse but not K-conducive for the members of the community as fundamentally unjust.

Crucially, it is important to conceive of the standards of K-conduciveness and self-development as working together to determine when an environment is epistemically just. A just configuration of epistemic resources, then, will be one that scores as well with respect to standards of K-conduciveness as it is mandated by the respect of the epistemic self-development of the community as a whole. As a result, a just environment will not be one where epistemic resources are most abundant or of the highest quality, but where the quality, variety and distribution will match the needs of its community. This will mandate standards for, say, the distribution of a particular piece of knowledge such that priority should be given to those members of the community that benefit the most from it —say, because it is necessary for their self-development or for the fulfilment of their social role. Similar standards will be in place for regulating the quality, variety and frequency of epistemic resources as well, which will have to be as rich as it is required for all community members to, say, be able to understand their own experiences, to express themselves and to be understood and so on.

So: switching from an agent-centred to an environment-centred image of epistemic normativity leads to a shift of the locus of normativity from facts about the agent (individual or group) and (derivatively) of their sociality, to deeply social facts about the community within which agents operate —what I have called their environment. Thanks to this shift, new facts obtain prominence that neither the psychological or the agential image were able to vindicate —namely, facts about the schemas, the systems and the structures that operate in deeply social contexts. These new facts, however, need not substitute old facts about epistemic agents and their sociality. The proposed image-shift need not compromise the key insights offered by existing images. In fact, to the extent that epistemic environments are constituted also and most importantly by agents and the networks that keep them together, it is natural to expect that the agent's fulfilment of their psychological and agential obligations, to the extent that it contributes to the epistemic self-determination of the collective, will be integral part of the new image[33].

[33] Most importantly, however, and precisely in virtue of the fact that it brings deeply social consideration squarely within the domain of epistemic theorisation, this shift lays down the groundwork for developing an epistemological framework within which to make sense of the epistemic good in terms as the inherently social idea of epistemic justice. For the purpose of this paper, I have proposed to make sense of the notion of epistemic justice in terms of the epistemic betterment of a community, where this is achieved when an environment is conduciveK in a way that harmonises with the demands of self-determination of its community. I don't doubt there will be more appropriate ways of cashing this out [again, maybe footnote non-consequentialist strategies]. More broadly, the hope was that, by sketching this framework, a conception of knowledge as 'just belief' would be made available that could be capable of capturing the crucial sense in which knowledge is, ultimately, a deeply social phenomenon.

III.2 Epistemic Justice and White Ignorance

Now, before I conclude, let me briefly outline a way in which I believe this framework can be employed to say something about the norms that regulate our environment, in particular with respect to our case and to the phenomenon of structural white ignorance more in general.

First of all, since the epistemic betterment of the environment depends, at least in part, also by the epistemic success of its individual members, some of the obligations and the norms imposed by this image will be familiar ones —like the duty to proportion one's belief to the evidence, say, or to respond appropriately to the evidence one has psychological access to, as well as to nearby evidence, or evidence one ought to respond to in virtue of their social role. In addition to these, however, this image will also make space for other obligations, like those regarding the way in which evidence and knowledge ought to be organised in an epistemic environment. These norms will concern things like the kind or the amount of evidence one ought to have access to, the kind of inquiry one should carry out, the norms and the systems of knowledge that should be preserved or discarded, and so forth.

The DOCTOR case, for instance, offers a good example of the advantages of adopting an environmental perspective. At a minimum, in fact, the environmental image makes available a normative perspective from which to make sense of how we judge that the doctor should know that *p*. Recall that our concern, in expressing this judgement, is not to request that the doctor does something they haven't done, or to offer a negative evaluation of the way in which the doctor has performed *as a doctor*. Rather, our concern is to criticise the position of ignorance the doctor has been put in by the way the environment is structured.

One way of understanding the scope of this obligation is as imposing restrictions on how medical research should be conducted. In this sense, the judgement that 'the doctor should know that *p*' would suggest, among other things, constraints about the way in which medical research should be carried out, and possibly also about the kind of research that ought to be conducted. More specifically, we can understand these obligations as concerning the fact that a particular piece of evidence (about the fact that *p*) ought to have been made available to the doctor, in virtue of the fact that they are a doctor[34]. Among other things, this would imply that, in deciding on which fraction of the population tests for Genome Wide Association Studies should be carried out, researchers have not only ethical, but also distinctively *epistemic* obligations —such as, say, the obligation to include as wide a sample of the population as possible. These sorts of obligations would originate from the particular epistemic constraints imposed about the amount, the diversity, the

[34] Let me get the following worry out of the way: Couldn't one say that, since I am appealing to the social function of the doctor, the same explanation I am offering now about the nature of the doctor's obligation is also available to proponents of the agential image? Surely not, since the obligation I am making the case for are not *the doctor's*. The agential image takes one's social role as dictating what one ought to do/believe. The fact that one is a doctor, however, doesn't mandate that one ought to form a belief about everything medical, especially when a medical fact hasn't been discovered yet. Instead, the fact that one is a doctor does mandate that a certain kind of evidence be made available to them. This, though, expresses a norm concerning the distribution of evidence, and so it is not an agential, but an environmental obligation. This becomes even clearer if we consider inquiry cases. One's role (say, a medical researcher) may mandate that research be carried our (say, about the response of a certain population to a drug), but won't mandate anything about the kind of research that ought to be carried out (say, about the response of this or that population to that drug). Environmental considerations allow us to do something that a purely agential perspective can't do —namely, to adjudicate the (epistemic) quality of different *kinds* of research.

quality and the distribution of evidence an environment ought to respect in order to be epistemically just, and that impose that doctors (like dr. D) have access to knowledge (like the fact *p*) that puts them in a position to offer equal medical treatment to everyone in their community. A failure to do so is a failure to respect the epistemic self-determination of some participants to the community, and so a failure to uphold a just epistemic environment.

Another example of an environmental epistemic obligation that can be extracted from this framework could be obligations to preserve certain systems of knowledge and skills —like, for instance, those of ethnic minorities. These duty would become relevant in cases where, say, hegemonic structures (say, educational systems imposing a viciously partial understanding of the world[35]) is contributing to erasing or endangering the cultural milieu of an oppressed group. The endangerment of particular domains of knowledge pertaining to oppressed groups is epistemically unjust because these resources are often not only necessary for members of minority groups, whose identity and self-determination depend on their preservation, but also because it directly influence the epistemic health of the environment as well, which, by losing them, loses in the heterogeneity and richness of its epistemic network[36].

If this is right, it is possible to see how, more generally, extending epistemic normativity to include evaluative judgement of a structural kind allows us to identify the distinctively *epistemic* significance of cases of sanctioned white ignorance. For the epistemic failure in cases of sanctioned white ignorance are failures at the level of the global structure of a system of practices —i.e., they involve practices that are caused by and that help promote racial oppression. Because systems of racial oppression are linked to unfair discrimination in the access to epistemic resources, or in the control over systems of knowledge production and dissemination, as well as in the access and distribution of information, they are simply incompatible with the norms imposed by an epistemically just environment. In other words, thanks to the theoretical tools afforded by the environmental image we can hope to carve out the normative space for a distinctive epistemic demand to end structural (racial) oppression.

CODA

In this paper I have made the case for a new image of epistemic normativity, one where epistemic environments are at the centre. For how I see it, the proposed shift, although somewhat radical, comes from the same kind of mindset that has led some philosophers to reject the psychological image in favour of the agential one and, I think, it represents a

[35] There is a more general aspect to this case that applies to almost any non-British/American culture which, in the general westernisation of education and culture (think of the hegemony of English and American systems of education, or of their cultural products, like music, films or literature), is always caught in between the danger of obliteration and the fear of assimilation. But there are also plenty of more specific cases that show this general trend in greater relief. The 2019 movie "In My Blood It Runs" by director Maya Newell, for instance, offers a striking example of how the dominant White educational system in Australia (as well as the horribly repressive justice system) actively contributes to the systematic erasure of Aboriginal Australian culture.

[36] What about those minority views that are (epistemically) *bad* —e.g., those of creationists, flat-Earthers, xenophobes, sexists etc.? Is the preservation of their view mandated by the promotion of a just environment? Obviously not, and for a simple reason: they are not systems of *knowledge*. Their preservation would impinge the quality of the epistemic environment (since they trade on false beliefs) and, given the discriminatory nature of their beliefs, also the frequency and distribution of positive epistemic resources made available.

natural extension of it. Still, I suspect that it will strike some as patently implausible. This resistance, I believe, is rooted in a certain reductive tendency that is still fairly common in epistemological theorisation, and that I take to be an expression of what I have called the psychological image of epistemic normativity.

Ultimately, the friction between the environmental and the psychological image concerns a disagreement about the scope of the normativity of the epistemic domain. With respect to this question, my view, as well as other individualist and socialist accounts, elaborates an answer that goes beyond simply *assuming* that the epistemic domain extends beyond the psychology of individuals. But I take it that sceptics might still be right. It may be that epistemic normativity is ultimately an *intra*personal matter, and that all these views, including mine, are wrong.

Indeed, imagine that they *are* all wrong. Assume that the sceptic is right, and that the scope of epistemic normativity *does* coincide with our ken. Obligations to respond to a particular piece of evidence, to take up a particular inquiry, or to regulate the access and distribution of information are never *genuinely* epistemic. Is this view plausible?

Maybe. But note this: *even if* the sceptical option were a live one, it would surely be a very convenient one. For isn't indeed convenient, say, for a husband to think that there is nothing *epistemically* problematic in ignoring their wife? Or for a White doctor to think that there is nothing epistemically problematic in not knowing how to treat an Indigenous Latina patient? Or for a particular fraction of a community to think that there is nothing epistemically bad about being systematically ignorant about the harms suffered by another part of the same community? Or —for what is worth— isn't it convenient for a particular academic tradition (say, the analytic tradition in philosophy) to think that there is nothing epistemically problematic in ignoring another academic tradition (say, the continental one, or any other really)?

For some, the sceptical option is convenient because it absolves them from recognising their obligation to listen to someone's testimony, to inform themselves about this or that species of harm, to come to know this or that injustice or the experiences of this or that group. Having the choice *not* to confront an injustice, or a harmful dynamic affecting members of a particular community, however, is a privilege. For some, injustices are not something one can choose to ignore. For these people, scepticism is not an option. It is only to those who have the choice to ignore, that scepticism presents itself as an option —the convenient option for the privileged to hold on to their privilege.

If my view is right, epistemic perfection is incompatible with absolute epistemic isolation. Rationality is not a private affair. There are epistemic norms that are imposed by the fact that we live with other people. One way to understand these obligations (although, I am sure, not the only one, and probably not the best) is as obligations to guarantee that positive epistemic goods are exchanged, upheld, and made available within our community. To the extent that it licences epistemic isolation, then, according to this view, this sceptical tendency *is*, ultimately, epistemically unjust.

Chapter Two

Conformism, Ignorance & Injustice: AI as a Tool of Epistemic Oppression

ABSTRACT

From music recommendation to assessment of asylum applications, machine-learning algorithms play a fundamental role in our lives. Naturally, the rise of AI implementation strategies has brought to public attention the ethical risks involved. However, the dominant anti-discrimination discourse, too often preoccupied with identifying particular instances of harmful AIs, has yet to bring clearly into focus the more structural roots of AI-based injustice. This paper addresses the problem of AI-based injustice from a distinctively epistemic angle. More precisely, I argue that the injustice generated by the implementation of AI machines in our societies is, in some paradigmatic cases, also a form of epistemic injustice. With a particular focus on AIs employed as gatekeepers of our epistemic resources, this paper shows how their epistemically conformist behaviour is responsible for the marginalisation and the ostracism of minoritarian perspectives. Because it clarifies key structural flaws and weaknesses of current AI design, this paper helps make headway in critical discussion of current AI technologies. And because it forges new theoretical tools to understand forms of epistemic oppression, this paper also contributes to the advancement of feminist theorisation.

I. AI, JUSTICE AND THE FUTURE OF RESEARCH

Consider the following examples:

> GOOGLE SEARCH Lately Irina, a young girl at the age of puberty, is experiencing new feelings for girls her age and, in an attempt to understand more about her emotions, she goes on Google. The overwhelming majority of information she finds, however, is evidently shaped by heterosexual and cis perspectives, and the search exposes her to overly sexualised content. As a result, not only she doesn't find an answer to her questions, but the research also instils in her a view of her sexuality that she doesn't feel as though it reflects her own.

> ASYLUM SEEKER Negasi, a young Black man migrating from Ethiopia by way of Sudan, Chad and Libya, is seeking asylum in Germany. Their asylum application is processed via a new fully automated procedure just implemented by the Home Office. Despite having all the right credentials, and their story being true, Negasi's asylum application is unjustly rejected.

The widespread implementation of machine learning algorithms in services we rely on in everyday life has heightened the concern about new automated forms of oppression. The examples above show just a few paradigmatic cases of AI-based injustice systematically affecting members of minority groups. But the list is much longer. In *Algorithms of Oppression*, Sofia Noble gives a detailed analysis of the wide-ranging forms of sexist and racist prejudices that have been consistently found by typing racialised qualifications of individuals on the Google Search engine. More recently, translations from Hungarian, Finnish, Filipino and other gender neutral languages to English have revealed that Google Translate automatically assigns female and male pronouns to genderless sentences according to stereotypical characterisations of genders. Translated to English, gender neutral sentences in Hungarian would read as follows: "She is beautiful. He is clever. He makes a lot of money. She bakes a cake. She is a cleaner. He is a professor. She is raising a child. She cooks. He is researching. He owns a business." (Ullmann 2021). But Google is not the only culprit. A study conducted by UC Berkeley on the algorithms employed to calculate targeted interest rates has found that information about borrowers (their geographical location, sexual orientation, spending habits etc.) allows the algorithms to profile ethnic minorities (who share comparable life conditions, such as living in financially isolated areas or being unable to do comparison shopping) and charge higher interest rates compared to White borrowers with comparable credit scores (Miller 2020). In criminal law, an investigation conducted in Florida by ProPublica (Angwin et al. 2016) on the scores assigned by AIs to rate a defendant's risk of future crime, has revealed that the machine was particularly likely to falsely flag Black defendants as future criminals, wrongly labelling them at almost twice the rate as White defendants, as well as mislabel White defendants as low risk more often than Black defendants.

These cases display situations where AIs failed to function properly. Google Search failed to provide adequate results for the search query inputted, the algorithms employed to calculate targeted interest rates failed to assess the creditworthiness of their applicants, and the risk scores have been found to unjustly favour White over Black defendants. These failures exacerbate unwarranted and unjust disparities, and generate harm. A working single mother that is denied a loan, for instance, is harmed financially, whereas Google's identity prejudices are liable to cause psychological or social harm.

Paradigmatic cases of AI-based injustice of this sort are now attracting the attention of the public and of the academic world, and have long been at centre stage for tech developers and researchers on the ethics of AI. Bracketing reactions of scepticism, the relevance of these cases is often taken to lie in the challenges that they present to the fast-growing practices of development and application of AI-based technologies in our everyday lives.

These challenges have contributed to shaping an understanding of AI not only as a useful tool that we can rely on, but also as a culturally and historically determined product

that we must learn to use responsibly. Indeed, concerns about ethics and social justice have always accompanied the history of technological advancement. Today, our culturally specific image of AI (Cave & Dihal 2020), its intrinsic biases (Noble 2018), and its connection with discrimination and harm (Bender 2021, Gandy 1998 and Adam 1998), are widely recognised to have a critical impact on our societies. These themes are now at the forefront of research on the ethics of AI, and constitute the theoretical premise of future development. The idea is that only by reflecting on the risks involved in its use can we hope to develop a more responsible relationship with AI in a way that can help us confront issues of social inequality, discrimination and oppression rather than exacerbate them.

Still, for the most part, critical theorising within AI has leveraged on a narrow and potentially damaging toolset, such as focus on singular 'bad actors' (Hoffman 2019). Take for instance the case of the report on the biases of AI proposed by Collett and Dillon in 2019. The report highlights concrete cases of gender prejudices in contemporary AI technologies. One of the cases discussed is that of automatic web-assistants, which, it has been found, are often characterised with stereotypical female attributes. The report proposes an informed and lucid analysis of the dangers associated with these kinds of practices, broadly connected with the perpetuating of stereotypical gender roles. In response, it is envisaged the possibility of overcoming this problem by suggesting practical solutions (i.e., changing the gendered attributes of the assistant) and encouraging collaboration between AI developers and gender theorists.

Examples of this sort show that the way in which specific AIs are designed and function must be scrutinised if we want to prevent them from inheriting the bias of their developers, and that this cannot be done without a tighter interdisciplinary collaboration. Indeed, the problem does sometimes boil down to identifying tech designers' and engineers' *dead spots* —that is, the unquestioned set of assumptions that is part of their cultural background (Snow 2018). But developer bias cannot be the sole cause of AI-based injustice. Oftentimes developers themselves fail to understand exactly why AIs develop certain prejudices. In these cases, there is a lack of interpretability of "black box" machine learning models —i.e., extremely long and complex sequences of algorithms whose functioning is impossible to predict for humans— that is not imputable to developer bias alone.

More in general, however, attention to developer bias has been criticised because it risks blurring our perception of the *structural* nature of the injustice at play —that is, both its connection with broader systems of oppression and in the sense in which AI-based injustice is necessitated by the very structure of AI systems in general. Contrary to this trend, an important strand of critical theorising within AI promotes a more systematic approach to AI-based injustice, interested in the multifaceted ways in which we interact with AI and actively contribute to strengthen and validate existing discriminatory social structures. As part of this 'structural turn' (Bagenstos 2006), work has been conducted to understand the limitations of narrow and mechanistic approaches to AI injustice (Hoffman 2019) and the importance of psychological (Krieger 1995) or cultural studies (Browne 2015) in giving central stage to broader concerns of social justice.

In line with this structural turn, I address the problem of AI-based injustice from a distinctively epistemic angle. More precisely, I will argue that the injustice generated by the implementation of AI machines in our societies is, in some paradigmatic cases, also a form of *epistemic* injustice —namely, affecting us in our role as epistemic agents. The following

discussion develops in three steps. First (§2), by looking at machine learning-based AIs employed as a gateway to our epistemic resources, I identify two interlocking concerns (i.e., what I call *toxicity* and *deficiency*) about their functioning and the training practices. These concerns, I show, stem from the adoption of a fundamentally flawed principle of *epistemic conformism* in the very design of machine-learning based AIs. §3 leaves discussions about AI design behind to focus more specifically on the epistemic harms arising from their implementation and the way in which they contribute to reinforce structural oppression. More precisely, I argue that machine learning-based AIs erect barriers against AI-users, specifically targeting members of minority groups in their capacity as epistemic agents. In particular, following Mason (2011), I show how, seen as a form of 'hermeneutical lacuna', the toxic deficiency of AI harms agents as knowledge *seekers*, while understood as a form of 'white ignorance' (Spivak 1999, Mills 1997, Martín 2021) it risks harming them as knowledge *givers*.

Here, the importance of the discussion for feminist theorisation is brought to light as two new forms of epistemic injustice are individuated: what I call *zetetic injustice* and *testimonial spurning*. The former, an expansion on Fricker's (2007) taxonomy, concerns agents who are unjustly obstructed in their attempt to carry out meaningful inquiry. The latter, building upon Kristy Dotson's (2011) notion of epistemic violence, and akin to her notion of testimonial quieting, concerns agents who are unjustly prevented from obtaining what it is in their right to obtain with their words.

II. BIASSED DATA AND EPISTEMIC CONFORMISM

The quantity of content produced and stored online is incredibly vast. According to rough estimates, it amounts to over 30 zettabytes. To give an idea of the size of this number, consider that streaming it using the fastest networks available would take over 2000 years[37]. The exponential increase, over the last few decades, of online data has urged experts to come up with clever solutions to help us navigate it comfortably. This challenge has been met by making recourse to intelligent 'sorting machines', trained to recognise and group together recurrent patterns of information among vast pools of data. Today, most of the streaming services we use everyday (Instagram, Netflix, Spotify, Youtube), systems of recommendation (Google, Baidu) and rating services (credit and assurance risk assessment, medical and legal services, etc.) is underpinned by the functioning of these machine, specifically designed to supervise and mediate access to specific epistemic environments —i.e., pools of online data. The rise of AIs of this sort has been possible thanks to the introduction, in the early 90s, of a sophisticated method of data analysis known as *machine learning* (ML). Machine learning is a term used to refer to a technique that consists in applying long strings of algorithms —long sequences of functions, or rules, that extract predictions from a given set of input values— and statistical analysis to numerical input values to produce numerical or binary (yes/no) outputs. More broadly, the term "machine learning-based AI" is generally used to refer to long strings of complex functions that have information (e.g., a search query) as input and output predictions (Hao 2018). ML-based

[37] Statista Research Department (2022)
https://www.statista.com/statistics/871513/worldwide-data-created/

AIs are thus essentially *predictive* machines. On the input of our online interactions (clicks and likes) and personal information (geographical location, gender, age, occupation etc.), ML algorithms extract and use patterns to predict our future clicks and likes.

To see how this works more precisely, consider the case of Spotify. Spotify is a music streaming service equipped with a content recommendation system powered by ML algorithms. When you listen to a song (album, artist or podcast), the system compares information about that song (e.g., the artist, producer, etiquette, genre, rhythm, melody, pitch) with patterns of information about content in Spotify's database that share similar characteristics. In this way, after we listen to a song by the Beatles, it may suggest songs by John Lennon (in virtue of the similarity between the song's artist and artist suggested) or by The Kinks (in virtue of a similarity between their pitch and melodies) and so on. Spotify's functioning depends on the fact that the machine has been trained to recognise the similarities between the input information (the question we ask Google, the song we listen on Spotify, the series we watch on Netflix, the digital request we submit for a loan etc.) and the trends and patterns of information present in their database ("hip hop music", "philosophy podcast", "drama series" etc.).

Patterns and trends are thus crucial to Spotify's ability to read and interpret the input information and output the prediction. In modern ML-based AIs, patterns are individuated through a procedure known as *data mining*, which consists in the sorting of information through a process of statistical analysis. Statistical sorting is a crucial part of ML-based AI functioning, as it provides a rationale (i.e., *statistical frequency*) for the identification of the trends and patterns that are then used to read and interpret the input information and finally output the prediction. AI's reliance on statistical analysis makes another factor crucial for its well-functioning, namely the *size* of the training data. Data is the raw material that is fed into ML-based AIs and that fuels its sorting engines. Because these machines function by selecting and identifying data patterns on the basis of their statistical frequency, the ability to identify diverse and reliable trends depends on the availability of large pools of data. The bigger the pool, the more solid and varied the trends available, and so the more accurate and adequate the machine's predictions.

To summarise, then, the more statistically robust a piece of information —i.e., the larger the amount of information that bears a relationship of close similarity with it— the more likely it is that the machine will be able to read and understand that piece of information (i.e., a search query), and provide accurate responses to it in the future. To simplify things, we can call the relevance a piece of information has with respect to the the machine's epistemic and hermeneutic abilities (i.e., the ability to read, understand and respond to that input information) *epistemic relevance*, and say that the epistemic relevance of a piece of information is just a function of its statistical robustness. The more common the input information, the more likely it is that it shares similarities with patterns already identified by the AI and present in its epistemic environment, and so the greater the machine's ability to read and understand it and output predictions that are adequate and accurate.

In what follows, the focus of my discussion will be on ML-based AIs regulating access, participation and contribution to shared online resources. Sometimes, I will be interested in this role as consisting in mediating the retrieval of information from online pools of data. In this case, the discussion will focus on search engines and recommendation systems, like Google, Spotify and Youtube, whose role is to help users navigate resources stored online.

Sometimes, I will be interested in ML-based AIs as regulating participation in epistemic environments and practices —like when, for instance, AIs are employed for the assessment and evaluation of the liability, creditworthiness, or credibility of their users. More in general, then, in this paper I will be looking at AIs as gatekeepers of particular pools of information within our broader epistemic environments[38].

In light of the increasing importance ML-based AIs are assuming in everyday life as gatekeepers of shared knowledge, ML-based AI's reliance on data mining and statistical sorting procedures has been the focus of harsh criticism[39]. In a recent article, Bender et al. (2021), refer to AI employed in the generation of text (like the recent GPT-3, BERT and Switch-C) as a *stochastic parrot*, on the grounds that machine functioning consists in "haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning" (2021, 617). It would be misleading, they warn us, to take AI intelligence to be based on the machine's ability to engage in genuine critical thinking, since all it boils down to is the mere parroting of the most common trends of information detected in its training data. Assuming that Bender is right, and that it is true that AIs do have features justifying the association between the hermeneutical abilities of ML-based AIs and stochastic parrots, I want to propose a characterisation of the functioning of ML-based AIs in analogy with conformist attitudes —i.e., as instantiating a tendency to value or endorse attitudes and behaviours that are commonly accepted *simply because* they are commonly accepted. More exactly, what I want to suggest is that ML-based AIs could be characterised as exhibiting something in the vicinity of the following feature[40]:

> *Epistemic Conformism* The tendency to only treat as epistemically relevant information that is statistically dominant *because* it is statistically dominant,

where the epistemic relevance of a piece of information is just a measure of the likelihood that that piece of information is understood and offered an adequate response by a ML-based AI. The thought here is that AIs' conceptual repertoire is based on the resources present in statistically dominant trends; by referring to AIs as epistemically conformist, then, my aim is to elucidate the idea, implicit in the idea of AI as 'stochastic parrots', that AIs simply *mimic* common trends present in their training data[41].

---

[38] Thinking of AI as gatekeepers of shared online epistemic environments does not exaggerate the importance of AI in our everyday lives. Considering that a great deal of information we possess today is stored online and accessed via AIs, their importance can hardly be overstated. Moreover, thinking of AIs as gatekeepers is not to think of AI as the sole gatekeepers of *all* epistemic resources.

[39] Bender et al (2021), Krieger (1995), Hoffman (2019), Gandy (1998)

[40] Clearly, I take the claim that AI machines do as a matter of fact possess this trait to be contentious as it depends, among other things, on the plausibility of treating AIs as epistemic agents. But this should not constitute an obstacle to the point I want to make here, which relies merely on the plausibility of recognising some degree of analogy between the functioning of ML-based AIs and conformist attitudes conceived along these lines.

[41] Note that, despite their similarity, the notions of 'stochastic parrot' and 'epistemic conformism' are importantly different. First, because the notion of 'stochastic parrot' is used by Bender to criticise the idea that ML-based AIs can be thought of as competent language users and that they can understand what they are saying. My notion of 'epistemic conformism', on the other hand, is neutral with respect to issues concerning whether ML-based AIs are competent language users, whether they understand

But referring to AI's functioning as conformist, to the extent that it may suggest that AI machines *merely* mirror the content and structure of our linguistic practices and conceptual resources, can be misleading. ML-based AIs are not neutral tools: they play an active role in shaping the resources to which they mediate access. Consider again the case of Spotify. For those who rely on Spotify as their main access to multimedia content, the Spotify recommendation system influences the distribution of the contents and their availability by singling out those patterns in one's listening preferences that bear closer similarity to the patterns that the system deems more relevant, and suggesting predictions based on that. What's more, because such relevance is measured in terms of statistical robustness, information will be distributed in such a way as to make more readily available 'trendy' information, while unpopular content will be more difficult to identify and retrieve. Think for instance of the different results you obtain depending on the kind of search query typed into the Google Search box. The accuracy and adequacy of search hits relating to common queries (e.g., "interpretation of the song 'Hey Jude', by The Beatles") are much higher and diversified than that of queries relating to a domain or a topic that doesn't get as many search hits (e.g., "interpretation of the song 'Gli Impermeabili', by Paolo Conte").

Because it measures epistemic relevance on the basis of statistical robustness, then, we can predict that ML-based AIs' epistemic conformist functioning will lead to the formation of knowledge-gaps and interpretative lacunae, affecting the machine's ability to read, understand and respond to minoritarian information (i.e., pieces of information that bear little to no similarity to statistically robust patterns). As a result of their conformist behaviour, then, ML-based AIs appear to manifest a fundamental lack of interpretative power —that is, a structural *deficiency*— with respect to minoritarian vocabularies, language norms and systems of meaning. Crucially, because it is due to its epistemic conformism, AI's deficiency is part of the machine's very *design*. It is the AI's conformist behaviour that, because it grounds the epistemic relevance of a piece of information on its popularity, encodes a fallacious epistemic principle leading to the systematic marginalisation of minoritarian information and the formation of lacunae in our epistemic environment[42]. This principle underlies the functioning of the sorting mechanism whereby patterns of information are identified, and that in turn determine the machine's ability to provide adequate and accurate predictions. Being marginalised, patterns of minoritarian information fail to be identified, and thus form part of the machine's interpretative tools, which is then in this sense importantly *deficient*[43].

---

what they are saying —or, for that matter, about the relationship between the two. With the notion of 'epistemic conformism', instead, I wanted to capture the distinctively epistemic principle underpinning the functioning of ML-base AIs. In this sense, and in line with the general scope of this paper, we could arguably say that 'epistemic conformism' could be taken to clarify the epistemic aspect of the notion of 'stochastic parrot'. (I wish thank an anonymous reviewer for bringing up this point)

[42] This is true even if conformist ML-based AI does, as a matter of fact, provide accurate responses in most cases. Indeed, the problem does not have to do with the overall rate of successful responses given by AIs, but with the conformist design itself, which causes the AI to make epistemically relevant distinctions between types of information on the basis of facts that should not matter *epistemically* —namely, their statistical frequency. I thank an anonymous reviewer for pressing me to clarify this point.

[43] Note that the word 'minoritarian' here is used in a strictly statistical sense. The content that is marginalised is simply content that fails to meet the threshold required for it to be read and

Notice at this point that all I've said so far tells us nothing about the normativity of the environment that is thus affected —that is, whether AI's conformism affects it for the better or worse. In fact, conformist attitudes are in some respects *neutral*: although they do impact the distribution of information in a determinate environment, they do so on the basis of a sorting principle that doesn't take into account its quality. Indeed, AI's conformist behaviour might uphold *good* just as much as *bad* epistemic environments —the minoritarian views screened off by the algorithms may be climate scientists' opinions on climate change just as much as neo nazis' claims about national identity, and whether AIs' conformism end up upholding either will depend on empirical facts about the epistemic quality of the statistically dominant strands of information.

A recent study conducted by Emily Bender and her team (Bender et al. 2021), focussing on Google's norms of implementation (although it refers to practices that are now widely standardised) has revealed that, ML-based AIs' need of large swathes of data is met by relying on the largest database available today —namely, online networks and communities such as Reddit and Wikipedia. In particular, the aim of Bender's article is to highlight the dangers that are associated with such practices. These span from the environmental costs of the data mining procedures (linked to the extraordinary processing power they require) to the way AIs are perceived in our society (ML-based AIs capacity to analyse and produce intelligible pieces of text gives the false impression that the machine can understand natural language). More importantly, however, their work draws attention to a fundamental problem connected to the quality of the information that is gathered from these sources. These concerns primarily stem from the consideration that access to the internet and its use are a prerogative of people from richer countries, and is more substantial among the wealthy White male youth (Bender et al. 2021, Roser & Ritchie & Ospina-Ortiz 2015). "GPT-2's training data" they argue, "is sourced by scraping outbound links from Reddit, and Pew Internet Research's 2016 survey reveals 67% of Reddit users in the United States are men, and 64% between ages 18 and 29. Similarly, recent surveys of Wikipedians find that only 8.8–15% are women or girls" (Bender et al. 2021, 613). Moderation practices regulating access to subsamples of the internet are also cited in this study as having a substantial discriminatory impact. A research conducted using digital ethnography techniques on Twitter (Jones 2020), for instance, has shown that people on the receiving end of online discrimination are more likely to have their account suspended than those perpetrating it.

Epistemic environments where discriminatory, biassed and harmful contents and norms prevail are *toxic* epistemic environments. Since empirical research gives an image of our shared online resources as expressing the world-view of dominant groups, reflecting their biassed, harmful, and often colonising view of the world, our shared online resources are thereby *toxic* in this sense —i.e., in the sense that they are permeated with contents and norms of bad epistemic quality.

adequately interpreted by the algorithms. A connection between minoritarian content and content expressing the world-view of non-dominant groups is proposed towards the end of this section.

In summary, then, I've pointed out two main concerns regarding ML-based AI design and implementation practices strategies. Because of the corruption of online resources that are employed to train AI machines, the epistemic environments to which AIs mediated access are often epistemically *toxic*; and because of the knowledge-gap generated by AI's conformist behaviour, such machines discriminate against trends of information that are statistically weaker, and is thus unjustly *deficient*. Note however how, although distinct, it is in combination with each other that toxicity and deficiency influence the implementation of ML-based AIs. In particular, in what follows I will be interested in the way in which toxically deficient AI are responsible for the epistemic marginalisation of the language norms and vocabulary of minority groups. How so? Consider again AI's deficiency. Because it measures epistemic relevance on the basis of statistical robustness, I argued, the epistemic conformism of ML-based AIs leads to minoritarian voices being systematically marginalised —that is, it prevents them from contributing equally to the formation of the shared meanings, concepts and interpretative tropes that operate within society. Because, on the other hand, the statistical weakness of online content expressing systems of meanings of minority groups is an empirical fact (what I referred to as the epistemic toxicity of AI), it is possible to see how, more often than not, the minoritarian voices that end up being marginalised due to AI's deficiency are precisely the voices of members of minority groups.

So much about AIs' functioning and their training practices. In the next sections I turn my attention from the design and function of AIs to issues arising from their implementation. In particular, the aim will be to identify the ways in which AI's toxic deficiency contributes to set up barriers to epistemic agents as knowledge *seekers* and knowledge *givers*.

III. HERMENEUTICAL LACUNAE AND WHITE IGNORANCE

Take again the two cases considered at the outset. In GOOGLE SEARCH Irina, a young girl who wants to learn more about her own sexuality, is not only unsuccessful at finding content that can help her understand own sexual experience, but throughout her research she is also repeatedly exposed to violent and overly sexualised content. ASYLUM SEEKER, on the other hand, describes the case of an Ethiopian man, Negasi, whose asylum request is rejected by a new fully automated system implemented by the German Home Office. In both cases something went wrong: Irina and Negasi's pursuit of their epistemic goals (i.e., to inquire into a topic or to acquire or transmit a piece of information) have been unjustly trumped by barriers set up by the technologies they have relied on to achieve them. Irina's inquiry was unsuccessful, and so was Negasi's application.

Crucially, these barriers have been erected by the toxic deficiency of ML-based AIs. Irina's search queries are interpreted in the light of the categories extracted from the toxic dominant trends which do not include the kind of statistically non-dominant information Irina is after. The same discriminatory content also constitutes the interpretative categories through which Negasi's application is evaluated and the grounds on which it is rejected, since ML-based AI assessed Negasi's testimony not only against its actual credential, but also as a function of prejudiced assumptions present in the training data —in this case, say,

the prejudiced thought that Black people are more prone to deception and violence, and thus less likely to give accurate testimony.

My goal in this section is to show that these two examples stand for two paradigmatic ways in which AI's toxic deficiency causes distinctive epistemic harms. I will point at two main ways in which this deficiency can affect the epistemic agency of the members of an epistemic community in harmful ways: as a hermeneutical lacuna, and as a form of active ignorance. Talking about AI's toxic deficiency as a *hermeneutical lacuna*, I will turn my attention to the way in which this deficiency sometimes impairs members of minority groups' ability to inquire into a topic or obtain knowledge regarding matters that are meaningful to them, thus harming them as *knowledge seekers*. With its identification with a form of *active ignorance*, on the other hand, I will be interested in understanding the way in which AI's toxic deficiency is responsible for perpetrating epistemic violence against members of minority groups by interfering with their ability as *knowledge givers*. Each of these barriers, I argue, becomes the source of a new form of epistemic harm. I call *zetetic injustice* the one resulting from barriers erected against epistemic agents as knowledge seekers, and *epistemic spurning* the one erected against epistemic agents as knowledge givers.

It is important to bear in mind, as the discussion goes on, that the aim of my argument is not to establish that the design and functioning of ML-based AI is detrimental to minority groups *exclusively*, nor that the harms I am concerned with here are the *only* AI-based harms we should look out for. The general scope of this part of the article is to advance feminist theorisation and critical race studies by showing some of the ways in which ML-based AIs risk contributing to worsening the oppression of minorities in our society.

III.1 Hermeneutical Lacunae and Zetetic Injustice

Based on the proposed characterisation of toxically deficient AI, the most obvious sense in which AI appears to be deficient is arguably with respect to the conceptual resources required for understanding, interpreting and adequately predicting requests pertaining to minoritarian preferences and patterns. How so? ML-based AI manifests conformist behaviour, which consists in a tendency to treat statistically robust patterns of data as epistemically relevant precisely in virtue of the fact that they are statistically robust. As a result, statistically weaker patterns of information, which fail to meet a statistical threshold, are systematically screened-off, and thus prevented from contributing to shaping the machine's interpretative resources. Because of the toxicity of the data scraped off the internet and used to train the AI, moreover, statistically weaker patterns invariably end up corresponding to the meanings, norms and interpretative tropes of minority groups.

ML-based AIs, then, lack the necessary conceptual competence to understand and interpret inputs from minority groups. If this is true, we should expect that attempts made from members of minority groups to access information that is relevant for them through ML-based AIs will fail systematically. In fact, this is precisely what goes on in GOOGLE SEARCH —because of the epistemically conformist behaviour displayed by Google's algorithms, which tends to read and interpret input information in the light of the categories extracted from the dominant trends, the overwhelming majority of information Irina gets access to concerns the heteronormative and often violent forms of sexual

expression that are most common among the majority of Google users, and that aren't helpful to her to make sense of her own sexual experience. In other words, the bias ingrained in the functioning of the Google Search engine prevents Irina from obtaining information that is relevant for her to understand aspects of her own identity.

Put this way, the case will strike those who are familiar with Miranda Fricker's notion of epistemic injustice as bearing close similarities to her characterisation of injustices of a hermeneutical variety. According to Fricker (2007), hermeneutical injustice is a particular form of injustice suffered by one as an epistemic agent, concerning one's ability to access meaningful information. More exactly, Fricker takes hermeneutical injustice to occur when prejudice ingrained in the body of shared interpretative resources hinders one's ability to obtain knowledge that is necessary to express oneself and to be understood. The prejudice is manifested in the form of gaps, or lacunae, in our hermeneutical resources —i.e., the tools, such as concepts or tropes, we use to make sense of our own experience. Now, since these lacunae occur at the level of our shared resources, and are formed and sustained by our collective meaning-making activities, hermeneutical injustice often concerns structural features of our communicative exchanges and social practices. The hermeneutical marginalisation of women, for example, is typically invoked to explain the lack, until very recently, of a fully-formed, shared concept of sexual harassment in our collective hermeneutical resources. Fricker's though is that, prior to its acquisition, victims of sexual harassment didn't have the conceptual resources required to come to know a fundamental part of their experience, and so to make sense of it.

Similarly, it seems plausible to describe GOOGLE SEARCH as a case where Irina is prevented from obtaining knowledge that is important for her to make sense of her own experience. Crucially, she is thus obstructed by a lacuna in the shared online hermeneutical resources, a lacuna that is due to the predominantly discriminatory language and biassed world-views ingrained in the data used to train ML-based AIs like the Google Search engine[44]. If this is correct, we can derive an important conclusion from this argument. That is: because the hermeneutical lacuna present in our shared online resources is a direct consequence of the very functioning and training practices of ML-based AIs, epistemic injustices of a hermeneutical kind like the one suffered by Irina, are not just unlucky byproducts of developers' biases, but a *systematic feature of the design of AI design*.

A closer look at this case, however, seems to suggest another sense in which Irina is harmed in their capacity as a knower. First of all, notice that the prejudice ingrained in the machine's interpretative resources doesn't just prevent Irina from *obtaining* the valuable piece of information she's after. Recall how, in her attempt to find out more about her own

[44] I believe that a criticism moved by Rebecca Mason (2011) concerning the limits of Fricker's model applies here. In a nutshell, this criticism is that "[a] gap in dominant hermeneutical resources with respect to one's social experiences does not necessitate a corresponding gap in nondominant hermeneutical resources." (2011, 300). Mason's point is even more obviously true in cases like GOOGLE SEARCH, where the pool of shared resources is the even restricted pool of online resources. While I agree with Mason, I think it is important to add how, even in the light of this consideration, it is still hard to overestimate the importance of dominant pools of information in one's epistemic life. This is largely because dominant knowledge is often also *sanctioned* knowledge, and is thereby granted special epistemic status. This I think is an important reason why the point made by Fricker retains special relevance even if, as Mason rightly points out (echoing Mills), hermeneutical resources are often already available outside through non-dominant channels.

experience, Irina not only fails to find what she's looking for, but her very attempt to *search* for it is repeatedly frustrated. Her queries, concerning vocabulary and concepts that aren't recorded in statistically robust trends, are systematically redirected to mainstream ones, often exposing her to violent heteronormative contents. On the face of it, then, it looks as though the hermeneutical injustice suffered by Irina is only the backhand of another barrier set up by the Google AI, this one against her attempt to conduct meaningful inquiry. To see better the kind of harm at play here consider the following case.

> SWEETGRASS Laure is a last year botany student, and she needs to find a topic for her dissertation. She has long been interested in indigenous harvesting practices, and over the years has collected various testimonies from indigenous experts regarding techniques of harvesting that, they say, would preserve and improve the quality of sweetgrass crops. She finds that experts are polarised on the topic —some say crops benefit from a harvesting technique involving the cutting of sweetgrass stems near the roots, while others favour the method of uprooting. Finally, she makes up her mind and decides to dedicate her thesis to settling this disagreement. When she presents her idea to the school, however, the academic committee refuses her research proposal on the grounds that, they say, it goes against the known scientific fact that harvesting *damages* crops. The committee also undermines the validity of the testimony of the experts gathered by Laure, on account of the fact that they are mostly old indigenous sweetgrass pickers and basket-makers, not scientists, and encourages her to focus her thesis on another project. As it turns out, research conducted several years later reveals that the scientific consensus is wrong and that, for some plant specimens like sweetgrass (like Laure had thought, backed by the knowledge of expert indigenous sweetgrass pickers) some types of harvesting *do* improve the quality and quantity of the crop[45].

Laure has evidence, gathered through years spent with people in communities in close contact with sweetgrass, suggesting a promising line of inquiry. Yet this evidence is present only in small centres at the periphery of the main streams of knowledge production and diffusion. Members of the academic committee, as gatekeepers of the mainstream, reject Laure's proposal on the ground of a conformist decision —i.e., the decision to consider as scientifically relevant and worthy of pursuit only research that complies with mainstream assumptions and knowledge. Despite promising, Laure's inquiry is thus unjustly frustrated.
Like Irina's, Laure's inquiry attempt is also threatened to be undermined or absorbed into more mainstream patterns. Like Laure's, Irina's attempt to conduct research is also unjustly frustrated by the conformist resolutions of the gatekeeping authorities. While the gatekeeping role in Laure's case is played by the scientific committee, in Irina's that role is occupied by Google algorithms. In both cases the academic committee and Google algorithms are equally responsible for perpetrating the same form of injustice: by getting in the way of Irina's and Laure's inquiry and obstructing exercise of their epistemic autonomy, they are responsible for harming the two women in their capacity as knowers. More precisely, because it concerns their distinctive ability to conduct meaningful research,

[45] This case is taken from Robin Wall Kimmerer's 'Braiding Sweetgrass' (2013)

question and, more generally, inquire into matters that are relevant for them, I propose to call this particular form of wronging *zetetic injustice*.

The concept of zetetic injustice I have in mind bears close similarity to Fricker's notion of epistemic injustice. For example, I take that, thus characterised, epistemic and zetetic injustices are structurally similar to each other, in such a way that the latter can be seen as another variety of the former, much like testimonial and hermeneutical injustices are kinds of epistemic injustices[46]. Like other forms of epistemic injustice, zetetic injustice also concerns one's epistemic conduct, and it too has identity prejudice as a key ingredient —although the examples discussed seem to suggest a pretty loose characterisation of prejudice as something that has less to do with one's cognitive commitment, as Fricker thinks, and more with structural flaws of one's epistemic environment.

On the other hand, zetetic and epistemic injustices naturally differ in important respects —most saliently, regarding the fact that zetetic injustice doesn't concern the obstruction of the transmission or acquisition of a piece of knowledge. In SWEETGRASS, the barrier put up by the academic committee against Laure's proposal does, as a matter of fact, prevent the acquisition of valuable knowledge —the knowledge that, at least for some plant specimens, harvesting can improve the quality of the crop. The zetetic wrong Laure is a victim of, however, doesn't depend on that. She would have been wronged in her capacity as an inquirer even if subsequent research confirmed the scientific consensus, or if it proved uninformative. What matters for the kind of injustice at play, instead, is merely that Laure ends up being obstructed in her attempt to carry out the research itself[47]. The (implicit or explicit) barriers raised against an inquirer will vary depending on the context, but will typically function to mislead or misdirect the investigation. In SWEETGRASS, for example, the obstruction is caused by the faulty functioning of conformist and sectarian academic practices, and involves things like discouraging the researcher from carrying out her research, offering alternative research opportunities, possibly cutting her funding and so on. In GOOGLE SEARCH, the obstruction (caused by the problematic functioning of the Google Search algorithm I have described, such as the toxic deficiency and the conformist mechanisms that systematically produce it) involved offering inadequate responses to the search queries, providing misleading information, and attempting to reconduct the investigation towards more mainstream topics.

In summary, then, looking at the epistemic impact of ML-based AI reveals that the structural faults of the machine's design lead to the systematic production of particular forms of injustice. More precisely, it looks as though the hermeneutical lacunae in our shared online resources, due to AI conformist behaviour, are susceptible to cause those who rely on them to suffer from injustice of *hermeneutical* and *zetetic* varieties. Because they concern members of minority groups' failure to obtain resources that are meaningful for them, or even to inquire into them, I take these injustices to broadly consist of impairments they suffer as knowledge seekers.

---

[46] I recognise that this may be contentious, as the relationship between the epistemic and the zetetic is a matter of ongoing debate. However, I don't think that anything substantial about my position here relies on this commitment.

[47] Naturally, the inquiry must also respect some basic zetetic norms —like, say, that one ought not to set out to inquire into whether X if one already knows that X.

III.2 White Ignorance and Epistemic Spurning

Because it is due to the toxic deficiency of AI design, the presence of hermeneutical lacunae, I have argued, tends to epistemically harm, for the most part, members of minority groups. However, it would be a mistake to think that minority groups are thereby relegated to a position of epistemic inferiority. This point, raised for the first time explicitly by Du Bois (1989 [1903]), and picked up and articulated more recently by Charles Mills (1998), reflects the fundamental insight of standpoint epistemology that "social privilege does not necessarily entail epistemic privilege" (Mason 2011, 301). In fact, the opposite is often and in crucial respects true: occupying a position of social disadvantage often puts one in a position of epistemic privilege. One influential way of explaining how this is the case is in terms of Charles Mills' notion of 'Racial Contract' (1998). According to Mills, dominant groups tend to think of their social organisation in terms of ideal, fundamentally *just* systems of meaning that exclude the possibility of the existence, from their very inception, of forms of oppression, injustice and discrimination. For this reason, an epistemic *asymmetry* is created between dominant and oppressed groups, whereby the former group, because these gaps and shortcomings are constitutive of their own world-view, tend to systematically fail to understand or (literally, according to Mills) perceive them. The latter group, instead, who often end up suffering from the lacunae in the fabric of the shared epistemic resources, and in virtue of the harm they often encounter (although not necessarily because of it, or not exclusively) become aware of them[48].

If true, Mills framework can offer a powerful interpretative key to our case. Recall that our online resources are constituted, for the most part, by content expressing the biassed, often discriminatory language and norms of wealthy White men. The toxic deficiency inherent in their own world-view, then, can become manifest to members of non-dominant groups as a consequence of the (hermeneutical and zetetic, for instance) injustices they suffer, and which are caused by the knowledge-gaps and lacunae in the shared online hermeneutical resources.

Crucially, however, despite the new awareness acquired, because of the very design of ML-based AIs —which are trained with content scraped from databases where languages and norms of the dominant groups are statistically preponderant— minority groups are systematically prevented from contributing to filling those gaps. This epistemic asymmetry leads then to a *fracture* in the shared resources between mainstream knowledge on the one hand, reflecting the world-view of the dominant groups, and informing and shaping the online resources; and non-dominant knowledges and practices on the other, which, in addition to mainstream knowledge, also include different kinds of awareness of minority norms and languages, of the gaps, the social realities and the injustices ignored by the dominant groups.

In this respect, then, the toxic deficiency of ML-based AI expresses not just a hermeneutical lacuna, but rather a form of *ignorance*. More exactly, a particular kind of ignorance that, prevalent among members of the dominant groups, is inherited by

[48] Note that this is not to say that, simply by virtue of being a member of a minority, one automatically obtains this kind of awareness, nor that all instances of injustice are revelatory of structural oppression. Yet, because they are oppressed, members of minority groups are in a position of natural advantage when it comes to obtaining awareness of the injustices and lacunae of dominant systems of meaning.

ML-based algorithms trained with content representing their (dominant) world-view. Moreover, this ignorance is not contingent, but rather a *systematic* feature of the shared online environment, produced and maintained as it is by the conformist attitude of AI design and training practices. And since it is an ignorance of the very oppressive systems that contribute to producing it, it is also not neutral, but plays an active role in upholding them, and in resisting its own erasure. Following Mills, then, I will refer to this *active* and *systematic* form of ignorance that contributes to sustain systems of oppression as a form of *white ignorance*.

In offering a characterisation of ML-based AIs as white ignorant, then, I propose to focus the attention on the following features of AI's toxic deficiency: *a*) its being part of the very design of ML-based technologies, *b*) its active resistance to erasure, and *c*) its being undiscerning of non-dominant languages, norms and systems of meaning. If this is plausible, I want to show how, while, as an *hermeneutical lacuna*, the toxic deficiency of ML-based AI tends to impair minority groups as knowledge *seekers*, seen as a form of *white ignorance* it tends to obstruct them as knowledge *givers*.

To do so let's first go back to ASYLUM SEEKER. In this example, the AI is employed to evaluate the testimony of an asylum applicant against certain parameters and, by assessing their credibility, accept or reject their request. In the process of obtaining asylum, people who have been forced to leave their own country and have often suffered trauma and violence are put through the humiliating task of providing evidence of their conditions to the authority of the host country. Evidence of trauma, fear and violence, however, often cannot be other than testimonial —asylum seekers have to provide a story detailing the circumstances that have led them to flee their country. Because this story ought to be believed for the claim to be accepted, the success of the application depends on the accurate assessment of the applicant's credibility.

In recent years, a few countries (including Hungary, Latvia, Germany, Greece, Canada, the US and the UK) have been trialling the implementation of ML-based systems to carry out such assessments (Fair Trials 2021). Perhaps unsurprisingly, these practices have sparked huge controversy over the norms and criteria employed to generate the predictions. For instance, algorithms employed by the Home Office in the UK have been shown to take the applicant's nationality as a risk factor, and to rely on face recognition systems unable track cultural and racial differences, or the impact that traumatic experiences have on the way one reports them, both at the level of one's facial expressions and in the language and vocabulary employed (Fair Trials 2021, van den Hoven 2019, Eckenweiler 2019).

When asylum is denied on such grounds, it is precisely the machine's (white) ignorance of all these factors that causes it to fail to assign the right level of credibility to the applicant's testimony. What I have in mind in this case, more exactly, is the machine's lack of resources apt to understand the system of meanings (such as the vocabulary and concepts as well as non-verbal cues and nuances of expression) of someone from a non-dominant background —like Negasi, for instance. In virtue of this lack, and owing to the prejudice ingrained in the machine, the categories applied by the algorithm to read and interpret Negasi's asylum application are inadequate to fairly assess the credibility of Negasi's testimony.

At the root of the injustice, then, a key role is played by a fundamental *communicative failure*. At bottom, that is, is the AI's failure to give a proper assessment of the applicant's credential that, in this case like many others, leads to the wrongful rejection of the applicants' request. Communicative failures of this sort, owing to the bias ingrained in a hearer's deficient conceptual resources, have been widely discussed in the literature on epistemic injustice. According to Kristie Dotson, for example, one can be a victim of a particular form of testimonial injustice (what she calls *testimonial quieting*) when a communicative failure is caused by a hearer's *pernicious* ignorance —that is, a kind of reliable ignorance that, in a particular context, tends to be harmful. More precisely, testimonial quieting involves cases where the pernicious ignorance of a hearer, in the form of negative stereotypes, or 'controlling images' (2011, 243), prevents them from perceiving the speaker as a knower, which causes them to fail to take up their communicative attempt. The communicative failure Dotson has in mind here is a form of illocutionary silencing, occurring when a hearer fails to take up a speaker's attempt to transmit a piece of information —for instance, when a woman's attempt to contribute to a conversation is taken to be a mere expression of her emotions[49]. The patronising interjection to "calm down, dear", uttered by the then UK Prime Minister David Cameron in response to a criticism by Angela Eagle's (then Shadow Chief Secretary to the Treasury), for instance, is a stark example of what Alessandra Tanesini calls 'haughty' illocutionary silencing. In order for a communicative attempt to be successful, it must be recognised (or treated) as the speech act it is (intended to be). In this case, the Angela Eagle is said to be *silenced* because her utterance is not successful at being recognised as the kind of speech-act the woman intended it to be.

Thus understood, the epistemic violence of testimonial quieting bears intuitively similarity with the kind of injustice described in ASYLUM SEEKER. In both cases, the communicative exchange fails, and in both cases (systematic and wrongful) ignorance plays a key explanatory role. More exactly, in our case, it is the AI's ignorance, rooted in the machine's biassed functioning, that causes the algorithm to fail to assess the applicant' epistemic worth, ultimately leading to the communicative failure.

True, the testimonial exchange in ASYLUM SEEKER may not be considered strictly speaking *testimonial*, because it takes place between a human and a machine, and human-to-machine interactions do not obey the same norms as human-to-human —or so one may think. But it is at least not intuitively obvious why this should be a problem, at least with respect to the conversational norms relevant to this case. Indeed, it seems reasonable to expect that the conversational norm that is at stake here doesn't apply only to human communicative exchanges. After all, it is difficult to see how AIs could, say, give us the right predictions if they didn't recognise our requests as such —if they took, say, one's asylum request as a greeting. Even so, I do ultimately agree that it would be a stretch to subsume this case under the notion of testimonial quieting —at least in the way in which Dotson understands it. The reason is that the communicative collapse in this case does *not* involve a *failure of uptake*. Negasi's application has been *rejected*, which means that, at the very least, his speech act *is* acknowledged for what it is —i.e., an asylum request. If this is so, however, ASYLUM SEEKER does not describe a case of *illocutionary* silencing.

[49] Case discussed in Tanesini (2016)

What's the issue in this case then? To a first approximation, I think that the problem can be understood as concerning the fact that the algorithm has prevented Negasi from obtaining the effect that, given their credentials, they were entitled to obtain with the communicative act they performed. If this is correct, the communicative failure at issue here does not concern *uptake* of the communicative act, but its *effect*. In other words, it is *perlocutionary* rather than illocutionary. The applicant has been perlocutionary silenced: they have been unjustly prevented from obtaining something that they were entitled to obtain with their words (Spewak 2023).

When considered in their capacity as receivers of information, then, ML-based AIs are liable, owing to their active ignorance, to perlocutionarily silence members of minority groups' communicative attempts. The harm caused by having one's perlocutionary attempt frustrated is very common, and has recently been aggravated by the increased implementation of ML-based AI technologies. Studies by UC Berkeley, for instance, have found that Black people were consistently refused house loans due to the bias present in newly automated systems employed to process loan applications, which unjustly discriminated against applicants based on their ethnicity. Similarly, an investigation conducted by ProPublica in 2016 on the fairness of the criminal law system in Florida, has revealed that Black defendant's testimony were evaluated against an assessment of their likelihood to reoffend, which was in turn produced by ML algorithms that systematically discriminated against all non-White defendants.

These cases present patterns of injustice similar to the one in ASYLUM SEEKER, where a member of a minority groups' attempt to obtain something through their communicative act is unjustly frustrated due to the systematic ignorance of ML-based AIs. Notice though how victims of AI-based perlocutionary silencing are clearly not *quieted*. Their communicative attempt doesn't go unacknowledged —instead, it is heard and taken up for what it is (a loan application, an asylum request, a non-guilty plea). The problem is rather that, in failing to obtain its goal, the attempt remains somewhat inert. Although it *is* heard, it is as though it wasn't. The Black woman who has applied for a loan, and whose request is being processed by the system, *has* been heard, and her communicative act has been taken up for what it really is —i.e., a request for a loan. Because it gets rejected, however, the request is unsuccessful, and she is unjustly prevented from obtaining what she had the credentials to obtain through that communicative act. Following this line of thought, then, we can say that the communicative attempts of victims of perlocutionary injustices, rather than being *quieted*, are unjustly shunned, or *spurned*. Expanding on Dotson's taxonomy, we can call *testimonial spurning* the kind of epistemic violence occuring when active ignorance systematically dismisses the perlocutionary effect one is otherwise entitled to obtain with one's communicative act.

Looking at the toxic deficiency of ML-based AIs as a form of white ignorance then reveals a distinctive form of violence that, for the most part, targets members of minority groups in their capacity as knowledge *givers*. Following Dotson, I have proposed to think of this violence in terms of a communicative failure occurring when (white) ignorance causes one to fail to recognise the epistemic worth of their interlocutor. Departing from Dotson's analysis, and in an attempt to adding to it, I have suggested that, when it comes to theorising about ML-based forms of epistemic injustice more specifically, the communicative failure is better understood as concerning the perlocutionary effects of the

speech act rather than its illocutionary force. Owing to this difference, I noted how the violence thus perpetrated concerns not the quieting as much as the spurning of one's testimonial attempt.

IV. CODA

In this paper I have tried to do two things. First, I have looked at the design and training practices of ML-based AIs, and tried to show how this seems to present systematic flaws, and how these flaws appear to be, to some extent, the result of the implementation of a fundamentally mistaken principle regulating the machine's behaviour —what I called *epistemic conformism*.

Honing in on these results, I then tried to show how these design flaws impact AI users in their capacity as epistemic agents. In particular, looking at ML-based AIs in their function as gatekeepers of the knowledge stored in our shared online resources, I focussed my attention in particular on two basic epistemic aspects of the users' agency: their ability to seek and to pass on their knowledge. What I have found is that, with respect to both their knowledge seeking and knowledge giving abilities, ML-based AIs tend to set up barriers mostly affecting members of minority groups. The reason, I have argued, ought to be found precisely in the specific structure of the design flaws of AI —particularly its toxicity and deficiency. More exactly, I have shown how the barriers erected against minorities' ability to give knowledge is connected to the white ignorance of AI, and how the barriers erected against minorities' ability to seek and obtain knowledge are connected to its hermeneutical lacunae.

If plausible, this seems to suggest a picture of ML-based AIs as systematically *ostracising* minority contributions. Considering the role that ML-based AI nowadays plays as gatekeepers of our shared online resources, and considering our increasing reliance on online content in our epistemic lives, the outright ostracism of minoritarian voices poses a serious threat to the integrity of our epistemic environments.

The growth of ML-based AIs, both in sophistication and extension of their application, is just at the beginning. The increase in implementation of ML-based technologies in everyday life is rapid and widespread. Since I started working on this article (in 2020, when my interest in machine learning was sparked by reading of the firing of Timnit Gebru from Google ethics team[50]), the boom of AI has been exponential —in terms of the technologies that have been made available to the public (e.g., chatGPT or dall-e); in terms of the political and financial attention it has raised (e.g., more and more funding opportunities made available by governments all over the world to secure leadership in AI-related areas of research); and in terms of the critical attention it has raised (e.g., regarding worries about online assessments, or the fights over creative rights). Still, very little is being done to match this enthusiasm with sufficient critical examination. If anything, when we hear of Google's decimation of their ethics team, followed by Twitter and Microsoft *en masse* suppression of theirs, the impression is rather that helpful criticism is being stifled.

[50] Hao (2020)

Yet, I don't think that pessimism about the future of AI in our society is fully justified. We already have the conceptual tools and critical capacities to understand the threats posed by these new technologies and to improve them. Attention to the relationship between the ways in which we design and use AIs and issues of social justice is steadily increasing. New work (e.g., Huang, et al. 2022; Simion and Kelp 2023; Rafanelli 2022) is shedding light on possible solutions and virtuous models we can follow to develop better and more just AI. This paper should also be seen as an attempt in this direction. If I am correct, one optimistic stance is not justified: the one endorsed by those who take AI to be a neutral tool. According to this stance, the problem is not to be found in the functioning of AI itself, but in the way in which we make use of it. If I am right, we shouldn't find this stance fully satisfactory. For if, on the one hand, it is true that better training practices, as well as wider participation to online pools of data, may make for more virtuous AIs and alleviate some of our current worries, a solution to the problem requires more than that. And this is because the problem I have identified concerns AI's very design. The epistemic conformism of AIs is a design flaw which needs to be addressed directly. Failing to address this worry, I have argued, leads to distressing epistemic worries, like the epistemic marginalisation and ostracism of minoritarian perspectives.

Chapter Three

What Is Mansplaining?

INTRODUCTION

In 2016 NASA astronaut Jessica Meir tweeted a video of herself in a space equivalent zone 63,000 feet above the Earth's surface observing how, in those extreme conditions, water spontaneously boiled. In response to her tweet a man, with the presumption of correcting her observation, explained to her the phenomenon in simple terms and with a condescending attitude. Undoubtedly, the man's tweet was all but necessary. In fact, it wasn't simply unnecessary, but unjust. And it exhibited the tweeter's disrespect for Meir. The tweet attracted the attention of the media, and its author was told off for his arrogance and disrespect.

Mansplaining is part of the digital discourse, and the vicious behaviour it portrays is now well-known and often discussed. Despite this, scholarly attention hasn't yet matched the increased visibility of this phenomenon. With few exceptions, most debates have so far been solely reactive, triggered in response to cases brought to the public attention by high-profile figures. Predictably, this has led to a hasty and often inaccurate appreciation of this phenomenon. As a result, it has been systematically underestimated both in its severity and extension.

With this problem in mind, this paper develops a fully fledged account of mansplaining as a form of epistemic injustice stemming from the violation of a norm of cooperative conversation. The account we propose also shows that the common understanding of the phenomenon of mansplaining is often inaccurate and prone to generate harmful mistakes.

The argument unfolds as follows: §1 offers a quick overview of the literature on this phenomenon, with particular attention to the work of Solnit (2014), Manne (2020) and

Johnson (2020), and concludes by identifying the analysis of mansplaining that is suggested by their proposals —what we call the Standard View. §2 follows up this discussion by showing that the Standard View is too narrow. A novel, more inclusive view of mansplaining is here proposed, and the relevance of the phenomenon of mansplaining for contemporary epistemology and feminist theorisation is finally brought to light.

I. MANSPLAINING: THE STANDARD VIEW

Consider the following examples:

> NASA Astronaut Jessica Meir tweets a video of herself in a space equivalent zone 63,000 feet above the Earth's surface, observing how, in those extreme conditions, water spontaneously boiled. In response to her tweet a sexist man, with the presumption of correcting her observation, explained to her the phenomenon in simple terms and with a condescending attitude.

> LAB Elohor and Ekon are in a laboratory, working on an experiment they have designed together. Also, Ekon, who is terribly sexist, believes he is epistemically superior to Elohor just because he is a man. While running the experiment, Elohor asks Ekon to confirm the exact temperature of incubation for an enzymatic reaction to proceed in the desired way. In response, Ekon condescendingly sets about explaining in simple terms facts about enzymatic reactions to Elohor.

These examples represent paradigmatic cases of mansplaining —they display instances of harmful behaviour that is typically inflicted on a woman by an arrogant, overconfident man. More in particular, the phenomenon appears to belong to a conversational setting where an explainee, typically a woman, is harmed by her interlocutor, the 'mansplainer' —typically, a man. There are many ways in which conversational exchanges can go wrong, but the particular failure involved in cases of mansplaining is commonly taken to involve two distinguishing features: the fact that the explanation is in some way *redundant*, and the fact that the mansplainer perpetrates some kind of *harm* towards their interlocutor. Accounts of mansplaining (including the one we propose here), then, typically take their task to be that of flashing out these features —that is, to clarify the sense in which the explanation is redundant, and the cause of the harm involved in the interaction, as well as its nature.

The explanation offered by the mansplainer, for instance, is typically understood to be redundant in virtue of being targeted to someone who is an *expert* on a particular topic, or because it is *unrequested*. The harm, on the other hand, most commonly thought of as epistemic and/or conversational in nature, is often traced back to the mansplainer's *false belief* that they are in a position of epistemic privilege, as well as to the *sexist prejudice* that this belief is often taken to be grounded on. To a first approximation, then, we can say that paradigmatic cases of mansplaining are typically taken to involve a sexist male speaker who, on the basis of his false presumption of epistemic superiority, gives an unrequested explanation to his female interlocutor despite the fact that she is an expert on the subject matter.

In the course of this paper, we will look at some examples that we think should bring us to question whether the way in which these features are often flashed out (in the literature as well as in public discourse) is in fact essential to capture the phenomenon of mansplaining. Before we do that, however, we would like to spend some more time sharpening this rough picture by looking at recent discussion of the phenomenon by Rebecca Solnit (2014), Kate Manne (2020) and Casey Rebecca Johnson (2020).

We observed that, in paradigmatic cases of mansplaining, one of the features that is usually attributed to the explainee is that they are expert in the relevant topic. Attention to this point has first been drawn by Rebecca Solnit in her influential article *Men Explain Things to Me* (2014), where she recounts her first-hand experience, as a woman and a scholar, of men assumingly explaining to her facts about her research expertise. More recently, in her book *Entitled: How Male Privilege Hurts Women* (2020) Manne gives central space to this feature in her account of mansplaining as consisting of "a man presuming to 'explain' something incorrect(ly) to a *more expert* female speaker or set of speakers." (Manne 2020, 406, emphasis added). Indeed, this seems to capture one of the common features we tend to intuitively attribute to this phenomenon across the board —that the man's presumed explanation is ultimately infelicitous because the explanation offered is in some sense *redundant*. This condition purports to explain this intuition in the following way: the explanation is redundant because the explainee is an *expert* in what she's being told.

Another often-invoked way to cash out this intuition consists in noting that in paradigm cases of mansplaining the explanation is typically unrequested. Johnson (2020) draws attention to this point when she notes that "[t]he problem with the mansplainers is that the explanations they offer are often [...] not asked of them, and/or made to a woman who is more of an expert than they are" (2020, 6). According to Johnson's account of mansplaining as 'speech act-confusion', it is precisely the woman's expertise, together with the fact that the explanation wasn't requested, which renders the mansplaining speech-act, which trades on a generally benign conversational pattern, ultimately pernicious.

In addition to being redundant, we noted that a mansplainer's speech act is also, and perhaps most importantly, often taken to be *harmful*. But what is the nature of harm, and what's its cause? A natural response to this question suggests itself once we consider any paradigmatic case of mansplaining. Take for instance LAB, where the male interlocutor (Ekon) assumes that his colleague (Elohor) doesn't know the first thing about enzyme reactions simply because she's a woman. In this case, it seems natural to think that Ekon's explanation is harmful because it is based on his gendered prejudice against his female colleague. In fact, it is precisely this one of the central features picked up by Manne in her characterisation of mansplaining. According to Kate Manne, the sexist prejudice of the mansplainer consists precisely in the *entitlement* he takes himself to have to explain to his interlocutor basic facts she is very familiar with. What Manne has in mind is a particular kind of entitlement, that is:

> "entitlement of the epistemic variety, which relates to knowledge, beliefs, and the possession of information. In particular, I believe that mansplaining typically stems from an *unwarranted sense of entitlement* on the part of the mansplainer to occupy the

> conversational position of the knower by default: to be the one who dispenses information, offers corrections, and authoritatively issues explanations." (Manne 2020, 48)

Following Manne, then, we can identify two factors that make the mansplainer's speech act harmful: first, the fact that the mansplainer holds the *belief that they are epistemically superior* to their (female) interlocutor on the matter at hand. Naturally, since the commonsense view takes mansplaining to involve an expert explainee being told something she already knows by a less expert male explainer, this belief will necessarily be *false*. Just the false belief won't do though, for one may falsely believe to be epistemically superior to their interlocutor without at the same time thinking that this grants them any privilege. In addition, then, it must be the case that the mansplainer also takes their belief that they are superior to *entitle* them to dispense information, offer corrections and authoritatively issue explanations —that is, to occupy, in the conversational exchange, the role of the explainer/giver of knowledge.

Wrapping it all up, then, we can say that there are four main conditions that our commonsensical view of mansplaining sits on:

> *The Standard View (SV)*: A man (M) performs a speech act of mansplaining in an exchange with a woman (W) if and only if M explains that *p* to W and:
> 1. *Unrequestedness condition*: W does not request M's explanation
> 2. *Expertise condition*: W is an expert regarding *p* (at least more so than M).
> 3. *Doxastic condition*: M falsely believes *q*: "men are epistemically superior to women".
> 4. *Entitlement condition*: M takes *p* to entitle him to issue the explanation,

where the first two conditions (the 'unrequestedness' and expertise conditions) can be seen as offering attempts at fleshing out the sense in which we take mansplaining to be a redundant explanation, and the latter two (the doxastic and the entitlement conditions) aim to clarify the sense in which mansplaining is harmful.

The problem now, of course, is to see whether these attempts are successful. We don't think they are, and to see why consider the following case:

> GRUMPY GRANDPA Elena is a last year highschool student struggling with her physics assignment. She definitely cannot get her mind around the fact that the speed of light is related in such a way with mass and energy in Einstein's special relativity. Her grandfather Louis has only very standard knowledge about physics, but nonetheless has a strong (and false) belief that, since he is a man, he knows better than his granddaughter. Looking for help, she resolves to ask him why the factors are thus related. Louis starts explaining the problem to her employing very basic terms as though meant for a 3rd grader.

Notice that, despite having retained only the doxastic and entitlement condition (Louis thinks he's better at physics than his granddaughter because he is a man), this immediately strikes us as a case of mansplaining. Louis is not an expert in physics, but neither is Elena

—she only has knowledge of the subject that any average high school student would have. Also, we have that Elena explicitly asks his grandfather for help on that particular topic. Still, despite being *requested* and directed to a *non-expert*, the explanation appears redundant in the way in which mansplaining typically does. If this is so, then two key features of what is commonly suggested to be a correct understanding of the phenomenon of mansplaining —i.e., the unrequestedness and the expertise conditions— need not be part of its analysis.

Now consider this variation of the case:

> GRUMPY GRANDPA* Elena is a last year highschool student struggling with her physics assignment. She definitely cannot get her mind around the fact that the speed of light is related in such a way with mass and energy in Einstein's special relativity. Her grandfather Louis is a physics major and, for this reason, rightly thinks he's more knowledgeable than his granddaughter on the topic. Still, when Elena resolves to ask him why the factors are thus related, he responds by employing very basic terms as though meant for a 3rd grader.

This case retains none of the conditions of the SV. Here, Louis' sense of entitlement to occupy the position of the knower —if at all present— is not grounded on the *belief* that men are epistemic superior to women but, we can suppose, on the accurate judgement of his expertise in relation to that of his granddaughter (although he may well have an implicit sexist bias against women's ability to understand physics). Still, this is a clear case of mansplaining. For not only the explanation, like in the original case, strikes us as redundant —Louis, with little regard for his female interlocutor's question and true epistemic needs, condescendingly explains to her something in a way that doesn't give her the credit she deserves for her level of understanding of the subject matter. In addition, and precisely because he offers an over-simplistic and degrading explanation that doesn't take into account Elena's true epistemic needs, Louis' speech act also strikes us as harmful.

Now, if this observation is correct, these examples show that the SV is too narrow. Or, more precisely: that there are ways in which mansplaining can be redundant and harmful that do not involve the conditions set out by the Standard View. Of course, since it is based on a hypothetical counter-example, one could resist this argument quite simply by rejecting the intuition that drives it —i.e., that the GRUMPY GRANDPA examples are genuine cases of mansplaining. But we are not sure there is any simple way of doing this. After all, both cases *do* respect the general structure of our commonsensical understanding of mansplaining —for instance, in both cases, Louis' explanation has been shown to be both redundant and harmful. Rejecting the intuition without further argument would thus be suspiciously *ad hoc*. And naturally, it would be question-begging to reject the intuition on the grounds that the examples fail to meet the conditions that the SV attributed to mansplaining.

But perhaps there is another, more general, reason why we find resistance to these counter-examples to be ultimately infelicitous. The Standard View attempts to flesh out mansplaining's harmful and redundant explanation by looking at the (mostly internal) features of the *interlocutors* —i.e., their beliefs, their expertise, their sense of entitlement. The picture that emerges, then, is a picture of mansplaining as a phenomenon that regards *certain kinds of people* —people, that is, that possess or fail to possess certain characteristics: a

belief, a skill, an attitude. An obvious limitation of representations of phenomena that are 'rigid' in this sense, however, is that they impose unnecessary restrictions on the way the phenomenon can manifest itself, depending on the circumstances and the characters of the agents involved. In contrast to the SV, in the next section we sketch an alternative image of mansplaining that is anchored to the *dynamics of the conversation*, its norms and the broader context in which the conversation takes place rather than to the features of the interlocutors. According to this view, mansplaining should be understood as a form of conversational disrespect that constitutes a particular kind of testimonial injustice. More precisely, we propose the following working definition of mansplaining:

> *Mansplaining* A man (M) performs a speech act of mansplaining in an exchange with a woman (W) if and only if M explains *p* to W and
> 1) M's speech act is haughty.
> 2) M's speech act constitutes a testimonial injustice.

In the next sections, we zoom in on the haughtiness and the injustice condition respectively by looking into Tanesini's (2016) characterisation of haughtiness as an epistemic vice and Hooker's (2010) expansion on Fricker's (2007) account of testimonial injustice. The hope is that this will equip us with more flexible conceptual tools, capable of sidestepping these issues by offering a more comprehensive and useful account of mansplaining.

II. MANSPLAINING & HAUGHTINESS

The received view has it that mansplaining is a form of conversational disrespect that appears to be down to the fact that the speaker's explanation is somehow *redundant*. However, we have shown that it seems incorrect to cash this out in terms of it being *unrequested* or *unnecessary*. But then how is the explanation redundant? A simple-minded way to respond to this question is to note that the mansplainer says more than it appears to be required by the conversation —more than the interlocutor asks, or more than the interlocutor needs to be told. In what follows, we take this idea seriously, and look at a way in which mansplaining can be understood as involving a speech act that breaks a conversational norm (and is thus faulty) because the speaker explains *too much*.

In her 2016 paper *Calm Down, Dear*, Alessandra Tanesini defends an account of arrogance as an epistemic vicious behaviour that involves a violation of norms of natural conversation. The notion of intellectual arrogance, which involves one's perception of one's abilities, is distinguished from that of *superbia*, or haughtiness, which is an essentially interpersonal behaviour manifesting in social interactions. More precisely, according to Tanesini, haughtiness is a form of arrogance whereby someone, who takes themselves to be intellectually superior, behaves as though they have privileges that are denied to others.

According to Tanesini, two main ingredients make up haughtiness, one doxastic and one behavioural. The doxastic part regards what Tanesini refers to as a 'feeling of superiority' that accompanies haughty attitudes, and that is unpacked as one's belief that one is intellectually superior to others. Importantly, this belief need not be true or justified; one may wrongly think of themselves as intellectually superior to others in some task and still be arrogant. The behavioural part, on the other hand, has to do with what one does,

verbally —the violation of the conversational norms that regulate the interaction between speakers and hearers. This will manifest differently depending on the conversational circumstances, but may involve things like cutting off someone's speech, talking over someone or taking more than one's fair share in a verbal exchange (Tanesini 2016).

The two ingredients are closely connected. Typically, one arrogates conversational privileges to themselves on the basis of their belief that they are superior to others. For this reason, Tanesini thinks that, in addition, haughtiness requires the presumption that one's feeling of superiority warrants their haughty behaviour. But this strikes us as too strong a requirement. Suppose someone has been repeatedly told off for their arrogance throughout their life, and have finally come to learn that, even when they are intellectually superior to others, and they believe so, this doesn't justify them to interrupt others when they are speaking, talk over them and so on. Yet, they continue to do so regardless. Perhaps unsurprisingly, our intuition is that, irrespective of whether they presume their superiority justifies their violations, and precisely *because* of their violation, they *are* being haughty. After all, it would be odd to exculpate one's haughtiness purely on the grounds that one's beliefs are not in the right epistemic relation with each other. In fact, it would seem odd to exculpate one's haughtiness on the basis of one's internal belief status *altogether*. For one thing, the presence or absence of a mental state may depend on factors that are only tangential to what may seem the right judgement of their behaviour. Suppose your friend suffers from a very peculiar brain injury that prevents them from forming any belief about their own capacities and how they relate to others'. By stipulation, they have no belief (true or false, justified or unjustified) that they are intellectually superior to you. Still (say, out of habit, because that's what they've seen their father and other men do in their family or in their community) they regularly interrupt you, talk over you and arrogate to themselves conversational privileges they don't have. Again, the specific 'haughty belief' being present or not, it looks like their acts should rightly be considered haughty. If this is true, haughtiness need not involve any feeling of superiority either —at least not when this is understood doxastically.

Although we cannot offer a full defence for our preferences here, for these reasons in what follows we assume this externalist variant of Tanesini's account of haughtiness. According to this variant, what determines one's haughtiness depends on what one *does* conversationally, and not (always, also) on what one believes —i.e., more precisely, it depends on whether one's speech act violates norms of cooperative conversation and disrespects their interlocutor[51]. More precisely, we will look at the practice of assertion, and identify the type of haughtiness involved in cases of mansplaining as a violation of one of the norms that regulate it.

[51] An obvious consequence of this is that this variant doesn't distinguish between systematic vs one-off instances of haughtiness —that is, on this view, the 'bad guy' who arrogantly steps into the conversation because he feels entitled to interrupt his interlocutors is just as haughty as the 'good guy' who only happens to break a norm of conversation on one occasion. While this distinction *matters* (for blame attribution purposes, for instance) we don't think this is central to the analysis of haughtiness. For one, we believe that the fact that the externalist account we adopt here can be adapted to different scenarios (i.e., systematic vs one-off instances) is only a benefit for our view. Also, notice that we discuss a variant of this objection in the final section, when we consider mansplaining as a testimonial injustice. The same reasoning we propose there can be applied, *mutatis mutandi*, also to this case.

There is wide agreement in the literature that assertion is a heavily regulated practice[52]. For instance, we normally assume that, in making an assertion, a speaker ought to have the epistemic standing that is required to assert a certain thing, and that the hearer ought to recognise that they have it —or, at a minimum, acknowledge the communicative attempt for what the speaker intends it to be. Failing to do so can lead to a variety of negative conversational outcomes, and result in harm. The phenomenon of (illocutionary) silencing, for instance, can be understood as a conversational failure whereby a hearer doesn't take up the utterance of a speaker[53]. A norm that is of interest for our present scope asks that, when making an assertion, a speaker makes a commitment to provide appropriate evidence in support of the content of their assertion, when challenged. Following Tanesini (2016), we call this the *answerability norm* of conversation.

Our proposal is that, in cases of mansplaining, a speaker (the mansplainer) fails to provide appropriate evidence in support of their assertion. In some cases, this may be in response to an actual challenge. In the GRUMPY GRANDPA cases, for instance, we can imagine that, in the course of the conversation, Louis offers his explanation in response to Elena's (proper) challenge to his assertions. But the challenge needn't always be real, nor proper. In some cases (and perhaps most commonly), in the absence of a challenge, it may be the mansplainer himself who *fabricates* the challenge —in response to which he proceeds to offer his explanation. Think for instance of LAB or NASA. These cases can be read as involving a situation where a question or a simple assertion are treated by the mansplainer as challenges to a (more or less explicit) assertion, which the mansplainer proceeds to support with their evidence. In these cases, the answerability norm of assertion is violated because the evidence provided by the mansplainer, which is targeted to their fabricated challenge, fails to respond to the *actual* challenge. The explanation, that is, simply misses the target.

But notice that not all explanations that miss the target count as 'mansplanations'. In fact, there is a peculiar way in which mansplainers fabricate and respond to a challenge, when they do, that determines the particular flavour of the harm that is distinctive of cases of mansplaining. Consider two main ways in which the answerability norm of assertion can be violated, depending on whether one fails to give sufficient evidence for their assertion or gives too much. Call violations of the former kind violations of the answerability norm *by deficit* of explanation. These include cases where someone (imagine for instance an academic philosopher) supports their claim by appealing to their authority (in this case, their authority as a professor), rather than actual evidence. Violations of the latter kind, instead, are *by excess* of explanation —i.e., cases where one gives more evidence than is required by the content of the exchange. We argue that the phenomenon of mansplaining belongs to the latter class of violations: mansplaining occurs when a speaker (typically, a man), in response to a challenge (actual or fabricated), proceeds to supply more evidence than it would be appropriate given the circumstances of the exchange.

Note an advantage of this proposal compared to the SV: since we take mansplaining to consist of a conversational failure by *excess* of explanation, we can capture the intuitive sense in which a mansplainer's speech act appears to be *redundant*. Besides, since the

[52] See mainly Lackey (2007)
[53] See mainly Langton (1993), Langton and Hornsby (1998)

'explanatory excess' is measured with respect to a norm of natural conversation, it depends only contingently on facts concerning whether the explanation is requested, or concerning the hearer's expertise or the speaker's lack thereof. In other words, if our proposal is right, we can see that the SV's failure to identify paradigmatic cases of mansplaining is due to the fact that it takes contingent facts about the interlocutors (whether the explanation is requested, for instance) as constitutive of the phenomenon. In addition, our view provides a framework within which the relevance of these factors is brought to light. The dynamic of the interaction, the relationship between the interlocutors or the epistemic standing of the interlocutors themselves, to name just a few, for instance, are all facts that may contribute to determining the right amount of evidence an explanation ought to supply in different contexts. Depending on the circumstances, very long explanations may not amount to mansplaining whereas very short ones may. For instance, a teacher who gives a detailed answer to a question from one of their pupils won't typically amount to a case of mansplaining, even if they go to a great length in their explanation. Instead, as in the NASA case, when an explanation is unsolicited and it disregards the competence of the explainee, even a single word may be too much.

But does a long sermon from my teacher *never* amount to a form of mansplaining? And does a short comment from a presumptuous man *always* be a case of mansplaining? Surely, contextual factors complicate the picture. What's not clear, however, is the criterion by virtue of which we can determine whether the evidence provided by the mansplainer is inappropriate.

To see this, go back to the norm the mansplainer violates. Following Tanesini, we said that the answerability norm involves a commitment to (being in a position to) backing one's claims with appropriate evidence. But when is the evidence appropriate? Start with a clearer and more often discussed violation of this norm —i.e., the one by deficit of explanation. An example of a violation of the norm by deficit of explanations is an *authoritative* speech act —i.e., a speech act performed by a speaker where nothing other than the act itself is offered in support of its content. Since one's say-so isn't normally[54] good evidence to support the content one's assertion, authoritative speech acts violate the answerability norm by deficit of explanation —i.e., the authoritative speaker leaves their claim unsubstantiated[55].

If this is correct, we can think of violations of the answerability norm by excess in a similar way as speech acts that offer to one evidence that they already have. Since we aren't normally supposed to explain to people what they already know, when we do so we are violating the answerability norm by excess of explanation. Granted, there are very clear exceptions to this norm. When we try to find some common ground with our interlocutor(s), for instance, we may well venture to say something that they are knowledgeable about. Academic conferences are a clear example of a dynamic of this sort: talk after talk, the (expert) audience of, say, a philosophy conference is presented with notions with which they are familiar, and with problems about which they are expert. The

[54] There are exceptions (e.g., utterance containing indexicals referring to the speaker being the speaker, or being present, and so on).

[55] Sometimes, this will be due to an evaluative error: the speaker's overestimation of the epistemic relevance of their authority. In speaking authoritatively, one may take something that doesn't have epistemic relevance (their say-so) to have epistemic relevance. Note though that it need not be the case that one *believes* that their say so is a good reason.

fact that the common ground is sought with the intention of presenting new ideas may be a reason why we don't take some such speech acts to be in violation of the answerability norm. Another reason, which probably applies more broadly, may be that the norm is overridden by other, stronger, social obligations, like perhaps that of giving wide access to conference participation, for instance.

In the absence of these factors, however, explaining to someone something they already know is just bad, epistemically. For, if nothing else, by explaining to someone something they already know we are *treating them as epistemic inferiors*. To do so to someone who *isn't* our inferior is an epistemic wrong. But we may also wrong someone who *is* our epistemic inferior by treating them as such —if, for instance, we treat them as though they were less epistemically worthy than they actually are. In fact, this seems to be precisely what goes on in cases of explanatory excess —the explainer provides their interlocutor with information they already possess, and so they *fail to hold the explainee to the epistemic standard they deserve*. Ekon's explanation, to the extent that it provides to her colleague information she already possesses, treats her as epistemically less worthy than she is. The man's twit, which presumes to explain to an astronaut a simple physical phenomenon, fails to hold Dr Meir to the epistemic standard she deserves. Louis' too fails to hold his granddaughter to the epistemic standard she deserves, and for this reason we think he is mansplaining —despite the fact that he is more knowledgeable than her, and despite the fact that it is her who asks for his help.

Clearly, these violations would look even worse if we were told that they were caused by the mansplainers' failed judgement of their interlocutors' epistemic staninding, especially if unjustified. In addition to the disrespect involved in *treating* their interlocutor as less knowledgeable than they are, the mansplainers would be responsible for an evaluative failure *as well* —the mansplainers would fail to *estimate* the right epistemic standing of their interlocutors. Crucially, however, it is not the evaluative failure itself that generates the epistemic wrong. So long as one violates the relevant norms of cooperative conversation by excess of explanation, thereby treating their interlocutor as less epistemic worthy than they are, then their act *is* haughty. And this is so irrespectively of the speaker's judgement[56] of one's epistemic standing, that of one's interlocutor, and the relationship between the two.

To sum up then, we have argued that mansplaining involves a form of conversational disrespect consisting in a violation of the answerability norm of conversation by excess of explanation. Following Tanesini, we think that this violation should be subsumed under the category of 'haughty' behaviours. Departing slightly from Tanesini's account, we have proposed that, in cases of mansplaining, haughtiness should not be understood in relation with anything internal to the mansplainer's ken, but simply with what he does verbally. Following this line of thought, we have proposed that the epistemic failure caused by haughty behaviours amounts to a failure to treat the hearer according to the epistemic standard they deserve. In the next section, we take a closer look at the distinctively epistemic ways in which one can wrong another agent, and identify the variety of injustice that better capture the harm mansplainers inflict to their victims.

[56] We use 'judgement' rather than 'belief' here to remain neutral with respect to the 'doxastic force' of the cognitive commitment one may have in mind.

Conversation is a natural place for epistemic exchange, and much of our epistemic lives have their place *with* others. Haughtiness is an interpersonal behaviour that involves the breaking of norms of natural conversation —it is a form of disrespect that often severely impacts our epistemic relations with others. Following Alessandra Tanesini (2016) and Nicole Dular (2021), in this section we look at the connection between dysfunctional conversational patterns and linguistic and epistemic harm, and how this plays out in the particular case of the phenomenon of mansplaining.

According to Alessandra Tanesini, the haughty individual who breaks the norms of asserting has a profound impact on the psychology of their victims, and it reinforces systematic patterns of epistemic oppression. In particular, Tanesini argues that by interrupting others, arrogating to oneself the right to speak first or taking up more than their fair conversational share, haughty individuals sustain the ignorance of their interlocutors and foster their intellectual timidity and servility by locutionarily or illocutionarily *silencing* them.

There are many ways in which the haughty individual can *locutionary* silence their interlocutor. For instance, one may fail to respect conversational rules of turn-taking by interrupting the interlocutor, speaking over them in such a way that undermines their attempt to speak. In the specific case of assertion, perhaps excessively long speeches may result in silencing the hearer insofar as they take up all the time available, or discourage one to speak. Since these are cases of violation of the answerability norm by excess, they seem to be compatible with mansplaining, and indeed we can imagine circumstances in which the mansplainer, by taking more than their fair share of conversation, coerces the hearer into silence. Still, this doesn't seem to capture what's distinctive about the phenomenon. None of the paradigmatic cases we've looked at are cases where the explainee is prevented from speaking. The man's twit in NASA or Ekon's explanation in LAB don't strike us as necessarily *long* explanations, nor explanations that in any way impact on the explainee's ability to respond or somehow intervene (although they could have, under some description of these cases).

Another way in which haughty individuals silence their interlocutors is when they fail to recognise their assertion for what it is. The 'patronising interjection' to "calm down, dear", uttered by the then UK Prime Minister David Cameron in response to a criticism by Angela Eagle's (then Shadow Chief Secretary to the Treasury), for instance, is offered by Tanesini as an example of 'haughty illocutionary silencing'. In order for a communicative attempt to be successful, it must be recognised (or treated) as the speech act it is (intended to be). When the hearer fails to take up an assertion for what it is, no actual assertion has taken place, and the speaker is illocutionarily silenced. In the case of mansplaining, a man who systematically takes women's assertions as requests for clarification, or challenges, will see himself called on by a woman's attempted communication to respond to or magnanimously rectify them. As a form of illocutionary silencing, mansplaining would then consist in a hearer's disrespect towards the speaker's assertion.

This seems to be a plausible model to account for a great deal of the cases discussed: the man's twit, for instance, could be seen as motivated by his mistaken perception of Dr Meir's observation as a request for clarification. Similarly, Ekon could justify themselves by saying that it was Elohor herself who challenged him, thus betraying his failure to recognise her speech act for what it really was. But it won't do for all cases. In GRUMPY GRANDPA, for instance, we have a case where Louis's explanation is offered in response to a proper challenge —to which Louis, disregarding the epistemic standing of his granddaughter, nonetheless offers a haughty response. More poignantly, on the other hand, because the disrespect is hearer-based, understanding mansplaining as a form of illocutionary silencing fails to capture those interactions where the mansplainer doesn't occupy the role of hearer at all. To the extent that it seems natural to take mansplaining to involve a disrespect that one incur into in their role as *speakers*, then, the illocutionary silencing view appears to be inappropriate. In summary, although mansplaining may (often) be connected to forms of silencing, it seems mistaken to think of silencing as a necessary ingredient for mansplaining.

Another illuminating attempt at connecting mansplaining and epistemic harm has been made by Nicole Dular in her 2021 'Mansplaining as Epistemic Injustice'. According to Dular, mansplaining consists in a special case of testimonial injustice involving a *forceful* and *dysfunctional* subversion of the speaker-hearer roles in a conversation. Unlike standard cases of testimonial injustice, where prejudice on the part of the information-receiver prevents the knowledge-giver to get across a piece of information, in cases of mansplaining "the epistemic role of speaker or giver of knowledge is *not even made available to them*" (2021, 11). More precisely, this is because, according to Dular, the explainer claims for themselves a role —that of the speaker— that ought to be the explainee's. What makes the subversion *forceful* is the fact that the explainer resists the explainee's attempts to reclaim her rightful position in the conversational exchange[57]. What makes it *dysfunctional* (and thus unjust) is the difference in knowledge between the interlocutors —the position of the speaker/giver of knowledge, which should be occupied by the more knowledgeable interlocutor, is instead usurped by the least knowledgeable interlocutor, who should occupy the position of hearer/receiver of knowledge.

Dular's view has considerable merits. For instance, to the extent that it understands the subversion as owing to gendered prejudicial stereotypes, it provides a view that embeds the phenomenon in the wider context of identity oppression. In addition, and unlike other accounts of mansplaining, Dular's gives centrality to its normative dimension, which is, we think, rightly located at the level of the conversational exchange. At the same time, however, we disagree with Dular regarding the source and the nature of the norm that the mansplainer violates. For her, the norm concerns the rightful allocation of speaker/hearer roles, and is sourced in the relative epistemic standing of the interlocutors: the speaker should be the more knowledgeable one, and the less knowledgeable the hearer. For us, as

[57] Here Dular refers to some specific cases of mansplaining, including the one discussed by Solnit where, despite her repeated attempts, she failed to claim back the role of speaker in a conversation where her interlocutor, the mansplainer, was attempting to explain to her the content of her own book.

we've shown, the norm concerns the rightful support an asserter ought to give to their claim, and it pertains to the rules that govern the ordinary practice of assertion.

Here's some reasons why we think our view is superior to Dular's. First of all, we are not entirely convinced that there are general norms regulating who is supposed to occupy the role of a speaker or hearer in a determinate conversation. Surely, we are convinced that conversations are regulated by (highly contextual) norms of turn-taking, which establish, say, how much one is roughly entitled to occupy the position of the speaker. But it would be bizarre to think that there is a particular hearer/speaker positioning that two people ought to assume in a conversation purely based on their epistemic standing. First, this is because the fact that one is less knowledgeable does not seem to suggest one ought not occupy the role of the giver of knowledge. For instance, one may have overall less knowledge than another on a topic but still have knowledge that the other doesn't possess. Or one may have less knowledge but a greater understanding on those matters, or a greater ability to expose them in a way that will be better received. On our view, this is hardly surprising, since we take the disrespect involved in cases of mansplaining to be independent of the *reciprocal epistemic level* of the interlocutors. Instead, our view measures the failure against fixed norms of conversation: mansplaining is the result of a failure to (conversationally) hold one's interlocutor to *their own* epistemic standard. The epistemic standard to which one ought to be held in a conversational exchange is not set by one's epistemic standing *in relation to their interlocutor*, but is based on what one actually knows.

In summary then: because it focuses on mansplaining as an injustice perpetrated by the speaker, our account of mansplaining fares better than accounts that take the phenomenon to be related to forms of (locutionary or illocutionary) silencing. Because it understands mansplaining as a violation of the answerability norm of assertion by excess of explanation, on the other hand, it fares better than accounts that attempt to take the failure to be connected to norms regulating the conversational roles one is entitled to occupy. The question though remains: if not in the way proposed by Dular, how, if at all, is it possible to conceive of mansplaining as a kind of testimonial injustice? In the rest of the paper we propose a way in which mansplaining can (and, we think, should) be thought of as an epistemic injustice of a testimonial variety.

To begin with, recall that, according to our view, what is going on in cases of mansplaining is that one treats their interlocutor as *less epistemically worthy* than they actually are. Naturally, this analysis will strike those already familiar with the literature as analogous in important respects to Miranda Fricker classic definition of testimonial injustice. According to Fricker (2007), injustices of a testimonial variety involve a failure of uptake of information occuring when prejudice causes one to give a *deflated level of credibility* to their interlocutors, and thus wrong them in their capacity as epistemic agents —as knowers. One way in which epistemic injustice of this kind is brought about is when, say, racism prevents one from believing testimony coming from a member of a minority group —say, when a police officer, apparently disregarding a Black man's attempted speech act, proceeds to perform extra careful checks that she would have spared to a White man. The officer's prejudiced stereotype that, say, Black men are dangerous and prone to deception, caused her to attribute a deflated level of credibility to their testimony.

Mansplaining, at least in the way we want to understand it, bears clear similarities with cases of this sort: in both cases, one's epistemic agency is degraded owing to their true epistemic standing not being given the consideration they are owed. Moreover, in both cases, this degradation is due to prejudiced stereotypes, which support, in the case of mansplaining, the perpetrator's norm-breaching behaviour. Sometimes, the prejudice takes the form of an outright belief and leads to an evaluative failure —i.e., a failure in the *assessment* of the credibility that their interlocutor is owed. Sometimes, the failure may be purely behavioural, based on the perpetrator's habits, cognitive schemas and implicit bias, and manifests itself in the way they conversationally *treat* their interlocutor.

For this reason, we feel inclined to think that mansplaining should be considered as a form of testimonial injustice. However, we agree with Dular that mansplaining cannot be made to fit the category without some important clarifications. As she notices, in fact, there seems to be a fundamental discrepancy between mansplaining and standard cases of testimonial injustice, since victims of mansplaining, unlike victims of testimonial injustice, need not "face issues in properly functioning as speakers in testimonial exchanges (having their word believed)" (2021, 11). Testimonial injustice affects one's epistemic agency so to speak 'directly': prejudice *blocks* their attempt to transmit a piece of information. Although this will often be true of instances of mansplaining as well, there are also clear instances where the mansplainee's testimonial attempt *is* successful (e.g., in GRUMPY GRANDPA, Elena's speech act is taken up for what it really is —a request for clarification).

For Dular, this calls for a distinction between the wrong perpetrated in standard cases of testimonial injustice and the wrong that is peculiar to mansplaining. In cases of testimonial injustice, the wrong consists in being unjustly treated as unreliable, untrustworthy givers of knowledge. Due to mansplaining's dysfunctional inversion of roles, on the other hand, a victim of mansplaining will at best be recognised as a *receiver* of knowledge. For this reason, she argues, the wrong suffered by the victim of mansplaining is instead that of *not being recognised as a speaker or giver of knowledge*. And because being recognised as a speaker is necessary for being thought of as unreliable, the wrong suffered by victims of mansplaining is, according to Dular, and crucially, even more basic than the one identified by Fricker.

This is a fascinating proposal, and it manages to capture what we take to be a key aspect of mansplaining —namely, the fact that, unlike testimonial injustice (where the victim, being objectified, has its agency hindered or denied), mansplaining is an agency-*degrading* phenomenon (Dular 2021, 14). Relegated to knowledge-receivers, victims of mansplaining see their agency (and, with it, the role they are entitled to occupy in the epistemic community) as fundamentally impoverished.

We too agree with Dular that something like a degradation of epistemic agency is what lies at the heart of the phenomenon of mansplaining. For how we see it, however, the distinguishing feature of the wrong inflicted is not that victims of mansplaining are never recognised as speakers and information-providers. Elena's proper challenge to his grandfather's assertion is successful, and is taken up for what it is. Dr Meir is successful at passing on a piece of information, even though it is then improperly treated as a challenge to which the man feels entitled to offer a haughty response. More generally, the point, it seems to us, is not much that Elena and Dr Meir are necessarily *denied* their role as knowledge-givers, but rather, and more broadly, that they are *treated as subaltern* agents.

Let us clarify this. The mansplainer breaches a norm of natural conversation with the effect of 'putting the hearer in their (wrongful) place', epistemically. Sometimes, this is done by denying one the role of knowledge-giver. Other times, this is done by coercing someone into silence. Yet other times, this is done by failing to recognise the credibility they deserve. But, crucially, this can also be done without blocking the victim's attempted speech act (like in cases of silencing or standard testimonial injustices) or altogether denying them the role as a speaker (as per Dular's forceful and dysfunctional subversion of conversational roles). Elena and Dr Meir are not excluded from contributing to their community's informational exchange. Rather, their contribution is treated as coming from a *subaltern* position —i.e., from the position of an agent who may have information, but is not able to fully appreciate its relevance; who may be a reliable informant, but not quite a trustworthy inquirer; who may have something to say, but not quite interesting, or appropriate, or worth engaging with —someone, that is, who may well be capable of participating in the epistemic community, but not quite at the same level as a *mature* epistemic agent. Epistemic inferiors, who can always benefit from some extra teaching, mentoring, explaining.

Now, to see this consider Christopher Hookway's distinction between what he calls the 'informational' and 'participant' perspectives. According to Hookway (2010), there are two ways in which we can assess one's epistemic contribution to an epistemic community. From a narrow perspective (what he calls the *informational* perspective) agents are bearers of information, and their contribution consists of their ability to transfer, accumulate and store knowledge and information. The *global* perspective (for Hookway, *participant* perspective), on the other hand, in addition to the agent's ability to act as information bearer/provider, takes into account a constellation of epistemic attitudes that include things like: one's ability to raise doubts, ask questions, recognise and debate the relevance of a piece of information, inquire into a topic, show sensitivity to the importance of responsibility and trust in epistemic exchanges, treat other epistemic players with appropriate respect, and so on.

This constellation of attitudes makes up the very fabric of the epistemic community, and it also plays a role in supporting and supplementing the agent's ability to store, transmit and receive information. In Hookway's words: "[m]uch of our participation in epistemic activities does not involve claims to knowledge; and much of it does not even serve as a precursor to the offering of testimony. Often, little in the way of claims to knowledge may be involved at all" (2010, 156).

Corresponding to each perspective, there are two different ways in which one can suffer an epistemic injustice. From a narrow informational perspective, one can be harmed epistemically only when they are prevented from playing their role as information bearer/providers. This is the kind of injustice Fricker and Dular have in mind. For Fricker, testimonial injustice occurs when one's attempt to transmit knowledge is unfairly undermined. For Dular, mansplaining occurs when one is prevented from occupying the role of the speaker, and so be a knowledge-giver.

Adopting a global perspective —that is, the perspective where the agent is considered more broadly as a participant in an epistemic community, and not *just* as an information bearer/provider— a whole new range of ways in which one can be harmed epistemically makes itself visible. In her discussion of the epistemic loss inflicted on Cassandra by Apollo

as a punishment for her refusal, Cynthia Townley offers a precious analysis of the spectrum of epistemic privileges that one may lose when excluded from participation in an epistemic community:

> "What Cassandra has lost is the capacity to participate in epistemic relationships that require acknowledgment and reciprocity, those things essential to epistemic agents understood as mutually dependent members of an epistemic community. Cassandra has lost her place as a *mature member* of her epistemic community. She cannot defend her claims to know; she has no discretion with respect to disclosure; she cannot entrust another with what she knows; and she is excluded from cooperative interactions" (2003, 108). (my emphasis)

The mansplainer's haughty speech act breaches a conversational norm in such a way as to treat their interlocutor as *epistemically subaltern*. The way in which the victim of mansplaining will be epistemically wronged, however, will vary depending on the circumstance of the interaction. Sometimes, she will be wronged informationally —when, for instance, the mansplainer prevents her from occupying the position of the giver of knowledge. Most of the times, however, she will be wronged in the multiple and varied ways in which (short of the total exclusion suffered by Cassandra) one can be wronged in their capacity as a participant in an epistemic community —say, by being prevented from defending one's claim; by having one's standing in a network of trust degraded; by not having one's attempt to open a dialogue taken seriously; by having one's question dismissed —and in many other ways. Consider for instance this case:

> TIDE Martina is a first year college student who's just moved to the Shetlands from Italy to study marine biology at the University of Scalloway. This is a new programme, with few students, and she is finding it difficult to make friends. One evening, one of her coursemates invites her to a friend's house party. She dreads the idea of being in a house full of strangers, especially giving her lack of confidence with the language. But she knows this is a chance for her to meet new people, so she resolves to accept the invitation. At the party, she is very nervous and struggles to make conversation with her coursemate's friends. After some time, she manages to include herself in a little group —mostly graduate students from the same uni. They are talking about nothing in particular, making inside jokes and nasty remarks directed at each other, and she's finding it difficult to jump in. At some point, one of the older guys begins to tell a story about the extraordinary tide they had just a few days before, and how it kept the incoming ferries from the mainland in check for more than eight hours. Surprised by the story, and seeing a chance for her to join in the conversation, Martina intervenes: "wow, I had no idea tides could reach those heights around here!". Turning to his other friends, the student who told the story rebukes her: "Aye, here comes the Italian! You're lucky if you get a tide of a few centimetres down there" and then, back to her, he goes on by explaining to Martina very basic facts about tides, meridians and lunar attraction.

In this example, the graduate student arrogantly imparts Martina a lesson, treating her as an epistemic inferior. The explanation is clearly redundant, even though Martina is patently *not* an expert on the topic and despite the fact that, presumably, he *does* know more than her. Notice, moreover, that Martina's position as a speaker has not been called into question or revoked[58]. In fact, it's not Martina's position as a giver of knowledge that seems to be primarily at stake here, but rather her attempt at establishing a connection with the group of people, at being considered as a participant in the exchange —or simply, and more profanely, at being included in the conversation. In other words, the mansplainer's speech act in this case does not affect Martina as an information bearer/provider but, more subtly, as someone who is attempting to being recognised as a member of a network of trust —that is, in yet another of the many ways in which we can be epistemically insulted.

Here's then the sense in which we take mansplaining to be, at its core, an act of epistemic *degradation*. Not though, as suggested by Dular, because we think that mansplaining denies an epistemic informant their ability to provide information —or, at least, not only for that. Instead, we take mansplaining to be (epistemically) degrading because it is a haughty speech act whereby one treats their interlocutor as an epistemic subaltern —i.e., a lesser epistemic agent, someone who doesn't belong to the epistemic community in the same way as other mature epistemic agents.

Because our understanding of the epistemic wrong involved in cases of mansplaining is compatible with a wide spectrum of ways in which the mansplainer can treat their interlocutor as epistemic inferiors, our proposal is more inclusive than Dular's. And because we take a global (rather than informational) perspective, the kind of testimonial injustice we take to be involved in cases of mansplaining differs from Fricker and Dular's to the extent that we understand the injustice to be 'testimonial' not merely because it involves an informational exchange but, more broadly, because it pertains to the conversational ways in which we participate in our epistemic environments.

CODA

We have proposed that mansplaining should be understood as a haughty speech act whereby a speaker treats their interlocutor as less epistemically worthy than they are. For this reason, we have suggested that mansplaining shares very crucial features with Fricker's notion of testimonial injustice. Indeed, as we have noted, we are not the first to have attempted an association between mansplaining and testimonial injustice. It just seems a very natural move to make. The notion of epistemic injustice, in fact, has the advantage of recognising the role of mansplaining in upholding systems of coloniality and oppression. By assigning a position of epistemic subalternity to epistemically marginalised individuals, mansplaining serves to reinforce a system that already systematically excludes them from

[58] If she feels intimidated by the guy's response, she may in fact end up being discouraged from attempting to share a piece of information with the group in the future. But we don't think this should entitle us to say that she's been refused to occupy the role of speaker —or at least, not any more than it entitles us to say that she's also been denied the role of the hearer. The mansplaining act functioned, in this case, as a way for the group to close ranks and take advantage of the social weakness of a newbie to strengthen existing social connections. To take the epistemic impact of this behaviour on Martina to consist solely in the fact that she has not been recognised as an information provider seems to us too reductive.

full participation in the epistemic community. In contrast with extant attempts, however, our proposal has made space for recognising a much wider range of ways in which mansplaining can be used as a tool for epistemic marginalisation. Adopting Hookway's participant perspective, we have proposed that the mansplainer's degrading treatment of their interlocutor can impact them in the most diverse ways: as information bearers and providers as well as in their more modest attempts at participating in our shared epistemic community.

Chapter four

Groups Believe In Many Ways
(and they're all fine)[59]

ABSTRACT

In recent years, a novel idea has taken root among epistemologists which understands knowledge as an essentially collective endeavour. This has brought to light a host of stimulating new epistemological questions. Here I focus on a specific issue concerning collective mental states. More exactly, the question I will be asking is the following: how do collective entities form their beliefs? Monist accounts of group belief are split into two main camps: deflationary summativists say that the doxastic status of a group simply reduces to that of its members, while strong inflationary non-summativists deny this. Pluralists, on the other hand, renounce to offer a unitary theory of group belief, and argue that sometimes groups form beliefs in summative ways, and sometimes in non-summative ways. In this paper, I argue that a functionalist analysis of group belief makes available a distinctively weak version of inflationism. In this way, my view manages to strike a fine balance between all the contenders in the debate: like pluralist views, my view has it that group belief can be formed in both summative *and* non-summative ways. Because it offers an *analysis* of group belief, however, it also retains the advantage of monist views of being capable of offering a unitary theory of group minds.

INTRODUCTION

Group belief attributions are pervasive in ordinary language. We say that juries believe that the defendant is guilty, or that gastroenterologists know that ulcers are caused by bacterial infection. Also, we naturally take beliefs to be produced in different ways by different types of groups. For instance, beliefs attributed to companies or organisations typically correspond to the views of their operative members, and political institutions often combine their members' beliefs according to some aggregation procedure.

[59] Title is inspired by, and a (friendly) challenge to, Pettigrew's (forthcoming)

Accounting for the variety of ways in which groups form beliefs *matters*. It matters because it gives us the key to useful predictions and explanations for the behaviour of important social actors like states or corporations. And it matters for holding them responsible for their actions, for attributing culpability and blame. We think it matters, for instance, to understand the politics of the affair involving the multinational corporation BP and the environmental disaster caused by the company's oil spill in the Gulf of Mexico in 2010. We think it matters to know how to attribute blame and responsibility to financial experts and economists' failure to predict and subsequently navigate the 2008 financial crisis. To do these things, we need to know whether BP *believed* that their safety protocols were up to standards, and whether financial experts *believed* that the markets were functioning properly. That is: to do so, we need to know what it is for a group to hold a belief. More exactly, we need to know what it is for different kinds of groups to hold a belief. If we can account for beliefs formed only in certain kinds of groups, our ability to attribute responsibility and blame is too narrow. What we want is a theory of group belief that helps us explain the behaviour and the dynamics of as wide a variety of groups as possible.

Yet, the variety of ways in which groups can form their beliefs seems to get in the way of the formulation of a single overarching account of group belief. A collection of experts and the members of a multinational organisation, for instance, are very different kinds of groups, and they believe in very different ways. In these cases, the very prospect of a unified theory may seem in principle misguided, or simply impossible to achieve.

The tension between the need to account for this diversity and that of doing so within a unitary framework has generated a fracture in the literature on group belief. Pluralists[60], who give central importance to capturing the diversity of group belief ascriptions, feel compelled, because of this, to renounce to offer a unitary *analysis* of group belief. Monists[61], on the other hand, give priority to this latter concern. Deflationary summativism (Quinton 1976), for instance, argues that the belief of a group *always* reduces to the individual beliefs of its members. This seems to offer a good explanation for beliefs hosted by loose, unstructured groups (e.g., the public opinion, the beliefs of cat lovers, etc.). But sometimes beliefs of organised groups are independent of their members' doxastic state. A jury may pronounce a verdict that differs from the individual jurors' opinion because of their bias, or because they had access to evidence inadmissible in a court of law. Centring their case on this latter sort of cases, strong non-summative inflationists hold that individual beliefs *never*[62] constitute group beliefs, and that they are instead a product of the members' joint

[60] In particular, Christian List (2014) Richard Pettigrew (forthcoming)

[61] Anthony Quinton (1976), Margaret Gilbert (1987), Tuomela (2007), Bird (2010), List and Pettit (2011)

[62] This may sound a bit too strong. Take for instance when the group members vote on an issue and they vote according to their individual beliefs, and the majority view gets accepted. Isn't this a case where an inflationary view would say that the group belief depends also in part on the beliefs of the individual members? I don't think so. The fact that they do personally believe what the group ends up believing is just a mere coincidence. Here I am not saying that group belief and individual belief never coincide, but that the former is never determined by the latter. This may be true even if, from time to time, the two coincide.

acceptance (Gilbert 1987), their organic collaboration (Bird 2010[63]) or their procedural aggregation (List and Pettit 2011, Pettit 2003).

Deflationary summativism and strong inflationary non-summativism ground their proposed analysis of group belief on the characteristics of some elected group-type[64]. Disagreement among monist accounts of group belief often boils down to disagreement over which type of group hosts *genuine* group beliefs[65]. For this reason, monist accounts end up having to discount the full variety of ways in which different group-types can form their beliefs.

My goal in this paper is to offer a unitary *analysis* of group belief in functionalist terms that is at the same time capable of vindicating the full spectrum of ordinary belief ascriptions to groups. Because it argues that there is a single unitary notion of group belief, defined in functionalist terms, my view belongs to the monist camp. In contrast to other monist views, however, and in natural continuity with functionalist analyses more generally, especially from within the pluralist camp, my view has it that group belief is multiply realisable —i.e., that it can be realised in different ways by different types of groups. More precisely, in contrast to both deflationism and strong inflationism, my view has it that *sometimes* groups form beliefs in a summative way and *sometimes* in a non-summative way. Since my view departs from deflationism, it falls within the inflationist camp. Since it differs from strong inflationism in that it allows for groups to hold beliefs in a deflationary way, it is a distinctively weak version of inflationism.

To be clear, then, the stakes are high: if I am right, the proposed view succeeds where others have failed in striking a very fine balance between all contenders in the debate, healing the fracture between monists and pluralists accounts of group belief while at the same time resolving the deflationism/inflationism dispute.

The argument develops as follows: in §I I situate my view within the monists' debate, starting from a brief overview of the main views in the literature (§I.1, §I.2), and closing with a discussion of their limits (§I.3). Motivated by the failure of extant monist views, in §II I put forward a hypothesis —what I call the Hypothesis of Multiple Realisability (HMR)— bring attention to two competing interpretations of it —one pluralist and the other monist— (§2.1) and show how a functionalist view can be borne out of the monist reading (§2.2). Finally (§3.1 to §3.4), I confront the main objections from pluralists and monists against the feasibility of this project.

I. MONIST ACCOUNTS OF GROUP BELIEF

Ordinary attributions of mental states extend beyond individual agents. Although the sciences of the mind have typically privileged the study of individuals, recent developments have challenged this tendency. In the biological sciences, psychological traits are

[63] Thinking of groups in an organismic way can also be found in Wray (2007) and List and Pettit (2011)

[64] Note that the line between monists and pluralists is blurred: some monists have defended pluralism —see List (with Pettit) 2011 vs List 2014.

[65] See for instance how, although Gilbert, Tuomela and Bird do acknowledge the existence of summative beliefs, neither of them think that a theory of group belief should be concerned with them.

systematically ascribed to aggregates of individuals: collective attributions have become fundamental in understanding the behaviour of hives, and collective memory appears to be an important feature of ant colonies (Wilson 2005). Studies on the psychology of crowds have brought the phenomenon of collective conations (e.g., collective hopes, fears and desires) to the attention of social scientists (Richer et al. 1987), and problems in decision-making theory have inspired research on collective reasoning (Pettit 2003).

In recent years, this trend has stimulated a surge of interest in the epistemology of groups, and in particular in the phenomena of group belief and group justification[66]. Since its inception, and under various descriptions[67], this new research project has consisted, for the most part, in extending to groups the tools developed by epistemological theorisation at the individual level (Longino 2022). In what follows, I will be concerned with one particular way in which traditional individual epistemology has been brought to bear onto theorisation about groups —that is, regarding the kind of method of inquiry that has been privileged. Traditionally, 21st century Western Anglophone epistemological theorisation has been preoccupied by the pursuit of a *dismantling analysis* of the concept of knowledge —that is, an analysis that proceeds by decomposing the analysandum into explanatorily prior elements. In a similar spirit, current attempts at understanding group belief take the notion of 'group belief' to be the target of a conceptual analysis very similar, in its dismantling character, to the one traditionally subjected to the concept of knowledge.

This tradition, at least with regard to its focus on dismantling analyses, will be the main target of my criticism in this first section. To do so, I will first start by providing a quick overview of the most popular monist accounts of group belief, and bring to light two aspirations common to these views —what I call their *aspiration to universality* and *aspiration to particularity*. Although I believe these are legitimate aspirations, my criticism will take shape by showing how the way in which extant accounts attempt to achieve them generates a problematic tension.

I.1 Complementary Accounts: Deflationism and Strong Inflationism

One of the most popular and influential accounts of group belief to date is Margaret Gilbert's collective commitment model. Gilbert's focus is on the particular kind of groups (juries, committees, reading groups, cabinets and so on) whose members are held together by what Emile Durkheim (1982 [1895]) calls 'mechanic collaboration'. Mechanic collaboration, according to Durkheim, is a principle of composition for groups whereby a collection of people are held together by the norms or rules that they decide (more or less consciously, depending on the details of the account) to comply with. In her seminal paper *Modelling Group Belief* (1987), Gilbert provides a strongly inflationist account of group belief of this sort in terms of *joint acceptance*, a group state resulting from the explicit and voluntary commitment made by each member of a group *qua* group member. Significantly, this state is irreducible to the sum of the mental states of its members, in that the group members' commitment to a proposition *P* (i.e., their tendency to explicitly and voluntarily act as if *P*

[66] See for instance Gilbert (1987) Bird (2010), List & Pettit (2011), Goldman (2014), Schmitt (1994), Lackey (2016, 2018, 2020).

[67] E.g., 'social social epistemology' according to Bird (2010), although it has become more commonly known (also and in part due to, Goldman) as 'group epistemology' (and sometimes, not uncontroversially, 'the epistemology of collective agents').

were true in their capacity as group members)[68] must necessarily be collective —that is, conditional upon the commitment of every other member. More formally, Gilbert's joint acceptance account is usually formulated as follows:

> *Joint Acceptance Account* (JAA): A group G believes that *p* iff the members of G jointly accept that *p*[69].

This model is typically contrasted with another dismantling analysis of group mental states proposed by Anthony Quinton. According to Quinton's ontological individualism (Quinton 1975) mental state attributions to social objects in general must be understood as merely metaphorical. This is the view that collective beliefs are constituted by the 'sum' of the beliefs of the individuals involved, like a wood is constituted by its trees. Hence the name summativism (or deflationism), in opposition to which Gilbert's joint acceptance account is usually referred to as non-summativist (or inflationist). In its simplest form, summativism is typically defined as follows:

> *Simple Summativism* (SS): A group G believes that *p* iff some or all group members believe that *p*.[70]

Crucially, both models provide *dismantling* analyses of group belief in the sense that their aim is to decompose a complex structure into simpler elements that are explanatorily prior to the analysandum —be these the group member's individual beliefs to which the collective one reduces to, or other mental states (i.e., their joint commitments). In this, both accounts closely follow the traditional project of individual epistemology, that is 'dismantling' in this sense because its aim is to decompose a complex structure (e.g., knowledge) into its explanatorily prior elements (justification, belief, truth).

A significant shift away from dismantling analyses in the epistemology of groups is due to Christian List and Philip Pettit's procedural account (Pettit 2003, List 2005, List & Pettit 2011) and Alexander Bird's distributed model (2010)[71]. Like Gilbert's JAA, both the procedural account and the distributed model are strongly inflationist accounts of group

[68] See Tuomela (2007) for a formulation of *implicit* joint attitudes —that is, attitudes that aren't conditional upon, but merely *assume* the commitment of the other members. Another very influential account of collective intentionality is Bratman's (1984, 1999). Differently from Tuomela and Gilbert, whose focus is, respectively, on the mode and subject of the collective belief, Bratman's is on its content ("I intend that we J").

[69] A more refined formulation of JAA is offered by Tuomela (1992, 1993 and 1995) which requires that only the operative members of a group jointly accept that *p*.

[70] This is a general definition of summativism that is often adopted (Lackey 2020, Pettigrew forthcoming, Bird 2010, 2019). However, notice that there are a variety of ways in which this general definition can be specified (Gilbert 1994 gives a very comprehensive overview).

[71] In some sense, Lackey's Group Agent Account could be taken to be an organismic view of group belief. Since I don't take the differences between organismic views of group belief to be relevant for the purpose of this paper, for reasons of space I will omit Lackey's view here. Another defender of a distributive view similar to Bird is Palermos (2016).

beliefs[72]. Unlike the JAA, however, they take other principles of composition as key to explaining the way beliefs are formed at the collective level.

List and Pettit's Procedural Account (PA), for instance, focuses on particular kinds of group agents, like committees and cabinets, where beliefs arise downstream of an intentional process of coordination between the group members. More exactly, according to the PA, we should understand this coordination as consisting in *procedures of judgement aggregation*. By the light of the PA, one salient aggregation procedure that doesn't lead to the group holding inconsistent attitudes is one where to be aggregated are not the votes on the matter itself, but on the *premises* on which the decision is made. On this view, then, group belief corresponds to the majoritarian aggregation of the judgements expressed by the group members on each premise.

To see this more clearly, take a sport's committee whose members gather to decide what football player should be awarded the title of 'player of the year'. Suppose there are three committee members (A, B, and C), and that the decision to award X 'player of the year' depends on their votes on three premises: (i) whether X is the best scorer, (ii) whether X is the best dribbler, and (iii) whether X has the best overall conduct. After voting, it emerges that:

	Best scorer?	Best dribbler?	Best fairplay?	Best player?
A	Yes	Y	N	N
B	Yes	N	Y	N
C	No	Y	N	N

Now, suppose that X is the best player for each committee member only if they think X is the best scorer, best dribbler *and* most respectful player but that, for each category (i.e., Best Scorer, Best Dribbler and Best Fairplay), only a majority of votes is necessary to determine the aggregate result (e.g., the aggregate result for the category "Best Scorer" is "Yes" since the majority of committee members have voted "Yes"). In this case no individual member of the committee believes that X is the best player (since no one thinks X satisfies all three desiderata). Crucially, however, because the majority of committee

[72] Admittedly, whether List and Pettit's view counts as strongly inflationist depends on the kind of aggregation procedure considered —e.g., it would be strongly inflationist if it allowed only a premise-based procedure of aggregation, but deflationist if it included other procedures. In this sense, their view can be seen as defending (or at least being compatible with) a form of pluralism about group belief (in fact, I take this to be the case for List (2014) in particular, where he explicitly defends a pluralist view). However, because they believe that reductive analyses of group belief characteristically fail to account for the distinctive kind of group rationality that they want to account for, in their (2011) they ultimately explicitly endorse a form of non-summative inflationism.

members believe that X is the best on each count (i.e., best scorer, best dribbler and best fairplay) *the group,* as a whole, *does*[73].

For this reason, proponents of the PA suggest that the neat discontinuity between the belief of the group members and that of the group in cases of premise-based aggregation procedures "make[s] vivid the sense in which a social integrate is an intentional subject that is distinct from its members" (Pettit 2003, p. 184).

Another attempt to introduce functionalist considerations in a theory of group belief is offered by Alexander Bird's distributed model (DM)[74]. Inspired by Hutchins' *Cognition In The Wild* (1995), Bird's DM takes group belief to be the product of the organic collaboration of group members. According to Bird, group belief is specific to *organic* groups —a particular type of groups identified by the principle of division of labour, whose "key feature […] is that individuals and organisations depend on others who have different skills and capacities" (Bird 2010, 37)[75]. Candidate social believers, according to this view, are only *organic* groups that possess the following properties:

> (i) they have characteristic outputs that are propositional in nature (*propositionality*);
> (ii) they have characteristic mechanisms whose function is to ensure or promote the chances that the outputs in (i) are true (*truth-filtering*);
> (iii) the outputs in (i) are the inputs for (a) social actions or for (b) social cognitive structures (including the very same structure [the structure that produces the output]) (*function of outputs*). (Bird 2010, pp. 42–43)[76].

Although functionalist in spirit, both the procedural account and the distributed model don't deviate significantly from competing dismantling analyses. Proponents of the DM define group belief by conjoining a functionalist analysis of belief at the individual level with a Durkheimian conception of organic entities, such that:

> *Distributed Model* (DM): A group G believes that *P* iff (a) *P* is the product of the members' organic labour, and (b) G is such that it respects the fundamental characteristics of organic social entities (i.e., conditions (i) to (iii)).

Similarly, List and Petitt's view offers a functional analysis of group belief in the limited sense that the belief is taken to be a product of a particular aggregation procedure in a particular kind of group:

---

[73] There are counterexamples to this. For a more detailed account of the ways in which choice of basis for the group belief see, other than List and Pettit (2011), also Pettigrew (forthcoming) and Lackey (2020).

[74] In recent years, this approach has reached a relatively wide consensus, and is now defended, among others, by Bird (2010, 2019), Wilson (2005), Giere (2002), Hutchins (1995) and Magnus (2007).

[75] The distinction between 'organic' versus 'mechanical solidarity' that, following Durkheim, Bird uses to distinguish Gilbert's and Tuomela's joint-acceptance-style accounts from his own, was originally brought to the debate in Wray's (2007).

[76] Note that membership in a Durckheimian group is not *fixed*. Who counts as a member of the group depends on who plays a role in the production of the belief according to the standards specified ((i), (ii), (iii) above).

> *Procedural Account* (PA): A group G believes that *P* iff (a) *P* is the product of the group members' system of deliberation according to a premise-based aggregation procedure and (b) G is an intentional agent.

The resulting analysis of group belief is thus still dismantling, although in a minimal sense. It is *dismantling* in structure, because it breaks down the complex phenomenon of group belief into its more basic components (i.e., the belief and the *organic* structure that hosts it, or the individual judgements and way in which they are aggregated); however, it is dismantling *in a minimal sense*, in that it is not only the beliefs of the individuals (like in other summative accounts), or their mental states more in general (like in the JAA), that figure in the analysis, but also other elements (like the particular organic structure of the group, or some aggregation procedure).

I.2 Between Universality and Particularity

Existing monist accounts of group belief seem to share two common features: (1) the fact that they propose a conceptual analysis of group belief that is dismantling in character, and (2) that they do so by appealing to the belief forming methods that are particular to some (one) elected group-type —i.e., mechanic or organic groups for Gilbert's JAA and Bird's DM, and more or less organised aggregates for List and Pettit's PA and Quinton's SS. A useful way to think about these features is as revealing monists' attempt at achieving two common aspirations: an *aspiration to universality* (i.e., the desire to formulate a fully general and comprehensive analysis of the concept of group belief) and an *aspiration to particularity* (i.e., the desire to vindicate the particular mechanisms of belief formation adopted by some group-types). In other words, it is because these monist accounts aspire to offer a fully general understanding of group belief applying indiscriminately to all instances of (genuine) group belief that they propose dismantling analyses; and it is because they aspire to be faithful to the internal workings of the social groups in which the beliefs are formed that they attempt to do so by drawing inspiration from the patterns of belief composition of actual groups.

I am convinced that the aspirations of current monist accounts of group belief are legitimate aspirations for anyone who wishes to offer a complete account of group belief. It is legitimate, I think, to aspire to offer a fully general analysis of group belief; and it is legitimate to want to vindicate the specific ways in which different types of groups form their beliefs. However, I also think that the way in which existing monist accounts attempt to achieve their aspirations generates a problematic tension. More precisely, this tension is the result of their attempt to achieve their aspiration of universality *by drawing inspiration from the workings of one elected group type* (say, only mechanic groups, or organic ones, or only aggregates, etc.). In other words, the attempt of each monist view to provide an analysis of group belief is in tension with their pretence of doing so by making leverage on type-specific features of their preferred group-type.

The reason why I think this tension is problematic is clear: a view that attempts to ground a fully general analysis on the characteristics that are peculiar to a particular group-type risks impartially reflecting these peculiarities in its analysis —that is, it risks being *idiosyncratic*. If that is true, we should expect monist analyses of group belief to end up

being undesirably *chauvinistic*, and run into trouble in trying to accommodate the variety of ways we typically take groups to be able to form beliefs. In the next section, I show how this problem, which I call *belief under-generation*, indeed affects indiscriminately both deflationist and strong inflationary accounts.

I.3 Idiosyncratic Monism

Deflationary summativism is typically[77] taken to suffer from belief under-generation to the extent that it fails to accommodate beliefs formed by collective entities such as mechanic and organic groups (juries, research teams, assemblies etc.), which are independent from their members' mental states. Take for instance the case, discussed by Hutchins (1995), of the U.S.S. Palau's crew members' collective effort to bring the ship into port. Hutchins argues that, collectively, the crew members have much knowledge essential to sailing the ship —such as, say, the ship's location and speed. Still, it is possible to imagine that no one at any one moment possesses any such information. Each crew member fulfils a particular role, and the relevant belief (and knowledge) is generated at the group level by virtue of their coordinated work.

Another case often discussed[78] is that of beliefs formed via deliberation processes in collective entities such as juries. Consider for instance the case of a jury whose members have access to conclusive evidence that the defendant actually committed the crime for which they are being prosecuted. This evidence, however, is not admissible in a court of law. Based on this evidence, each juror individually believes that the defendant is guilty. However, because the admissible evidence available to them is insufficient for conviction, they pronounce an innocence verdict. In this case, we standardly say that the jury believes that the defendant is innocent, and we do so irrespectively of the personal beliefs of the jurors[79].

Both examples present cases where beliefs formed at the collective level cannot be accounted for simply by looking at the doxastic status of the individual agents making up the group. If this is so, the thought goes, deflationary summativism fails to offer necessary conditions for group belief.

The observation that there are beliefs formed at the group level that 'float freely' from the mental states of the group members, on the other hand, gives *prima facie* support to inflationary non-summativist views. After all, at a minimum, inflationism can be understood as the negative claim that group beliefs do not reduce to a mere 'sum' of individual beliefs. Current inflationary views, however, commit to the stronger claim that group beliefs *never* depend on their member's doxastic states. For this reason, a very similar belief under-generation worry can be seen to arise for this strong version of inflationary non-summativism.

---

[77] In their discussion of deflationary summativism, for instance, both Gilbert (1987) and Bird (2010) quickly dismiss SS on account of its inability to accommodate beliefs generated in more complex groups —that is, precisely those groups whose beliefs they both take to be paradigmatic cases of group belief.

[78] Gilbert (1987), Bird (2010).

[79] Notice that this is not only confirmed by our intuition. In the American legal system, a verdict is explicitly referred to as the *belief* of the jury (Ho 2008, chapter 4).

Consider for instance the case of a newbie making their first appearance in a high school class where everyone already knows each other. Based on their appearance, the classmates form a prejudiced belief about them. *Social stigma* typically affects individuals or minorities who are socially discriminated against in virtue of more or less perceivable characteristics that distinguish them from the rest of the group (Goffman 1963). In cases of social stigmatisation, the group is said to possess a certain set of prejudiced beliefs directed against a minority, which in turn determines the group's discriminatory behaviour against them, or against other groups within or outside the group itself (Major & Laurie 2005, Smith 2012). These beliefs have been observed to inform and dictate distinctively collective discriminatory attitudes brought about automatically, without collective agreements and independently of the member's role in the group. If this is so —if there can be group beliefs formed despite what the group members jointly accept, their organic collaboration within the group, or some accepted belief aggregation procedure— then, the objection goes, what is jointly accepted or organically produced sometimes comes apart from what is collectively believed.

In summary then, the idiosyncrasies of extant monist accounts seem to generate a problem —what I have called the problem of belief under-generation. Perhaps there are ways in which these views can repel these objections, in one or the other of its incarnations. If I am right, however, there is no trivial way for them to resist the pressure of this problem more in general, since the idiosyncrasy that generates this problem stems from the unresolved tension between their conflicting aspirations, which lies at the very heart of these views.

In order to overcome this problem, extant monist accounts would have to legitimate their idiosyncrasies, and provide detailed stories explaining what makes all and only the beliefs formed according to their preferred belief formation practices, and not others, count as *genuine* group beliefs. In section III.2, I consider attempts to come up with some such stories. Before I do that, however, I want to turn my attention to another (and to my mind, much more natural) strategy for resolving this tension. In a nutshell, instead of reducing the number of genuine group beliefs (in order to match those beliefs that can be explained by appeal to a group-type specific belief forming practice), this strategy has it that *we should just accept that there are multiple ways in which groups can form their beliefs*.

II. RADICAL FUNCTIONALISM

II.1 The Hypothesis of Multiple Realisability

Current monist accounts of group belief unjustifiably discount the variety of ways in which we ordinarily ascribe beliefs to collective entities. For this reason, I argued, they are undesirably *chauvinistic*. Despite their failure to provide necessary conditions for group belief, however, it could be argued that they still offer plausible explanations of how different group types *sometimes* form their beliefs. I do not take this to be a very controversial claim. It is true that, along the lines of a deflationary summativist framework, doctors' collective belief that ulcers are caused by bacterial infection can fruitfully be cashed out in terms of a belief that is shared by most doctors. For other types of groups, it is true that there is no function that has individual beliefs as inputs and that can give us the desired outputs when it comes to assessing beliefs hosted by mechanic groups. Whether

Philip Morris (a large group whose members are connected together by a complex network of social and financial relations) believes that tobacco is harmful does not depend on the personal opinion of the entirety of its employers and employees. When we consider a belief held by large structured groups like this one, it is natural to account for it as arising from some sort of joint commitment —that is, the product of a process of deliberation that involves a collective agreement on a particular matter. Similarly, it appears natural to understand beliefs held by organic groups (such as researchers working together in a laboratory) to depend on the organic collaboration between the group members. This is how organic groups believe, which is different from how mechanic groups, collections, aggregates, statistical groups and categories believe[80].

More generally, we could summarise these intuitive observations by saying that simple summativism, joint acceptance, organic collaboration and procedural aggregation all describe group type-specific ways in which beliefs can be realised at the collective level. Call this the Hypothesis of Multiple Realisability:

(HMR): Group belief can be realised in different ways in multiple group-types.

Note that, thus formulated, the HMR lends itself to two interpretations: it can be read as suggesting either (1) that there are many different types or notions of group belief, each of which is realised in different ways in different types of groups; or (2) that there is a single kind of group belief, and that this is multiply realised in all the various ways in which beliefs are formed in different group-types[81].

Call these two interpretations the *Pluralist* and the *Monist* reading of the HMR. Each of them offers its own way out of the conflict between the ways in which monist accounts have attempted to achieve their aspirations. Reading (1) resolves the tension by giving up the aspiration to universality *altogether*. Once the pretence of offering an *analysis* of group belief is abandoned, the group-type specificity of current monist accounts ceases to be problematic, and can be appreciated as reflecting the variety of ways in which groups form their beliefs. For those who favour this reading, the problem with existing accounts of group belief is their very *monism*.

In fact, it is precisely this pluralist intuition that inspires Christian List's appeal to carefully disambiguate between different group belief formation practices. According to List (2014), the ambiguity occurs between three main ways in which patterns of individual attitudes and actions can give rise to group-level belief: by aggregation (e.g., beliefs obtained by surveys), by common belief (e.g., beliefs everyone is aware that everyone else holds) and corporate belief (i.e., beliefs formed by groups that are themselves agents in

[80] Notice that this is compatible with the observations that the same group might form beliefs in different ways. This is because the same collection of individuals can be grouped in different ways —e.g., A, B and C may constitute a group of friends, or of people who share the property of being brown-haired, or a sports committee, and so on. When I say that it is natural to take different 'groups' to believe in different ways, then, I take 'group' to include the individual members *and* the principle of composition under which we are considering them. I discuss this in greater detail in 4.4 below. (I wish to thank an anonymous referee for pressing me on this point).

[81] I wish to thank an anonymous reviewer for pointing out the ambiguity between these two interpretations.

their own right). A similar point, although prompted by more pessimistic motivations[82], is made also by Richard Pettigrew (forthcoming), whose argument expands on List's by dramatically exploding the number of possible group belief ascriptions. In fact, it is precisely in the acknowledgement of the great variety of ways in which groups can organise themselves to form beliefs that the point of pluralist accounts like List and Pettigrews ultimately lies: belief ascriptions to groups are easy to come by, and a theory of this phenomenon should be able to capture the full range of possibilities —that is, even if this comes at the cost of abandoning the aspiration to offer a comprehensive analysis of group belief.

The latter reading, on the other hand, retains the monist flair of current accounts while taking issue with their idiosyncrasies. According to proponents of this reading, the monist aspiration to universality is not incompatible with the pluralists' endorsement of the multiple realisability of group belief. How so? How is it possible to reconcile the aspiration to universality with type-specific peculiarities of different groups?

For how I see it, the problem of extant monist accounts is that they think that the only way to achieve their aspiration to particularity is by starting from the mechanics of belief formation of *one* elected group-type. This is what makes monist analysis inherently idiosyncratic, and leads to the problem of belief uner-generation. An alternative route, however, and a much more straightforward one, I believe, would be to *start from the belief itself*, and *then* take the mechanics of the group-types as possible ways in which that belief is realised. One way of doing this (and the one I attempt in this paper) is to take different practices of belief formation as the various ways in which group belief can be realised across different group-types. In this way, a functionalist analysis of group belief would make itself available that, despite being a fully general *analysis* of group belief, would not have to give up the particularity of any of the ways in which beliefs are realised in different groups.

If this is correct, there are two ways out of the problem afflicting extant monists accounts of group belief: give up their aspiration to universality and endorse pluralism, or target their idiosyncrasy and claim that there is a single notion of group belief that is radically multiply realisable across a variety of group-types. In what follows, I show how, if we take this latter reading seriously, the HMR can motivate endorsement of a novel functionalist approach that can overcome the idiosyncrasies of competing monist accounts while retaining the pluralist intuition about multiple realisability of group belief.

II.2 Radical Functionalism about Group Belief

Multiple realisability arguments are often employed to motivate functionalist analyses of the mind. The core idea expressed in these arguments is the following: given that mental states can be realised in different ways in systems with different physical structures, then mental states should not be identified with the physical structure of the system, but with the role

[82] Pettigrew believes that such variety causes ambiguity of a vicious sort, and suggest on this basis that we abandon group belief *ascriptions* altogether (although note that Pettigrew isn't sceptical about the fact that groups may indeed be genuine believers: his point is rather a cautionary note about language use).

they play in it. In a slogan: mental states are not identified by what they *are* (i.e., their material composition, or their particular structural organisation), but what they *do* (i.e., their functional role). Because it invites thinking of mental states as individuated by their functional role rather than their material constitution or structural organisation, then, functionalism is a natural approach to account for multiply realisable states. For the same reason, I take that our natural commitment to the HMR lends intuitive support to a functionalist approach to group belief.

The first thing that needs to be clarified when developing one such account is how to define a group belief based on its functional role. Different stripes of functionalism give different responses. In broad terms, and following a classic line of thought, I propose to understand the functional role of a collective mental state in terms of the causal pattern it mediates within the system of which it is part. Take for instance pain[83]. Pain can be given a rough functionalist analysis in this sense as the mental state that tends to be generated by physical damage and, when connected with the right beliefs and desires, to cause avoidance behaviour. In a similar vein, we can take, say, a hiring committee's belief that *P* (=Victoria is the best candidate for the job) to be caused by the group's reflection on Victoria's skills and character, and causing the group to decide that she should get the job.

Following this line of thought, we can define group belief in terms of a Ramsey sentence —namely, a sentence including a collection of statements that quantify over a variable[84]. In this case, the variable corresponds to the group belief, and the collection of statements include terms that refer to external stimuli, other mental states, behaviour, and to causal relations among them. The idea here is that it is possible to analyse group belief along these lines as what tends to be caused by the right stimuli, and in turn tends to produce a corresponding typical group behaviour.

To put some flesh of the bones of this proposed analysis, more should be said about (1) the *functional profile* that characterises beliefs and distinguishes them from, say, desires or pains, and (2) about *the way in which this is realised* in different group-types[85].

(1) *about their functional profile*: at least in broad outline, providing a general idea of what the functional profile of group belief looks like is fairly straightforward. In fact, I take it to be a key advantage of thinking of group belief in functional terms that it makes available an understanding of the role belief plays in a group as analogous to that played in an individual. For instance, in individuals, it is common to think of belief as strongly connected with truth and as playing an important role with respect to action. Following Loar (1981), this connection may be expressed in terms of the tendency, associated with the state of believing X and its combination with the right sort of mental states, to *utter* X or *act so as to bring about* X —given the right circumstances and the appropriate input

[83] The example of pain as a functional state —picked up often times by Putnam— appears both in the Stanford Encyclopaedia of Philosophy (entry by Janet Levin (2018)) and the Internet Encyclopaedia of Philosophy (entry by Thomas Polger).

[84] Carnap, R. (1950) "Empiricism, Semantics, and Ontology" in Paul Moser and Arnold Nat, Human Knowledge Oxford University Press. (2003).

[85] This is no easy task, and functionalist views that clarify these questions are often hard to come by. Schwitzgebel (2023) goes so far as to say that "[p]hilosophers frequently endorse functionalism about belief *without even briefly sketching out* the various particular functional relationships that are supposed to be involved" (my emphasis). Given the plausibility that functionalism enjoys among philosophers, however, I take this observation to excuse my own lack of clarity on these issues, whose full explanation very obviously lies beyond the scope of this paper.

stimulation. Typically, appropriate input stimulation will include things like: directing one's perceptual attention to particular properties of observable objects (e.g., perception of the redness of a flower will tend to cause the belief about the flower being red) or reflection on the logical relationship between propositions (e.g., directing one's 'mind's eye' to the logical relationship between proposition $p \vee q$ and proposition $\neg p$ tends to generate belief). As for their functional profile, then, beliefs at the individual and group level have much in common —chiefly, in their connection with truth and action. Now: what about the way in which this role is realised?

(2) *about the way in which the functional profile is realised*: At least on the face of it, one should expect the two stories to come apart at this point. After all, individuals perceive with their sense organs and reflect thanks to their cognitive abilities. Groups, it is usually thought, have no eyes or brains of their own. Here's a direction one could take to characterise the realisation relation in the special case of beliefs held by a collective entity: *take facts about the type of group considered* (i.e., facts about the principles of composition holding group members together) *to impose natural restrictions on the way beliefs are realised*. This, I take it, is a fairly uncontentious claim: think of the functional property of being a woodwind instrument for example. Typically, musical instruments are woodwinds when they produce tones by the player's blowing through a mouth hole and causing an air column to vibrate. Although anything that occupies this role possesses the functional property (i.e., is a woodwind), the physical structure of the object imposes restrictions on *how* the role is occupied. In brass instruments the vibration is obtained thanks to the sensitivity of the metal, while wooden instruments (like recorders) employ a wooden reed that vibrates when the player blows into the mouthpiece.

Similarly, the functional property of being a group belief will be implemented differently depending on the type of group we consider. The details will vary slightly from case to case, but here's a general schema for how this might work. Consider three main types of groups: *aggregates* (sets of individuals sharing a common feature: e.g., doctors, married people etc.), *mechanic groups* (people who share acceptance of certain norms: e.g., cabinets, juries, parties etc.) and *organic groups* (people who work together via division of labour: e.g., companies, teams, etc.).

For aggregates, the case is quite straightforward: an aggregate perceives that *P* or behaves in *P*-related ways when all or most members of the group perceive and behave accordingly. Consider for instance a case where a political and economic crisis comes to influence the lives of a significant fraction of the population of a country. We can imagine that the effects of the crisis independently affect a large number of individuals' faith in their political representatives, who are thus independently caused to take individual actions —from small scale ordinary decisions (like verbally deprecating politicians) to more impactful ones (like taking part in rallies and demonstrations).

Organic groups, on the other hand, can be said to perceive that *P* when those members who play the same role that sense organs play in individuals do so. For instance, we say that a ship crew sees land when the lookout does, and the information is transmitted to and acted upon by the other members; characteristic behaviour in organic groups may include

cases of group coordinated assertions, whereby, say, a group of people stranded on a desert island work together to write an SOS message in the sand[86].

In mechanic groups, collective reflection may take the form of an open discussion where reasons in favour of and against a particular claim are brought to light and debated in a collegial manner; the belief of a jury, for example, can be seen as the result of some such mechanism; mechanic groups' behaviour, finally, can be seen as springing from their joint commitment —think for instance of coordinated military strikes that occur downstream of a collective commitment to shared plans, rules and goals.

Finally, then, according to this view, we can say that a group believes X when the belief thus attributed is individuated (via a Ramsey sentence) by a set of inputs (e.g., perception of X or reflection on X) and outputs (typical corresponding behaviour, like utterance or other corresponding actions aimed at bringing X about) that identify the role the belief occupies in the group host[87]. The principle of composition of the group (aggregation, joint acceptance, organic collaboration etc.) will then dictate the implementation strategy in such a way that, for example, aggregates and categories will generate group beliefs summatively via aggregation of individual beliefs (plus some aggregation procedure in some cases), and mechanic and organic groups will do so via more elaborated systems involving some sort of mechanic (joint commitment) or organic (division of labour) collaboration among group members[88].

If this is plausible, the particular version of functionalism I have presented here offers an understanding of group belief that can satisfy the aspirations of competing monist accounts without at the same time falling prey to their idiosyncrasies. First of all, because it provides a fully general, (functionalist) analysis of group belief, it aligns with other monist accounts in its *aspiration to universality*. Then, because it vindicates the particular ways in which different groups-types come to form their beliefs, it aligns with their *aspiration to particularity*.

So the version of group belief functionalism defended here is then *radical* precisely in this sense —that is, in the sense that it extends to include not only beliefs of mechanic and organic groups, but also beliefs, like some common beliefs and beliefs of unstructured aggregates, that no other functionalist view (like Bird's DM and List and Pettit's PA) can account for. In the next section, I will discuss two main ways in which scepticism about the success of this radical functionalism can be motivated. One comes from a pluralist perspective, and it considers the reasons proponents of other functionalist accounts offer to think that functionalism simply *can't* be as radical as I make it out to be. Another comes from competing monist accounts, and it aims to undermine support for my view from the

[86] This case is discussed in Lackey (2018)

[87] One way in which this could be formalised is the following:

> $\exists x \exists y \exists z \exists w$ ¦ x tends to be caused by perception or reflection of X & x tends to produce mental states y, z, and w & x tends to lead the system to express x or act in such a way to bring x about.

[88] Note that neither my analysis of belief in terms of a Ramsey sentence or the proposed conditions of its realisation are necessary ingredients for a functionalist analysis of group belief. The view I propose here is simply one possible way of adapting the functional approach to the case of group belief.

HMR by arguing that *genuine* group beliefs are not as common as our ordinary ascriptions may lead to think.

III. AGGREGATES, SOCIAL GROUPS AND OTHER CONCERNS

III.1 A functional analysis of aggregate beliefs

So far, I have argued that we ordinarily ascribe beliefs to all sorts of different groups (what I called the Hypothesis of Multiple Realisability), and I have shown that, if this is the case, the kind of radical functionalism I have defended is an excellent candidate for a comprehensive account of group belief. Indeed, if correct, the radical functionalism defended here would enjoy an enormous explanatory advantage over existing views, since it would be the only view on the market able to accommodate beliefs formed in most of the groups to which we typically attribute them.

This is a point worth stressing. Philosophers working on collective epistemology regularly highlight the importance of theorising about groups on the grounds that group attitudes play a fundamental role in our everyday life[89]. Our lives are shaped by group level decisions —think for instance of the deliberation of a jury, a big company's resolution on environmental matters, or a political party's commitment to some course of action and so on— and we want these groups to be accountable for them. Being able to recognise groups as genuine believers not only helps us make sense of our ordinary attributions, but also to reason normatively about group actions and attribute responsibility. Along this line of thought, then, it becomes clear how being able to attribute beliefs to different *kinds* of groups puts us in a position to understand and intervene in the deliberations and manoeuvres of groups that would otherwise be left out by theories that focus only on beliefs held by some kind of group but not others.

Still, there are some important concerns that a view like the one proposed here might at first appear to be vulnerable to. After all, being persuaded of the plausibility of the HMR doesn't yet suffice to grant commitment to my view. Pluralists, for instance, who favour reading (1), take the HMR as evidence for a very different claim —namely, that there are different kinds or notions of group belief. Christian List[90], for example, makes the case for a functionalist view of group belief that closely resembles my own, when he suggests that

> "[i]f we understand intentional attitudes such as beliefs and desires in functionalist terms […] then an agent's beliefs are simply those states of the agent whose functional role is to represent certain features of the environment […]. The beliefs […] of a group agent are thus whichever states of the organized collective play the relevant functional roles." (2014).

At the same time, however, he rejects the idea that this functionalist view can be made to work for most of the ways in which groups can form their beliefs, which leads him to endorse a form of pluralism about group belief. Why so? I can see two main reasons that can motivate one to be sceptical about extending some such functionalist analysis to most group beliefs. The first concerns the (sometimes great) difference between the ways beliefs

[89] In particular, Lackey (2020), List and Petitt (2011) Pettigrew (forthcoming) and List (2014)
[90] On his own (2014) and with Pettit (2011)

are realised in different group-types. The second has to do with the difficulty of extending a functionalist analysis of group belief to aggregate groups in particular. The rest of this section will be dedicated to addressing each worry in turn.

About the difference between the ways beliefs are realised. The first worry, in a nutshell, is the following: the instances of 'aggregate', 'common', and 'corporate' group belief are so radically different that one may worry that it is simply impossible to identify a unified functional profile capable of subsuming them all. For instance, the functional profile involved in (say) the belief of Gen Z that protecting the environment is important, is *very different* from the functional profile involved in (say) the belief of British Petroleum that it should increase its profits[91].

I think that this worry originates from the ambiguity between two ways in which belief forming practices can differ from each other —i.e., with respect to the physical composition/structural organisation of their realisers, or with respect to the role beliefs play in the system in which they are part. I agree that the ways in which groups form their beliefs can vary radically among group-types. Attributing a belief to Gen Z often amounts to nothing more than identifying a belief that is shared by a particular set of individuals. Something very different is going on, on the other hand, in the case of beliefs attributed to corporate groups like BP which, we might expect, often involve some level of collaboration among group members (whether mechanic, organic or both).

Notice though that this difference, however profound, only concerns the physical composition of the two groups (say, the number of members, or their qualities) or their structural organisation (i.e., the difference between the principles of compositions holding them together). Since functionalism defines its analysandum in terms of what it *does* (i.e., the role it occupies) and not by what it *is* (i.e., its physical composition or structure), however, this difference hardly poses any problem. The beliefs of Gen Z and BP are very different in the sense that they *have very different structures*: what realises one belief is very different from what realises the other. However, they still *do similar things* (i.e., play similar roles) in the systems of which they are part (of course, compatibly with the fact that they have a different content) —namely, (and very roughly) they (tend to) result from the groups' interaction with their external environment, and (tend to) guide the behaviour[92].

About the problem of incorporating aggregate beliefs in a functionalist analysis. Another worry that could be raised for my functionalist account is that it can't be extended to include beliefs formed in aggregate groups. One way of motivating this worry —championed, among others, by List and Pettit (2011), as well as by Pettigrew (forthcoming)— is to argue that only the beliefs of group *agents* can be analysed functionally. This idea comes from the

[91] I wish to thank an anonymous reviewer for pressing me to clarify this by raising this objection.

[92] Consider a case where Gen Z and BP have the same belief X ("that we should care about environmental matters"). Since the two groups have different powers, goals and responsibilities, the same beliefs will have very different functional profiles in the two groups. Does this mean that they can't be given a common functional analysis? Of course not, since the variety in powers, goals and responsibilities is built into the functionalist analysis as what determines the functional profile of a belief. That's precisely why the functionalist analyses are normally formulated in terms of what *tends* to be caused by X and what *tends* to cause X. In fact, the same variety in functional profiles occurs also among individuals —depending on the powers, dispositions, responsibility, social positions etc., the same belief will have very different causes and outcomes in different individuals.

observation that to think functionally about belief is also, and importantly, to think of its close connection to intentional action. Since intentional action is a prerogative of intentional agents, the thought goes, one cannot extend to groups a functional analysis of individual beliefs that draws so importantly on individuals' ability to act intentionally *unless* such groups are themselves agents. Since aggregate groups aren't agents, they conclude, functionalism can't be applied to beliefs formed in those groups.

But why think that aggregate groups are not agents?[93] Here's List and Pettit about what it takes to be a (group) agent: "[a]ny multi-member agent must be identifiable over time by the way its beliefs and desires evolve. So there must be a basis for thinking of it as the same entity, even as its membership changes due to someone's departure or the addition of new members" (2011, ch. 1.3). Mere aggregates or collections, they think, do not satisfy these criteria. Why though? For surely there are aggregates that satisfy them. Gen Z, for instance, does persist over time, even if some members add or depart from it, and its beliefs and desires, and their change over time, are the target of market speculation and scientific study. So why not think of it as a group agent?

In another passage, List and Pettit say more in support of their conclusion:

> "So how do we draw the line between non-agential and agential groups? Our discussion suggests that we regard a group as an agent just when we think something is amiss if those attitudes are inconsistent, or otherwise irrational. [...] We assume that only group agents (as opposed to mere groups) should acknowledge that this is a fault that should be rectified". (2011, ch. 1.3).

Again, this seems to say nothing that at least some aggregate groups can't satisfy. Think for instance of an electorate. An electorate is a mere collection of individuals based on their entitlement to vote in a political election. Yet, electorates don't just persist over time, don't just play an important social role in the political life of a country, but they are also treated as rational agents —when, say, they are swayed by populist appeals, and are then criticised for being so gullible, and it is expected, or demanded of them that they don't do that in the future. Electorates respond to the results of political decisions by changing their political allegiances, demanding the resignation of an MP or urging political intervention on some relevant matters. Similar observations also apply to Gen Z, when, say, they are criticised for being too sensitive, too attached to the internet, too obsessed with environmental issues, and so on. Both groups have intentions and desires (like, say, stopping the rise of populism in Italy, reframe the discussion surrounding mental health, create more trans-friendly environments in our society) and take actions to achieve them (by, say, voting for a particular political party, taking part in community initiatives, adopting the use of trans-friendly pronouns and so on). If this is true, aggregates and collections *should*, at least sometimes, be considered as genuine group agents[94].

---

[93] In what follows, I adopt List and Pettit's (2011) notion of group agency. For other ways of thinking of group agency, see e.g., Epstein (2019).

[94] In fact, List and Pettit themselves ultimately agree with this. They consider the possibility of what they call 'coalescent agents' —i.e., aggregate groups that display agent-like features, and even offer an example of how this could work, by considering the network structure of a 'terrorist organisation', established to achieve some goal, but composed by members who are unaware of it. "The

But this may still not be enough. For even if we granted that some aggregate groups are genuine group agents, one may worry that my functionalist analysis still wouldn't be able to apply to them, since in aggregate groups *there is no belief at the group level* that is *distinct* from the individual beliefs and desires of the individuals that form part of the group. If this is the case —that is, if there is no group-level belief that could be distinguished from the individual beliefs of the group members— then there is literally no *collective* belief to which my functionalist analysis could be extended —or so the thought would go.

To see where this worry comes from, think of a group of people waiting at the bus stop. Here it seems possible to come up with a functional profile for the group's belief that the bus is due to arrive soon —say, as the state that tends to be caused by the group checking the bus timetable, and that has the tendency to cause them to be on the lookout for the bus, have their tickets at the ready, jump on it as soon as it arrives and so forth. The problem here is that even if this functional story can be successfully put together, and the group could (very loosely!) somehow be treated as an agent, there seems to be nothing *distinctive* about the functional role of this belief that justifies its attribution *to the group itself* rather than simply *to its individual members*.

In fact, it is far from clear that when a bunch of people are waiting for the bus, in addition to the beliefs that each individual person has there is *also* a corresponding state of the group which realises the functional role of the belief that the bus is due to arrive soon. So, the thought goes, unless we are given reasons to think that there *is* something special about this particular kind of case, a functional analysis of mere aggregates that fits the weak inflationism I defend in this paper doesn't seem viable.

I agree that it would be redundant to stipulate the existence of an additional belief at the collective level for groups (like the set of people waiting for the bus) when its functional profile is just the same as the member's individual beliefs. Sometimes, beliefs formed in summative ways don't have anything distinctive about their functional profile. Sometimes, however, they *do*. Consider for instance this case. Suppose there is a shared belief among people living in Scotland that there is a cost-of-living crisis[95]. Like before, the functional profile of the group can be identified quite easily: we can imagine, for instance, that Scots have come to hold this belief by reading their daily newspaper, watching television, discussing the current political and economic situation with their friends, or more directly because they experienced an increase in rents or in the price of basic consumables. In turn, this will affect their collective behaviour in specific ways. For instance, with respect to ordinary matters, such as their shopping habits, but also in relation to less ordinary and more socially impactful ones, like being willing to organise themselves in protest groups, or participating in demonstrations and so on. In this latter respect, in fact, it is easy to see that the functional profile of the collective belief (i.e., the belief which is the sum of the beliefs of those Scots who think that there is a crisis) differs substantively from that of the individual beliefs that constitute it. The former has the tendency of bringing about radical

organization" they conclude, "would be composed of a group of people, in perhaps a thin sense of group, and would function as an agent" (2011, end of ch. 1.3).

[95] A discussion of this case has been proposed by an anonymous referee, whom I wish to thank.

social changes, like the formation of protest groups or the rise of a political party, while the latter doesn't[96].

This is not an isolated case. Take the beliefs of Catholics that (say) same sex marriage is a mortal sin. This could be given a functional description in terms of a state that tends to be acquired through preachings (or, more generally, through exposition to Catholic principles and practices), and that, when suitably related with other mental states, is liable to cause Catholics to assume antagonistic attitudes towards the adoption of inclusive practices in their communities, to favour conservative policies and resist social and progressive political initiatives[97]. Likewise (although it may be difficult to trace a specific profile in this case), we may schematise rappers' appreciation of hip hop, roughly, as a set of states, including beliefs, individuated by the group's exposition to this genre, and a tendency to listen to it, attend, organise and get involved in thematic events that reinforce and perpetuate the culture.

The perpetuation of a musical genre, its growth and popularisation, or the rooting of an ideology and its weight in a country's political outlook, are (at least in part) plausible outcomes of beliefs that are predominant in a population, and that can be explained precisely by virtue of how pervasive they are. Such beliefs are capable of determining radical social changes precisely because they extend to a population that is sufficiently numerous.

So there *are* some aggregate groups that not only can be ascribed group-level mental states, but whose ascription cannot be reduced to its members without an explanatory loss[98]. If this is true, then, also the latter motivation for thinking that my functional analysis could not be extended to include aggregate groups —i.e., the one insisting on the necessity of a distinction between the functional profile of belief at the group- and individual-level— is disarmed too.

Now, one may wonder whether, in suggesting that *there is* a difference between the causal profile of the belief at the individual and group level, I am not thereby giving up summativism altogether. Given the substantial difference between individual and group minds, however, I don't think —with Gilbert (2014)— that it is good practice to expect beliefs at the collective level to match exactly the shape of individual beliefs. More in particular, however, this is obviously wrong when it comes to their functional profile.

[96] One may argue that I am here ruling out the possibility that individuals, collectively or as a group, may bring about social changes. Since my view is weakly inflationist, however, it is perfectly compatible with my view that the belief of the group is nothing other than the sum of the beliefs of the individuals that constitute the group —and hence that it is indeed the collection of individuals who are bringing about radical social change. The distinction I am making here is merely between the functional role of the belief at the individual and collective level: so long as the functional profile of the sum of the beliefs of the individuals and the individual beliefs themselves differ, this justifies a distinction between them.

[97] Note that this does express a different Ramsey sentence than the one given earlier. Rather, this is a specification of the way in which that Ramsey sentence is realised in a very particular group.

[98] Maybe there aren't many (although this seems implausible since it doesn't seem too difficult to come up with aggregate groups that satisfy the functionalist criteria; most categories seem to be good candidates, as well as other random collections like 'cat lovers', 'flat earthers' and so on), but even so, that'd still be enough —so long as there are some collections (aggregates or categories) whose beliefs have the right sort of functional profile (i.e., one that is distinctive of the group itself), then it seems plausible to take some such groups to be compatible with the sort of functionalist view I defend here.

Groups don't have eyes or brains of their own, but it would seem silly to deny that they can think and see *purely on this basis* (unless one is inclined to accept some rather controversial form of radical exceptionalism about individual minds).

More positively, I would say that the difference between the causal profile of the two beliefs shouldn't encourage one to think that the beliefs themselves are different any more than the difference between the sound of a G note played by an orchestra and by a single flute justifies one to think that they are different notes. The properties of a sum may sometimes differ in some respect from the properties of the entities that constitute it, even if the sum *just is* a sum of those entities. The summative group formed by the Scots who believe that there is a cost of living crisis is nothing more than the summative aggregation of the individual beliefs of a large number of Scottish people. Still, it is not unless we assume the existence of an overarching collective belief, with its own profile, that we can make sense of the changes brought about by there being a sum of people sharing the same belief.

In conclusion, then, not only aggregates and categories do often count as genuine group agents, but their functional profile also presents characteristics that justify the attribution of the belief to the group itself, and not just to its members. Resisting the functional approach to the weak inflationism defended here on the grounds that functionalist analyses cannot be extended to aggregate groups, then, cannot be motivated by these reasons. As long as it makes sense to treat aggregate groups like genuine agents, and so long as there are cases where a relevant difference between the functional profile of the belief held by the group and its members cannot be eliminated or ruled out, the additional attribution of a belief to the group is warranted, and its functional analysis legitimate.

III.2 Social groups

At the end of section §I.3 I have noted how extant monist accounts cannot surpass the problem of belief under-generation unless they attempt to address the tension between their aspiration to universality and particularity. To do so, I briefly sketched, they would have to provide a story legitimising the idiosyncrasy of their view —i.e., of legitimising the 'universalisation' of their preferred way of forming group beliefs. A natural way of doing that (and I suspect one that has been implicitly endorsed by most inflationists about group belief) would be to claim that their chosen belief forming mechanism is the only one yielding beliefs that are *truly collective*.

In what follows, I will give space to an attempt to provide some such story by a proponent of the joint acceptance account. In short, the argument aims to establish that only beliefs formed via joint acceptance by mechanic groups should be considered *genuine* group beliefs. Importantly, because it takes beliefs formed in other ways by different types of groups not to be *genuine*, this argument implies that group belief isn't multiply realisable. If true, then, this story wouldn't only successfully defend competing monist accounts, but it would at the same time risk jeopardising one of the main motivations for my view[99].

[99] The argument I propose takes the moves from a consideration advanced by Margaret Gilbert (1987), and is thus phrased as an attack to my view from a JAA perspective. However, a similar criticism is also hinted at by Alexander Bird in his (2019), and the same objection could be rephrased, *mutatis mutandi*, as being levelled at my view from a proponent of the distributed model as well. As will

To start, notice how a lot hinges, for the plausibility of the argument, on the difference between genuine and non-genuine group belief[100]. Here's Gilbert's attempt at making sense of this distinction:

> "[...] one might expect that for statements of the form 'Group G believes that p' to be really apt, they would refer to a phenomenon *involving a group in a more than accidental way*." (Gilbert 1987, 189) (my emphasis).

The suggestion here is that the grounds for distinguishing genuine from non-genuine beliefs should be found by looking at the *groups* —or better, at the way in which groups are involved in the formation of the belief. According to Gilbert, it is the fact that it involves a group 'in a non-accidental way' that makes the belief *genuinely* collective. But what does this mean more exactly? One way to cash out the idea that a belief involves a group in a non-accidental way is by reference to the distinctively *social* nature of the group hosting genuine group beliefs[101]. Genuine group beliefs, it is argued, possess a social dimension that non-genuine beliefs lack because they are formed in a *social group* —that is, they are produced by a population where the fact that they are a social group is essential for the existence of the belief. In other words, what distinguishes genuine from non-genuine belief is that in the former, but not in the latter, a social group figures in the formation of the belief[102].

Now, note that, even if correct, the proposed distinction between genuine and non-genuine beliefs doesn't constitute a threat to the plausibility of the HMR by itself. After all, it may very well be that categories, aggregates and organic groups (where, typically, beliefs are not formed via joint acceptance) do constitute social groups —and so that genuine group beliefs are not only those formed in mechanic groups via joint acceptance. With this aim in mind, then, JAA theorists should argue that only those types of groups that host beliefs via joint acceptance, like mechanic groups, are genuine social groups, whereas other, 'accidental' types of groups, like aggregates, categories and organic groups, where beliefs are formed via other means, aren't. More schematically, we can summarise the argument as follows:

1. *Genuine* group beliefs are such when they play a distinctive social role.
2. Group beliefs play a distinctive social role when they involve a social group.
3. Aggregates, categories and other 'accidental' groups are not social groups.

---

become evident, the reason why I limit my focus in this way is that the general strategy of my response would remain the same in both cases.

[100] Notice that my question doesn't ask what the difference between genuine and non-genuine group beliefs is according to Gilbert —this question has a clear answer: the fact that one involves joint acceptance and the other doesn't. The question I am asking here is instead the following: "what motivates Gilbert to say that group beliefs that involve joint acceptance are genuine, and that others aren't?". That is, what I am looking for is a story explaining why we should think that group beliefs formed in mechanic groups via joint acceptance are *genuine* group beliefs while others aren't. On pain of circularity, the answer to the latter question cannot be the same as the answer to the former.

[101] This is suggested by Bird in his (2019).

[102] For Gilbert there are a variety of ways in which a group may figure in the formation of a belief, from a formal discussion to more informal agreements like a nod between strangers.

(C) Group beliefs hosted by aggregates, categories and other 'accidental' groups are not *genuine* group beliefs.

If the support I take the HMR to offer to my functionalist view depends on the fact that group beliefs can be realised in different ways in different group-types, and if mechanic groups are the only group-type where genuine beliefs are realised (via joint acceptance), then the conclusion of this argument is meant to suggest that the HMR fails at doing that.

But is this argument plausible? First of all, notice that acceptance of premise (iii) depends on a substantial sociological claim about the kinds of entities that ought to count as social groups. Now, does sociological practice justify a distinction between mechanic groups on the one hand, and aggregates, categories and organic groups on the other? The short answer is that it doesn't. Sociologists distinguish a vast array of group-types depending on the scope and interest of their studies (Reicher 1982, Tajfel & Turner 1979, Forsyth 2010). Based on different principles of compositions (e.g., the members' interactions, their goals, interdependence, unity, structure, objective similarity, etc…), they recognise that group sociality comes in all shapes and forms: street gangs, mobs, communities, peer groups and crews, for example, are all considered fully-fledged social groups according to sociological taxonomy. If this is so, however, it is not at all clear why only one group-type, typically associated with a particular non-summative way of forming beliefs, should have any right to be considered more distinctively social than other social groups. Categories (doctors, plumbers, women, Catholics), as well as organic groups (the scientific community at large, the French society) have the same social status mechanic groups have. They are simply tied together by different principles of composition.

But one may point out, as a reason to draw the distinction, that mechanic groups are more tightly connected, and occupy a 'higher position' than other groups in terms of their social complexity[103]. Based on this consideration, they might argue that only 'sufficiently sophisticated' groups count as genuine social groups, and that aggregates, categories and organic groups aren't as sophisticated as mechanic groups.

Now, an obvious worry with this attempt has to do with identifying the criterion to determine when a group is sufficiently sophisticated to count as genuinely social in this sense. This worry is particularly pressing given how broad and varied the domain of mechanic groups itself is. For example, in Gilbert & Pilchman (2014), the authors offer the example of six people sitting in the same train compartment and agreeing with a nod that the train carriage is hot as a case of a mechanic group forming a genuine group belief. If the criterion to make the distinction between social and non-social groups is their tightness, it is natural to ask why strangers in a train compartment would display a higher social complexity than members of a community, people that share the same office every day or populations of individuals that share common beliefs and desires.

But most importantly (even granted that groups can be ordered by complexity, and that a neat separating line can be found that traces the distinction between categories and organic groups on the one hand, and mechanic groups on the other) the main worry here

[103] Admittedly, this way of defending Gilbert's account may not be her preferred one. In the absence of a strategy that she explicitly endorses, I consider this one here.

is that it is not clear why considerations about the social complexity of a group should matter for determining the 'genuinity' of their collective belief. On what grounds should we consider a nod between strangers in a train carriage conducive to group belief in a more genuine or distinctive way than a belief shared by a collection of people? So long as a group satisfies the conditions, set by sociologists and social scientists, for being regarded as a social entity, then such group counts as a social group, irrespectively of whether it is an established group or any other group-type.

If this is so, the argument is unsound, and Gilbert's attempt to resolve the tension between the aspirations of her view (that is, her ambition to offer analysis of group belief, and to do so by looking at the ways in which beliefs are formed in some elected type of group) is bound to fail. And if it fails, so does too the attempt to undermine the radical multiple realisability of group beliefs and, with it, a powerful objection to my view.

III.3 Problems of Inheritance

If this is correct, the attempt of proponents of the JAA to undermine the legitimacy of the hypothesis of multiple realisability on the grounds that groups formed by aggregates, categories and organic groups are not genuine group beliefs appears to be infelicitous. But there are still other worries that a defender of a functionalist version of weak inflationism must address.

One of these relates to the very structure of this proposal, and in particular with its reliance on both summative and non-summative ways of forming beliefs. For one may argue that by allowing beliefs to be formed in summative and non-summative ways, my view makes itself vulnerable to some of the same problems that summativist *and* non-summativist accounts are also vulnerable to. Call this the *inheritance problem*. In what follows I consider two of the main objections that have been moved against strong inflationary accounts and show that the functionalist proposal defended here can handle them.

The first one has been raised against the distributed model, and it concerns the way the view identifies the subject of collective believing. Call this the problem of *group-membership over-generation*. According to proponents of the distributed model, only Durkheimian group-types are capable of hosting genuine group beliefs. The problem consists in the fact that there often seems to be a mismatch between the boundaries of Durkheimian groups and of the groups to which we would intuitively ascribe the belief. Take for instance the case of a team of physicists working at the Large Hadron Collider at CERN to demonstrate a theory that affirms X. The work is divided in subtasks, and each task is assigned to a team member based on their area of expertise. The result of the research is then published and becomes accessible to the wider public. Intuitively, it looks as though belief X should be attributed to the group whose members are physicists working at CERN.

However, it is easy to come up with scenarios where someone who wouldn't otherwise be included in the group instead figures as a member of the Durkheimian group simply in virtue of the function they play in the belief-forming mechanism. Suppose, for instance, that the scientists live in different countries, and have to share their results via mail. It then turns out that a fundamental role in the production of the result is played by the mailmen who diligently delivered to each member of the group of scientists the partial results of their colleagues. If this is correct, it looks as though the restriction imposed by the

Durkheimian properties on the groups that can host the belief allows members that are extraneous to the group to which we would intuitively attribute the belief to be included anyway[104].

The distributed model, the case shows, fails to attribute the belief to the group of scientists, and attributes it instead to the larger group including the mailman. This problem, the objection goes, would extend to functionalist views of the kind defended here, and threatens to undermine their capacity to comply with our intuitions when it comes to attributing beliefs to organic groups.

Functional and distributive accounts of group belief differ in important ways. Recall that the latter provides a (minimal) dismantling analysis that includes two main conditions; the first concerns the way in which the belief is realised (namely, the organic collaboration of the group members). The second has to do with the characteristic Durkheimian structure that implements such collaboration. This latter condition imposes a strong restriction on the groups that are apt to host beliefs, thus drawing fire from cases such as the mailmen. The kind of functionalism defended here, on the other hand, imposes no such restrictions on the group realiser. On the contrary, according to my view, it is the group agent that dictates the way in which beliefs are realised. This means that it is open to a proponent of my view, in assessing a group's epistemic status, to first identify the boundaries of the relevant social group, and then determine the way in which such group realises the belief accordingly. The possibility of a mismatch between the boundaries of the group under consideration and those imposed by the definition of group belief, is therefore ruled out *ex ante*.

A different problem is posed by the implementation of the joint acceptance account. 'Rejectionists' find the JAA troublesome because it posits a belief-forming mechanism that has more traits in common with the attitude of acceptances than beliefs; and belief, rejectionists argue, is relevantly different from acceptance in two main respects: it is involuntary, and it typically aims at truth[105].

Gilbert has persuasively argued against rejectionism and in defence of her account at length (2004, 2013, 2014). Her point is that "[o]ne should not assume that accounts and distinctions arrived at within individual epistemology are appropriately applied within collective epistemology, however central they are to individual epistemology" (Gilbert & Pilchman 2014). Although it is true that, on her view, group members must accept a proposition in order for the belief to be formed at the group level, the collective mental state thus produced shouldn't thereby be thought of as a form of acceptance, on pain of reducing her view to a deflationary summativist one.

It is true though that the two inquiries (at the individual and collective level) are not entirely separated, and the problem then emerges as to how to assess their relationship, and how to decide when insights at one level also apply to the other. However, it is difficult to correctly evaluate how this could be done, and Gilbert's suggestion to leave the decision to the details of individual cases hardly seems decisive.

[104] I am grateful to Mona Simion for suggesting this line of argument against Bird-style Durheimian group-types, and for helpful reflections on how to address this problem.
[105] Most notably Cohen (1989). But see also Hakli (2007) Wray (2001).

The joint acceptance accounts and the organismic ones are popular accounts of group belief, and it is in the interest of a defender of this functionalist view to show how they can work together in a unitary framework. This is what I've tried to do here, although the inconclusiveness of the debate between defenders of the JAA and rejectionists, I grant, does not speak in favour of the JAA (even though it does not set the matter in favour of rejectionism either). The key advantage of the kind of functionalism defended here, however, is that its plausibility is independent of commitment to any particular view about the ways in which group beliefs can be formed. Those who find the JAA proposal inadequate, for instance, may prefer summative or distributive ways of forming beliefs, and attempt to explain how collective entities like established groups, which are taken to form beliefs via joint acceptance, instead form beliefs in this alternative manner.

III.4 Over-generation

A final problem I wish to address here stems from the consideration that the functionalism defended here is liable to lead to an over-generation of group belief. One of the advantages of my functionalist view is that it can account for group beliefs being realised in multiple ways by different group-types. Such versatility, however, can become problematic if it leads to counter-intuitive attributions of group beliefs. In the jury example, for instance, we noticed that a natural way to assess this case is to say that the jury believes that the defendant is innocent even if all the jurors personally believe that they are guilty. Still, by functionalist light, beliefs can be formed via joint acceptance as well as via aggregation of individual beliefs. My view's prediction, then, should be that the jury, via summative aggregation, does believe that the defendant is guilty as well (given that this is what the jurors' personally believe), even though we wouldn't intuitively attribute such belief to the jury[106].

In order to counter this objection, it is important to notice that the same collection of individuals may fall under various descriptions depending on the social relations we take to be relevant. Consider for instance a group of people that form part of the board of a music magazine and that, during a meeting, they collegially decide X (= *x* is the best track of the month). The committee members, however, are also old-time friends, and meet every day at the local pub. While discussing the decision they took as committee members, they find out that they all personally believe Y (= *y* is the best track of the month). At the pub, they constitute a group that has the common belief Y, which is different from their view as committee members. However, this doesn't seem to be problematic at all —indeed, it may well be the case that the same collection of individuals believes different things when grouped according to different sociological principles of compositions.

This distinction disarms the threat from belief over-generation. Suppose that the facts we are interested in about the group of people that constitutes the jury are the social bonds, determined by the law, that derive from their being members of the jury. In this case, such facts will restrict the ways in which the belief role will be occupied so that the mere fact that they share a common belief won't suffice for belief attribution (in compliance with ordinary language attributions). If, on the other hand, we are interested in the mere

[106] I wish to thank an anonymous referee here who has noted that the case of coextensive groups originates from Gilbert (1987).

collection of people that constitute the jury rather than the jury itself, beliefs are rightly attributed to the jurors via simple summative aggregation. This also explains why the natural way in which we assess this case is by attributing to the jury the belief of the established group, given that, when we attribute beliefs to a jury, we typically refer to the legal institution and not the mere aggregate of individuals.

CODA

The aim of the view I have defended in this paper —namely, healing the fracture between pluralists and monists about group belief, while at the same time resolving the debate between deflationary summativism and strong inflationary non-summativism— is an ambitious aim. Naturally, ambitious aims set high stakes, and I am sure that the view I defend here in many ways falls short of achieving them. In particular, there are worries (concerning the threat of belief over-generation, for instance, or the risk that my view would inherit the problems of the belief forming practices that it incorporates) that I have not had the time to address in this paper, but that would have to be part of a complete defence of this view.

Even so, I do think that the arguments provided in this paper are successful in other, more important, ways. For instance, I do think that my view succeeds at recognising that this ambitious goal can be achieved, and in sketching a view —namely, my functionalist approach to weak inflationism about group belief— that is able to do so. How so? First of all, this view supports, with other inflationist accounts, the intuition that organised groups (e.g., mechanic and organic groups) are the proper subject of genuine belief attributions. As a *weak* version of inflationism, however, it allows beliefs formed in deflationary summativist ways to count as genuine group beliefs *too*. For this reason, my view also manages to strike a fine balance between monism and pluralism about group belief. Because it allows beliefs to be formed in both inflationary and deflationary ways, it can accommodate the great variety of belief attributions, like other pluralist views. Because it provides a functionalist *analysis* of group belief, however, it does so from within the monist camp.

Finally, by extending to groups a functionalist approach standardly adopted at the individual level, the view defended here has the advantage over competing accounts of integrating nicely with current scientific and philosophical research, and of offering a unitary picture of the nature of the mind.

Chapter five

Gender, Race, and Group Disagreement

ABSTRACT

This paper has two aims. The first is critical: it argues that our mainstream epistemology of disagreement does not have the resources to explain what goes wrong in cases of group-level epistemic injustice. The second is positive: we argue that a functionalist account of group belief and group justification delivers (1) an account of the epistemic peerhood relation between groups that accommodates minority and oppressed groups, and (2), furthermore, diagnoses the epistemic injustice cases correctly as cases of unwarranted belief on the part of the oppressor group.

INTRODUCTION

A hotly debated question in mainstream social epistemology asks what rational agents should believe when they find themselves in disagreement with others.[107] Although special attention has been paid to disagreement between *individuals*, recent developments have opposed this trend by broadening the focus to include cases of disagreement between *groups*.[108] We argue that this shift is interesting because the phenomenon of inter-group disagreement (such as e.g. the disagreement that occurs between opposing political parties, or countries) raises some distinctive challenges for our methodological choices in the epistemology of disagreement. To do that, we look at two cases of group disagreement, one involving gender discrimination, the other involving the marginalisation of racial and religious minorities, and argue that mainstream epistemology of peer disagreement essentially lacks the resources to explain what is going wrong in these cases. In this paper, we advance a two-tiered strategy to tackle this challenge by drawing on an inflationist account of group belief and an externalist account of the normativity of belief in the face of disagreement.

[107] Lackey (2010), Christensen (2009), Feldman & Warfield (2010), Matheson (2015), Kelly (2005) and Elga (2007).
[108] Carter (2016), Skipper & Steglich-Petersen (2019).

Here's the structure of this paper. We start off the discussion by presenting two examples of discrimination in cases of group disagreement, and then offer a diagnosis of the distinctive form of epistemic injustice at play (#2). We then proceed to examine the prospects of extant views in the epistemology of peer disagreement to address the problem raised in the first section, and conclude that they have difficulties accounting for what went wrong in these cases (#3). We suggest that the problem lies at methodological level, and advance a two-tiered solution to the problem that relies on an externalist epistemology and a functionalist theoretical framework (#4).

I. GENDER, RACE AND GROUP PEER DISAGREEMENT

Consider the two following cases:

> SEXIST SCIENTISTS: During a conference on the impact of climate change on the Arctic Pole, a group of male scientists presents their most recent result that *p*: 'the melting rate of ice has halved in the last year'. In the Q&A, a group of female scientists notes that *p* doesn't take into account the results of a study published by them, which supports not-*p:* 'it is not the case that the melting rate of ice has halved in the last year'. Not-*p* is, as a matter of fact, true, but the group of male scientists continue to disregard this option solely on the grounds that her research group was entirely composed of female scientists.
>
> RACIST COMMITTEE In a predominantly Christian elementary school, the RACIST committee convenes to discuss what food should be served for lunch the uncoming semester. As it turns out, white schoolteachers of Christian faith exclusively compose the committee. After a brief discussion, the committee comes to believe, among other things, that *q*: 'Children should be served pork on Wednesdays.' A small group of non-white Muslim parents, informed of the outcome of the meeting, raise a number of independent formal complaints against the RACIST committee on the grounds that the decision doesn't respect the dietary restrictions of their religion and arguing for not-q: 'It is not the case that children should be served pork on Wednesdays.' Due to racial prejudice, however, the RACIST committee ignores the complaints, and no action is taken to amend the decision.

In the first case, the group of male scientists dismisses a relevant piece of evidence based on their prejudice against women. Because of their gender, the women's team fails to be rightly perceived as a peer. In the second case, the group formed by the parents of the school kids is discriminated against because they constitute a racial and religious minority.

It is crucial to note that, although moral harm is definitely at stake in these cases as well, the kind of harm perpetrated is distinctively *epistemic*, in that both discriminated groups are harmed in their capacity as knowers (Fricker 2007). What is common between the two cases is that both manifest some form of *epistemic injustice*– i.e., the discriminated groups fail, due to their hearers' prejudices, in their attempt to transmit a piece of information they possess. Moreover, the epistemic harm at stake here is the result of a fundamental epistemic failure on the part of the oppressive groups. The group of scientists and the

school representatives don't simply *happen* to fail to notice some relevant piece of information, nor it is the case that they aren't in a position to easily access it. Instead, upon being presented with the relevant piece of evidence, they discount it for no good epistemic reason; in this, the oppressor groups fail to be properly responsive to evidence (Simion 2019a).

The above cases represent instances of disagreement between groups, whereby the disagreement is resolved in a bad way: the oppressor group ignores or dismisses the information the oppressed one attempts to transmit, and this happens in virtue of the social dynamics that are particular to the two types of case: it is the prejudiced belief that the male group of scientists have towards women, and the RACIST committee has towards minorities, that prevents them from perceiving their interlocutors as their peer.

We strongly believe that the epistemology of disagreement should be able to account for what is going wrong in these cases. Furthermore, we think that if our epistemology is not able to do so – i.e., if we don't have resources to explain the arguably most ubiquitous and harmful among epistemic failures, of which these cases are prime examples of – our epistemology requires a swift and radical methodological change. For this reason, an important question that such examples raise is the following: are extant accounts in the epistemology of disagreement sensitive enough to actual social dynamics to be capable of explaining what went wrong in these problem cases?

II. A (PROBLEMATICALLY) NARROW METHODOLOGICAL CHOICE

Epistemology at large is concerned with what is permissible to believe;[109] given this, it is a matter of surprising historical contingency that the vast majority[110] of the literature in the epistemology of disagreement concerns itself with a much narrower question, i.e.: 'What is rational to believe in the face of disagreement with an epistemic peer?.' (henceforth, the question).[111] The question is narrow in two crucial ways. First, in that it is explicitly conceived as concerning an internalist accessibilist notion of rationality[112]: the version of the question that the vast majority of the literature concerns itself with is: 'Given all and only reasons accessible to me, what is rational for me to believe in the face of disagreement with an epistemic peer?'

A second crucial way in which the question is narrow is in that it is not primarily concerned with real cases of everyday disagreement, but rather restricts focus to highly idealised cases in which one disagrees with one's epistemic peer. The thought is that if we answer the question for perfect peerhood, we can then 'upload context' and figure out the right verdict for cases of real-life disagreement as well. Here is how David Christensen puts it:

> The hope is that by studying this sort of artificially simple socio-epistemic interaction, we will test general principles that could be extended to more

[109] See Step and Neta (2020).
[110] But see e.g. Broncano-Berrocal & Simion (2020) and Hawthorne & Srinivasan (2013) for exceptions.
[111] Lackey (2014)
[112] Internalist accessibilism is the view that epistemic support depends exclusively factors that are internal to the subject and accessible through reflection alone (e.g. Chisholm 1977, 17)

> complicated and realistic situations, such as the ones encountered by all of us who have views–perhaps strongly held ones–in areas where smart, honest, well-informed opinion is deeply divided. (Christensen 2009: 231).

One notable difficulty for these accounts concerns how to define the notion of peerhood at stake in the question. In the literature, epistemic peerhood is typically assessed along two main lines: cognitive or evidential equality.[113] Agents are taken to be evidential peers if they ground their confidence in a proposition *p* on pieces of evidence that are epistemically equivalent, while cognitive peers are typically taken to have the same cognitive abilities[114]. No matter the correct account, though, it is crucial to note that, *as a matter of principle*, on pain of normative misfit, the notion *cannot* feature externalist elements. After all, if the question regards a purely internalist notion of rationality, the corresponding notion of peerhood should follow suit: it should concern perceived peerhood rather than *de facto* peerhood. To see this, consider the following case:

> EXPERT CHILD My six-year-old son (weirdly enough) disagrees with me about whether the closure principle for knowledge holds. Intuitively, it seems fine for me to hold steadfast: after all, discounting him as an epistemic peer on the issue seems like the rational thing to do. Surprisingly, however, my son is, as a matter of fact, and unbeknownst to me, my epistemic peer on this topic (he is extraordinarily smart and he's been reading up a lot on the matter).

If we allow this unknown fact in the world to matter for our peerhood assignments, on conciliatory views of disagreement we're going to get the implausible result that I'm internalistically irrational to discount his testimony. That seems wrong. An internalist question about peer disagreement requires an internalist notion of peerhood.

On the other hand, a purely internalist notion of peerhood obstructs the prospects of coming to account for the phenomenon of disagreement between groups. For consider again the problem cases presented at the outset, SEXIST SCIENTISTS and RACIST COMMITTEE. By stipulation, in both cases the oppressor groups are not taking the oppressed groups to be their peers in virtue of sexist, respectively racist prejudice. As such, views on how to respond to peer disagreement internalistically conceived will not even straightforwardly apply to the cases above, since they will not count as cases of peer disagreement to begin with.

Recall, though, that focusing on the narrow question was not supposed to be the end of the road in the epistemology of disagreement. After all, cases of perfect peer disagreement are rare, if not even non-existent. The thought was that, as soon as we figure out the rational response in these idealized cases, we could upload context and get the right result

[113] Lackey (2010).

[114] There is still ongoing debate on how to spell out the notion of cognitive or evidential equality. The former is typically understood in terms of sameness of reliabilist (i.e., a well-functioning cognitive system) or responsibilist (e.g., open-mindedness, humility) virtues. The latter is sometimes taken to require 'rough sameness' of evidence and mutual knowledge of the relevant differences (Conee 2010). However, neither route is fully satisfactory. For a useful discussion of the prospects and problems of this problem see Broncano-Berrocal & Simion (2020).

in real-life cases as well. So maybe once we do that for the cases at hand – i.e., upload context - things will start looking up?

Unfortunately, there is reason to believe otherwise. There are two broad families of views in the literature on peer disagreement: *conciliationist views*[115] and *steadfast views.*[116] Conciliationists claim that disagreement compels rational agents to decrease their confidence about p when faced with peer disagreement; steadfasters deny this claim, and argue that, in such situations, rational agents are entitled to hold on to their beliefs.

What is the verdict these views give us on the examples discussed at the outset? The case is quite straightforward for steadfasters: if a rational agent (in this case, a group) is entitled, in the face of disagreement with a *peer*, to stick to their guns, then, *a fortiori*, they are also entitled to do so when they disagree with someone whose epistemic position they take to be *inferior* to theirs. Such is indeed the case in both examples above. In SEXIST SCIENTISTS, the team of female scientists is not perceived as a peer by the group of male scientists in virtue of gendered prejudice; similarly, in RACIST COMMITTEE, the school representatives judge the complaint not worth of consideration precisely because it is made by a group they take to be epistemically inferior to them in virtue of racial prejudice. Steadfasters then would conclude that both the group of male scientists and the school representatives are entitled to hold on to their beliefs and discount the minority groups' testimony on the grounds that such testimony isn't recognised as being produced by a peer group.

According to conciliationism, in the face of disagreement with a peer, one should revise one's beliefs. What ought one to do, epistemically, when one doesn't take the disagreeing party to be their peer, though? The question remains open: Conciliationism does not give any prediction: peerhood is sufficient for conciliation, we don't know, though, whether it's also necessary.

In conclusion, then, it looks as though the two main accounts of peer disagreement in the literature aren't able to explain what is going wrong in the two examples presented at the outset. Even worse, in fact, we have identified two major, interrelated methodological problems that prevent the vast majority of our epistemology of disagreement to explain what is going wrong in garden-variety group epistemic injustice cases. First, in virtue of solely asking a question pertaining to internalist standards of rationality, the oppressor groups come out as justified to discount the testimony of the oppressed groups. Second, in virtue of employing an internalistic account of peerhood moulded out of disagreements between individuals, the literature fails to account for the intuition that the oppressed groups are, intuitively, the epistemic peers of the oppressor groups on the question at hand irrespectively of their social features.

We take these two problems to motivate the corresponding two desiderata for any satisfactory account of group peerhood and group disagreement. Here they are:

> *Peerhood Constraint*: Accounts of the relation of epistemic peerhood among groups should be able to account for peerhood in cases of minority groups and socially oppressed groups.

[115] Bogardus (2009), Christensen (2007), Elga (2007), Feldman (2006), Matheson (2015).
[116] Kelly (2005), Bergmann (2009), van Inwagen (2010), Wedgwood (2010), Weintraub (2013), Weatherson (2013), Decker (2014), Titelbaum (2015).

> *Normative Constraint*: Accounts of peer disagreement should be capable of providing the normative grounds on which the beliefs of oppressive groups in cases of epistemic injustice can be negatively evaluated (namely, that they be capable of recognising that the oppressive groups believe something they should not).

The two desiderata are independent, in that they concern different spaces in theory: the first desideratum sets a minimal requirement for accounts of group epistemic peerhood, in that it asks that they be capable of identifying minority groups and groups discriminated against as epistemic peers when they are so. The second desideratum, in turn, asks that accounts of group disagreement possess the required normative toolkit to identify the epistemic harm at play in frustrating the attempt of a peer group to transmit a piece of information in virtue of prejudice against them.

III. A FUNCTIONALIST SOLUTION

In what follows, we make the case for a functionalist theoretical framework that, with the resources made available from an inflationist account of group belief and an externalist account of the normativity of belief in the face of disagreement, can deliver both goods. In previous work (Miragoli 2020, Simion 2019b, Broncano-Berrocal & Simion 2020), we have independently developed (1) a functionalist account of the nature of group belief, and (2) a functionalist account of the normativity of belief in the face of disagreement. In the following sections, we will show how our functionalist accounts deliver on both the desiderata identified above.

3.1 The Peerhood Constraint: A Functionalist View Of Group Belief

To begin with, it is important to note that, even if we move away from an essentially internalist overall notion of the peerhood relation – i.e. targeting perceived peerhood - to an externalist one – targeting de facto peerhood - , the latter might not yet be fitting to capture the epistemic dimension of the social dynamics at play in the examples above. We want minority groups – which, by definition, are smaller groups, numerically – to be able to count as epistemic peers – i.e., we want that groups that are *numerically* inferior are not thereby also considered inferior *epistemically*.

Furthermore, the disagreement might occur between different types of groups: it must be possible, on the account at stake, to recognise cultural minorities that do not form established groups (either because their structure isn't sufficiently sophisticated or because they are not recognised to be such) as being the epistemic peers of more highly organised collectives. We can take this as suggesting that it must be possible for the relation of peerhood to hold between different group-types.

The debate surrounding the epistemology of groups features two main camps: *deflationism*[117] and *inflationism*[118]. The former argues that the belief of a group is nothing more than the sum of the individual beliefs of the group members. To say that Swedes

[117] Quinton (1975), List & Pettit (2011)
[118] Gilbert (1987), Lackey (2016), Tuomela (2013) and Tollefsen (2015)

believe that Volvos are safe is equivalent to say that all (or most) Swedes believe so[119]. According to deflationism, then, group belief obtains when individuals are held together by the principle of composition of aggregation. Although other sociological principles are available to explain how individuals get together to form collective beliefs, deflationists claim that genuine group beliefs are those and only attributed to aggregates - i.e., groups of people that share a common trait (such as, in this case, a common belief).

In contrast, inflationists argue that group belief is independent of the beliefs of the group members. The jury's belief that the defendant is guilty, for instance, is typically taken to hold irrespectively of the individual belief of its members[120]. There are two main inflationist views available on the market: on these views, groups form beliefs either by the joint acceptance[121] of a common view, or distributively, by collaborating organically to the production of a belief.[122] The former generalises over instances of beliefs formed in established groups such as juries, committees, institutions and so on, and rely on the sociological principle of acceptance of common norms or sanctions. So for instance, according to the Joint Acceptance Account (or JAA), we have a genuine group belief when the European Commission representatives agree that the member states will halve the CO2 emissions by 2025, and their agreement is conditional on the acceptance of the other members. The latter, instead, takes as paradigmatic the beliefs formed by organic groups, such as teams, agencies, crews, cooperations. Proponents of the Distributive Model (or DM) argue that genuine group belief is the result of the group members' collaboration, and rely on the sociological principle of division of labour. Take for instance a team of scientists working together: the work is divided among the group members according to their expertise, in such a way that the final belief is the product of their organic cooperation.

It is easy to see that deflationist views will have trouble meeting the Peerhood Constraint. After all, deflationism suggests that the belief of a group *deflates to* the individual beliefs of (some of) its members. This means that, when we compare the beliefs of two groups that are equal on every other respect (i.e., cognitively or evidentially), we are still comparing two unequal sets of beliefs. That is because, according to deflationism, group belief *just is* the sum of individual beliefs (plus some aggregation function, in some formulations). This means that when there are two groups that disagree with each other, the clash between two group beliefs is, in deflationary terms, a clash between two sets of individual beliefs, each constituted by the sum of the individual beliefs of the group members.

From the perspective of deflationism, then, it is hard to see how the two groups can qualify as peers. To see why, note that numbers do matter, epistemically: if one reliable testifier tells me that *p*, while four other reliable testifiers tell me that not-*p*, all else equal, it

[119] The number of individuals that suffices to make up a group belief differs depending on the aggregation function adopted by the group. For instance, in a dictatorial state the belief of the group corresponds to the belief held by a single individual (see List & Pettit 2011).
according to different formulations of deflationism. For instance, if the aggregation function

[120] Take for instance a case where, due to their prejudice, none of the jurors can form the belief that the defendant is innocent. However, based on the evidence brought to light in the trial, they collectively judge that she is innocent.

[121] Gilbert (1987)

[122] Bird (2010)

is intuitive that I should lean towards believing not-*p*. As such, if we reduce group belief to the beliefs of individuals, it is mysterious how the Peerhood Constraint can be met.

Inflationism, on the other hand, seems, at first glance, to fare better than deflationism on this score. Inflationists take group belief to be irreducible to individual belief. For them, it is by relying on some distinctive principle of composition (joint acceptance or organic labour) that the group members *collectively* (i.e., *as one epistemic agent*) form a belief. So, while for deflationists the believing subjects are as many as the *believers* in each group, for inflationists they are as many as the *groups* involved in the disagreement, irrespectively of the group-size. As a result, all else equal, on an inflationist reading, beliefs formed by minority groups won't be considered epistemically inferior to majoritarian ones simply by virtue of being backed by an inferior number of believers.

However, on a closer look, not just *any* inflationist account will do the work. To see this, recall that, in RACIST COMMITTEE, the group of the parents don't file a collective complaint, but rather each family raises the issue with the school individually. Here, you have an example of disagreement between a formalised group – the committee - and a mere aggregate (the sum of individual parents). If our account doesn't recognise that different group-types can host genuine group beliefs, it will also fail to recognise that such groups can be epistemic peers on the matter at hand. On the Joint Acceptance account, for instance, since the parents do not get together to 'shake hands' on the issue, they don't count as being a believing group to begin with. As such, an account that cannot accommodate aggregates delivers the result that what is at stake in RACIST COMMITTEE is, once more, a series of disagreements between a group and separate individuals. It is easy to see how the peerhood relation might not obtain under such circumstances: after all, it seems intuitively right that, if I disagree with my entire group of friends on a topic of common expertise, it is I that should lower my credence in the relevant proposition. Clearly, however, it must be possible to recognise minorities that do not form established groups (either because their structure isn't sufficiently sophisticated or because they are not recognised to be such) as peers. What we are looking for, then, is an inflationist account that is versatile enough to accommodate different types of groups.

In previous work, one of us has developed a functionalist view of the nature of group belief (Miragoli manuscript). In a nutshell, Group Belief Functionalism (henceforth, GBF) defines group belief in terms of the role the belief plays in the agent host. On this view, a group believes something when the belief attributed is individuated via a Ramsey sentence by a set of inputs - e.g., perception or reflection - and outputs - typical corresponding behaviour - that identify the role it occupies in the group host[123]. The principle of composition of such agent (aggregation of individual beliefs, joint commitment or organic labour), then, imposes restrictions on the way in which the role is implemented. As a result, for example, mere aggregates will generate group beliefs via simple belief aggregation, and established and organic groups will do so via more elaborated systems involving some sort of mechanic or organic collaboration among group members.

[123] A ramsey sentence is a sentence that includes a collection of statements that quantify over a variable. In the case of group belief, the variable corresponds to the mental state of the group, and the collection of statements includes terms that refer to external stimuli, other mental states, behaviour, and to causal relations among them.

A special advantage of relying on a functionalist framework is the versatility it affords. GBF licences that beliefs are attributed to each group-type according to the belief forming mechanism that is most suitable to their sociological structure. For example, if the sociological principle of composition of a group is the acceptance of a certain system of norms or sanctions, then GBF allows that such group can naturally form beliefs via the joint acceptance of a common view. On the other hand, where the sociological structure of the group is such that its members are held together by a common goal and the fact that they work together to achieve it, in this case GBF allows that the group will be able to form beliefs via organic collaboration. On this view, it is sometimes the case that a group forms beliefs via a 'deflationist' mechanism, meaning that the main condition the group has to satisfy in order to count as a believing subject is that all group members have the relevant belief. Sometimes, the belief will be formed in an inflationist way, meaning that other more sophisticated conditions will have to be met (i.e., as noted earlier, that all group members jointly commit to the propositions at hand, or that they cooperate organically).

GBF meets the Peerhood Constraint nicely precisely in virtue of its functionalist details. Since it denies the deflationist claim that group belief reduces to the sum of individual beliefs, GBF enjoys the inflationist advantages with respect to the group-size. Furthermore, since it offers a functionalist analysis of group belief, it accommodates multiple realizability, which allows that genuine group beliefs can be formed by the aggregation recipe peculiar to any group-type (aggregates, categories, established and organic groups).

Going back to our examples, then, we can see how GBF gives the right verdict in both cases. As we noted, in SEXIST SCIENTISTS and RACIST COMMITTEE, the belief of the oppressed group was discounted on the grounds that it was formed by a racial or gender minority. According to GBF the doxastic status of a group agent is determined independently of its numerical and sociological characteristics (i.e., the size and the type of the group). As such, granted that the symmetric epistemic conditions are in place, GBF can accommodate our peerhood intuitions in the cases above.

3.2 The Normative Constraint: A Functionalist View of the Epistemology of Disagreement

In previous work, one of us has developed a functionalist account of the normativity of belief in cases of disagreement, the Epistemic Improvement Knowledge Norm of Disagreement (Broncano-Berrocal and Simion 2020, Simion 2019b). In a nutshell, the account looks into what has been left out of the equation so far in the epistemology of disagreement and what, arguably, defines the subject matter: the fact that the doxastic attitudes of disagreeing parties never have the same overall *epistemic status: one of them is right and the other one wrong.* This *fundamental asymmetry* present in all cases of disagreement is an asymmetry concerning evaluative normativity – i.e., how good (epistemically) the doxastic attitudes of the disagreeing parties are. In this way, by accounting for the rational response to disagreement in terms of what all cases of disagreement have in common, the account can easily address all possible cases of disagreement, independently of whether they are instances of peer or everyday disagreement. Indeed, that a given case is a case of peer or everyday disagreement is orthogonal to the distribution of epistemic statuses.

On this view, knowledge is the function of the practice of inquiry. Social epistemic interactions such as disagreements are moves in inquiry, therefore their function is to generate knowledge. If that is the case, in cases of disagreement one should make progress

towards achieving knowledge.

On the Epistemic Improvement Knowledge Norm of Disagreement (EIKND), one should (i) improve the epistemic status of one's doxastic attitude by conciliating if the other party has a doxastic attitude with a better epistemic status and (ii) stick to one's guns if the other party's doxastic attitude has a worse epistemic status. In turn, the quality of the epistemic status at stake is measured against closeness to knowledge: given a value ranking R of epistemic states with respect to proximity to knowledge, in a case of disagreement about whether p, where, after having registered the disagreement, by believing p, S is in epistemic state E_1 and, by believing not-p, H is in epistemic state E_2, S should conciliate if and only if E1 is lesser than E2 on R and hold steadfast iff E1 is better than E2 on R. The view has several crucial advantages over extant views in the disagreement literature, e.g.: a. it accounts for the epistemic significance of disagreement as a social practice, i.e. its conduciveness to knowledge; b. it straightforwardly applies to everyday disagreement rather than to idealised, perfect-peer disagreement cases, and thus does not face the transition problem exemplified above.

It is easy to see that the view will also give the right results in the cases of gender and race group discrimination we are looking at: by stipulation, both of the above cases are cases in which the asymmetry in epistemic status favours the oppressed groups: the epistemic status of their beliefs is closer to knowledge that the epistemic status of the beliefs of their oppressors. After all, by stipulation, the opressed groups are wrong about the matter at hand. As such, in these cases, EIKND delivers the right result that the oppressors should conciliate in order to improve the epistemic status of their beliefs.

CONCLUSION

This paper has put forward a two-tiered functionalist account of group peer disagreement. This strategy is primarily made possible by a radical methodological shift: *contra* extant accounts, that rely on internalist notions of epistemic peerhood and belief permissibility, we have advanced an externalist approach motivated by cases of epistemic injustice in group peer disagreement (SEXSIST SCIENTISTS and RACIST COMMITTEE). We have shown that such cases set two desiderata (what we called the Peerhood and Normative Constraint) that can be elegantly met by appealing to a functionalist view of group belief (GBF) and group justification (EIKND). GBF guarantees that minority groups are considered epistemic peers despite the social prejudices to which they are systematically subject in real cases of disagreement. EIKND, in turn, provides the normative framework to evaluate the conduct of the disagreeing parties and to recognise instances of epistemic injustice.

Chapter six

A Final Word on Hinge Epistemology?

ABSTRACT

Hinge epistemology's main claim to fame lies with its purported advantages in dealing with the problem of radical scepticism. In this paper I argue that the *framework reading*, one of its most promising formulations, is unsuccessful. In a nutshell, the framework reading argues that the system of our rational evaluation is essentially *local* —i.e., resting on a set of arational propositions —hinges— that constitute the limits and the conditions of validity of our epistemic practices. The discussion develops in two main parts. First, I show that, unless important clarifications are made, the framework theory is incapable of offering any solace against the problem of radical scepticism. I then present two ways in which framework theorists may want to clarify their view —following lines of argument found in Coliva (2015) and Pritchard (2016)— but find them both wanting. To the extent that hinge epistemology represents one of the most relevant options available to internalists to avoid sceptical collapse, the results of this discussion contribute to cast a grim light on the chances of a successful defence of internalist epistemic justification more in general.

INTRODUCTION

Hardly anyone is a sceptic. Yet, sceptical arguments are very popular. How so? What is so bewitching about scepticism? One way of answering these questions is to think of sceptical arguments as picking up on features of our epistemic relationship to the world that lead us to question its very nature and scope. In its simplest, most crude version, the problem of scepticism can be framed as starting from two main questions about *how* we know and *what* we know, and then proceeding to show that our epistemic practices are fundamentally faulty, and that we ultimately don't know many, or perhaps any, of the things we normally think we do. To give a very rough idea[124]: External World scepticism (EWSK) is the sceptical challenge that compels us to endorse the latter conclusion —that is, that we don't know any of the *facts* we ordinarily take ourselves to know or believe justifiably about the world. Pyrrhonian scepticism (or PSK), on the other hand, is that strand of scepticism

[124] I will have more to say about these sceptical challenges in the next section.

casting doubt not on the facts we take ourselves to know, but on the *way* (i.e., justification) we come to know them[125].

Hinge epistemology —roughly, the revisionary view that the system of our rational evaluation sits on a set of (more or less) fixed propositions known as *hinges*— stands out as an anti-sceptical strategy in that it promises a *charitable* and *unitary* treatment for both incarnation of the sceptical problem. It is unitary because it promises to soothe our worries with respect to our mundane knowledge of ordinary empirical propositions *while at the same time* making sure that the epistemic tools we use to obtain this knowledge are valid tools. And it is charitable because it promises to confront the sceptic by granting them some key sceptical assumptions about the structure of epistemic justification[126]. Since the pull towards scepticism is often motivated, in large part, by the strong intuitive plausibility of its starting assumptions, hinge epistemologists' promise to win the sceptical challenge on the sceptic's ground constitutes the main appeal for this view.

For the most part, in fact, contemporary solutions to sceptical worries involve recourse to principles that override the sceptical challenge. Epistemic externalism[127] is a good example of this. Some of its proponents believe that both EWSK and PSK can be set aside if we abandon the inherently internalist picture of epistemic justification that seems to underwrite their challenge. However, even those who are willing to accept that externalism does, by and large, get it right, are often recalcitrant to accept the idea that we can (or indeed we should) get rid of internalistically conceived notions of evidence and justification *tout court*. Indeed, the thought that internalistically conceived reasons do play a non-negligible role in our epistemic lives, as well as the accompanying sentiment that purely externalist solutions to the sceptical problem beg the question against scepticism, are both still relatively widespread[128]. Because it promises to carve out a space for an internalist notion of knowledge and justification that is impermeable to sceptical worries, then, hinge epistemologists' undercutting strategy offers a uniquely appealing solution to the sceptical problem.

So hinge epistemology truly aspires to be the panacea to the gamut of our sceptical worries. Can it achieve its aspirations? The answer I give in this paper will be negative, and I motivate it in the following steps: in §1 I give the set up —I quickly present a popular version of hinge epistemology (what I call, following Coliva, the framework reading) and break down its solution to the sceptical problems. This section concludes by showing how commitment to the framework reading threatens sceptical collapse, and it is followed up in

[125] Given my formulation, this variety of sceptical concern may better be identified as 'Agrippan' scepticism, a subspecies of Pyrrhonian scepticism. Since it won't play a crucial role in my argument, I will gloss over this distinction in the rest of the paper.

[126] That is, the broadly internalist and foundationalist conception of epistemic justification.

[127] Here I am referring in particular to epistemic externalism about the nature of epistemic justification, of the sort that knowledge requires. Alvin Goldman's much praised and harshly criticised early formulation of reliabilism (1979) is perhaps the best example of the kind of 'simpleminded' epistemic externalism I have in mind here.

[128] This form of 'gonzo' externalism (Brandom 1995) has been harshly criticised, both by internalists (e.g., Brandom 1994, 1995, McDowell 1995, Bonjour (1985) Fumerton 1998, Wright 2005, Conee and Feldman 2001) and externalists (e.g., Pritchard 2016).

the succeeding two sections by a critical review of the two main ways framework theorists might attempt the rescue. Ultimately, however, neither is found to be satisfactory: broadening the notion of rationality, on the one hand, leads to a commitment to a particularly bad form of relativism (§2) and, on the other, leaning into Pyrrhonian scepticism simply fails to offer a genuine solution to the problem (§3).

I. THE FRAMEWORK READING

The title of "hinge epistemology" is often used somewhat broadly to include a constellation of anti-sceptical strategies inspired by Wittgenstein's remarks in On Certainty, which developed into a wide variety of neighbouring epistemological views —from Wright (1985, 2004) and Williams' (1991) entitlement (or epistemic) views, Conant (1998) and Strawson's (1985) therapeutic accounts, as well as Coliva (2015, 2016), Pritchard (2016) and Moyal-Sharrock's (2004, 2005) non-epistemic views[129]. Very roughly, this family of views can be seen as sprouting from the various attempts made at drawing the full richness of thought inspired by the leading intuition that:

> […] the questions that we raise and our doubts depend upon the fact that some propositions are exempt from doubt, are as it were like *hinges* on which those turn. (OC 341) (my emphasis).

From this shared point of departure, different hinge epistemologies can distinguished based on the way they characterise (*a*) the features they attribute to these special 'hinge propositions' (or *hinges* for short), and (*b*) the role they take hinges to play with respect to the rational system of which they are part. In what follows I will be primarily concerned with one of the most popular formulations of hinge epistemology to date, known as the 'framework reading'. Proponents of this view, popularised by influential work by Coliva (2015) and Pritchard (2016), can be seen to converge over the two following two claims:

> *Non-Epistemicity Thesis* (NET): Hinges are very general, *sui generis* propositions that are not the target of epistemic appraisal.
>
> *Basic Hinge Commitment* (BHC): The system of rational evaluation is relative to a set of (more or less) fixed assumptions, called hinges.

The NET is supposed to capture the framework theorists' commitment to the idea that hinges (which are taken to be very general, non-empirical commitments such as that 'I am not radically deceived', 'I am not a brain in a vat', 'there are physical objects', 'my perceptual faculties are generally reliable' and so on) are rationally *inert* —i.e., that they are 'visceral', 'non-optional' commitments that are not responsive to positive *or negative* rational evaluation (like justification or doubt). The BHC, on the other hand, captures the framework theorists' broad foundationalist commitment about the structure of rational evaluation.

[129] For a more exhaustive taxonomy, see Coliva (2016)

Combining the NET and BHC together, we obtain the following concise summary of the main commitment of the framework theory:

> *Locality:* Our system of rational evaluation is grounded on a set of fundamental commitments that are not themselves the target of epistemic appraisal.[130]

In summary, then, the framework theory is a particular version of hinge epistemology that takes hinges to be the (*a*) non-epistemically assessable (*b*) ground of our rational system. The picture of the structure of our rational evaluation that the framework reading invites us to buy, then, is one where a set of propositions, hinges, are, at one time the *ground* and the *limit* of our rational system. They are the ground in virtue of being those basic assumptions from which our belief system draws its rational validity; and they are its limit in virtue of lying outside epistemic appraisal.

Now, how does the framework theorists' commitment to the Locality claim help them offer the unitary and undercutting response to the problem of scepticism they promised? Start with EWSK. The core challenge of this form of scepticism is directed at *what* we can sensibly claim to have justification or knowledge of, and it is about the *scope* or *boundaries* of our epistemic practices. To carry out this challenge, the Cartesian sceptic typically begins conjuring radical sceptical scenarios where nothing is as it seems, and where the gap between appearance and reality appears insurmountable (e.g., think of Matrix-style scenarios, evil demon possibilities, dream-like experiences and so forth). On the basis of the ineliminability of these sceptical hypotheses (SHs), a Cartesian-style type of scepticism about the external world can be raised that makes leverage on the closure principle (or CP)[131] for knowledge or justification in the following way:

> *The External-World sceptical argument from closure*
> (P1) We are not justified in believing that SHs do not obtain (¬SH)
> (P2) If I am justified in believing E, then I am justified in believing that ¬SH[132]
> ★ I am not justified in believing E[133]

[130] Commitment to locality is offered more explicitly by Pritchard (2016), and endorsed by Coliva (2020, 9). Endorsement of locality is what sets non-epistemic readings of hinge epistemology, like Pritchard and Coliva, apart from epistemic ones (e.g., Wright 2004), according to which hinges *are* proper subject of epistemic appraisal.

[131] The principle of closure formalises the idea shared by epistemologists that knowledge and justification can be extended via (competent) deduction. A standard formulation of the closure principle (for justification) goes like this: "CPJ: If one is justified in believing that P and competently deduces Q from P, thereby coming to believe Q while retaining their justification for P, then one is justified in believing that Q."

[132] Assuming (as it seems plausible) that E entails ¬SH, if one accepts the general principle of closure, it follows that, if one is justified in believing E, then one is thereby justified in believing what is entailed by E (i.e., ¬SH) —provided that one competently draws the inference from E and ¬SH.

[133] Given that knowledge requires justified belief, this argument naturally entails lack of knowledge of ordinary empirical propositions. Although this argument is inspired by Pritchard's (2016) own reconstruction of the external-world sceptical challenge, I phrased the argument as focussing on justification rather than knowledge because it is issues surrounding the former, more so than the latter, that I will focus on in this paper.

Because it can virtually be applied to any empirical proposition we take ourselves to know or believe justifiably, this argument provides a powerful, universal challenge to our knowledge of the external world.

In some sense, the advantage that endorsing the framework reading can give with respect to scepticism of the external world is clear. The EWSK challenge relies on the closure principle, which is about extending one's justification via competent deduction. Proponents of the framework reading start from this observation to derive the following condition for the applicability of the CPJ:

> *Condition for the applicability of CPJ* Given two propositions P and Q, where P entails Q, it is possible for a subject S to extend (via competent deduction) their justification from P to Q only if Q can be epistemically appraised by S.

This is intuitive: after all, how would it be possible to extend one's justification via inference if the entailed proposition wasn't available for epistemic assessment? In such a case, the CPJ wouldn't be applicable. This is a trivial point, but the framework reading draws a surprising conclusion from this. Crucially, a consequence of buying into the framework view is that the inference that goes from our justification of everyday propositions to our justification of the denial of the sceptical hypotheses (e.g., that I am not radically deceived) is precisely one where the consequent (i.e., a hinge proposition) is not in the market for epistemic appraisal. It follows then that the CPJ cannot be applied to the couple of propositions (i.e., E and the denial of SH) that make up the EWSK.

By making a distinction between the denial of closure and its failure of application, then, the framework reading can meet the challenge by appealing to the unique features of hinge propositions. At the cost of buying into a revisionary story about the structure of our rational evaluation, then, we can obtain a powerful anti-sceptical argument that allows us to retain our justification and knowledge of the external world in spite of our inability to know the denial of the sceptical hypotheses without, with this, denying closure

So far so good. But what about the other sceptical challenge? Recall that PSK is not (primarily, at least) interested in undermining our grasp of empirical facts —instead, its doubt is cast on the methods we use to grasp them in the first place. In other words, the PSK challenge is directed at whether we can sensibly claim to possess a valid way of rationalising our beliefs, and is thus concerned with their *standard of validity*.

The challenge gets started from the observation that epistemic justification is a matter of believing on good grounds, and that there can be such a thing as a justified belief. On this basis, the Pyrrhonian sceptic carries out the following considerations about rational support: if we take our beliefs to be justifiable, then they are justified either independently of reasoning (i.e., they are *basic*) or by another belief (i.e., they are justified *inferentially*). The Pyrrhonian sceptic is not convinced that there could be any basic belief, since they would either be arbitrarily picked or, if not, and a reason why that set of beliefs should be basic could be given, they would not be basic after all. When we attempt to justify a belief by means of another belief, however, we find that the support ultimately turns out to either be circular (*mode of circularity*) or terminate in an unjustified belief (*mode of hypothesis*) or continue

endlessly (*mode of infinite regress*). Given that none of these 'modes' to end inquiry into the grounds of our beliefs is acceptable, the sceptic concludes, there is no valid way we can provide justification for our beliefs.

Very roughly, we can identify three main claims from which the Pyrrhonian sceptic mount their challenge:

> *Cliffordian Principle* Justification is a matter of believing on good grounds.
>
> *Pyrrhonian Demand* There are only inferentially justified beliefs (i.e., there are no basic beliefs).
>
> *Agrippean Modes* No belief is justified by (1) an infinite chain of reasons (*mode of infinite regress*), (2) a circular chain of reasons (*mode of circularity*) or (3) one terminating in an unjustified belief (*mode of hypothesis*),

where the latter two (the Pyrrhonian Demand and the Agrippan Modes) can be seen as specifying when the grounds mentioned in the former (Cliffordian Principle) are *good* grounds. Hence, the Pyrrhonian challenge can be roughly characterised as setting the following desideratum that, according to Pyrrhonian sceptics, cannot be satisfied:

> *Pyrrhonian Desideratum* In order for our justificatory practices to be valid, it must be possible to find a non-arbitrary ground for our beliefs capable of offering support that is not circular or leading to an infinite regress.

Can the framework reading help us deal with this challenge? The beginning of an answer here would notice that by endorsing Locality, the framework reading is committed to a response to PSK that is foundational, but only *in spirit*. The response is foundationalist to the extent that it aims to resist the challenge by stopping the regress —hinges, according to the framework reading, are "regress-stoppers", the solid foundation on which our rational system is built. But it is nonetheless a foundationalism *sui generis*, since standard foundationalist views normally take the grounds to be themselves *rational*[134].

In fact, while foundationalist views normally offer overriding responses to PSK by arguing that there *are* some beliefs that are basically justified (i.e., by denying the Pyrrhonian Demand), the framework reading on the other hand concedes that justification is a matter of believing on good grounds. The problem, at this point, is how to understand exactly the kind of response one such *sui generis* foundationalism aims to provide to the Pyrrhonian challenge. For notice that, if the solution offered by the framework reading simply *is* that such propositions ought to be taken for granted, then this is tantamount to

[134] Here rational should be interpreted broadly to give space to internalist or externalist notions of rationality. This is to give space to externalist versions of foundationalism according to which, e.g., basic beliefs are justified in virtue of being produced by a reliable mechanism.

embracing the Agrippan *mode of hypothesis*[135]. Clearly, however, this would not be a very desirable result. Not only because it would consist in a capitulation, more than a solution, to Pyrrhonian scepticism —which in itself is a rather bad result for any self-proclaimed anti-sceptical strategy. Even worse, this would give the sceptic a new way of reinstantiating scepticism of the external world —the one that targets justification and knowledge of everyday empirical beliefs.

How so? Assume, with the Pyrrhonian sceptic, that the only route for a belief to be rationally grounded is via inference from another belief (recall the Pyrrhonian Demand). This means that a belief B is rationally grounded only if it rests on some ground, G. As per the Pyrrhonian Desideratum, however, B won't ultimately be rationally grounded unless G, its ground, is itself justified by another belief, G*. Iterating this reasoning within the foundationalist structure the framework theorists endorses, we have that any everyday empirical belief B is ultimately grounded on hinge propositions, which are, by definition, rationally unresponsive. Since the ultimate ground of any empirical belief is not rationally held, then, and since a belief ought to sit on a rational ground in order to be rational itself, none of our empirical beliefs would be rationally grounded.

You can put it this way: the framework reading shares with foundationalism a picture of the rational validity of the system of our beliefs as flowing upstream from the rationality of its grounds. At the same time, however, since it is a *sui generis* form of foundationalism, it doesn't enjoy the resources available to standard foundationalist views of drawing rational force from basic justified beliefs —hinges are the ground, yes, but they are *not* rationally responsive[136]. And from the groundlessness of hinges, nothing prevents one from inferring the groundlessness of all the beliefs resting on them, including ordinary empirical propositions about the external world.

This is a problem the framework theorist must avoid. I will call it The Problem, and reconstruct it thus:

> *The Problem*
> (P1) Hinges are the ground of our belief system (*from commitment to the framework reading*)
> (P2) The only way a belief can be rationally grounded is via inference from another belief (and in such a way that the support is not circular, leading to an infinite regress or resting on an unsupported assumption) (*from the Pyrrhonian Demand*)
> (P3) Hinges are assumed without support, and are not themselves inferentially grounded (*from commitment to the framework reading*).
> ★ Our beliefs are not rationally grounded

[135] Coliva herself acknowledges this point when she says that "If this were the situation, since we can provide neither immediate nor mediate justifications for these propositions, it would seem that the skeptical outcome would ensue. That is to say, it would seem that the only plausible alternative would be to hold that these are just a-rational assumptions and that, even if we think we are justified in believing ordinary empirical propositions, we are not." (2020, 12). This ensuing scepticism, which is embedded in The Problem I introduce here, is what Coliva refers to as 'Humean scepticism'.

[136] Note that this is compatible with Coliva's claim that hinges are rational since, for her, hinge's rationality is not to be intended in terms of their rational responsiveness.

Since premise 1 and 3 come with acceptance of the framework reading, there are only two possible ways its proponents can address the Problem: deny premise 2, and show that hinges can be rationally held, even if not in the way envisaged by the Pyrrhonian Demand; or deny the validity of the argument, and show how our belief system can be rationally valid despite the arationality of hinges. In the next sections I will address each option in turn.

II. EXTENDING RATIONALITY

Denying premise 2 allows proponents of the framework reading to respond to the problem by arguing that, despite lacking inferential justification, and thus being in *this* sense arational, commitment to hinges is still somehow rationally compelling. But how? One way would be to take hinge propositions to be basically justified and give a response to PSK in a broadly foundational spirit. This option, however, leads to familiar problems about identifying what exactly would make hinges justified —problems that would be complicated by the fact that standard foundationalist strategies normally take experience to provide basic justification, whether in a primitive or non-primitive fashion, while framework theorists take hinges to be *non-empirical*. Despite being about the world, like common empirical propositions, framework theorists think hinges differ fundamentally from empirical propositions in that there is no experience that has as its content the claim "that there is an external world", or "that I am not being radically deceived".

A solution then could be to extend the notion of rationality to include not only propositions that enjoy some kind of support (whether inferential or non-inferential), but also hinges. The only problem is, of course, that such a move may appear irritatingly *ad hoc*. The framework theorist has made up a new class of propositions ('hinges') that *it just so happens* to occupy such a unique role that it licences a fundamental restructuring of our rational system in order to include them as part of it. Why should we buy that?

One way to motivate this move has recently been defended by Annalisa Coliva in a series of influential works (2015, 2016, 2020). Her general idea is to take hinges to be rational not because they draw their justification from other beliefs or from other facts, but in virtue of the role they play in the very structure of our rational system. To see this more clearly, consider an analogy with chess: the game of chess includes chess pieces (kings, bishops, pawns, queens and so on) and a board where the pieces move following fixed patterns. These patterns are defined by specific rules —i.e., the instructions that determine the moves a player can make at a particular time. Now, it is natural to think that the rules of the game of chess are an integral part of the game itself —with board and pieces, we take them to *constitute* what we call 'the game of chess'. Proponents of this strategy take hinge propositions to play a similar role that rules play in the game of chess (or any game really) —they are the instructions that make it possible to determine the practice of forming, assessing, and withdrawing from empirical beliefs. Like rules are constitutive of games, hinges are constitutive of our system of rational evaluation.

Following this line of thought, Coliva motivates a fundamental restructuring of our rational system to include, in addition to inferentially supported beliefs, also those assumptions in virtue of which such support is possible in the first place. In this way,

proponents of the framework reading can arguably have their cake and eat it too: they can preserve the sense in which hinges are not responsive to rational consideration (since there is no *supporting reason* in virtue of which they are rational) while at the same time proposing a way hinges can still be considered rationally held (that is, because they are constitutive of the notion of rationality)[137]. Because it does the former, the framework theorist is capable of retaining the advantage of their position with respect to the closure-based challenge; and because it does the latter, the Pyrrhonian regress can be blocked in a way that preserves the rational validity of the system as a whole. If this is correct, and so long as one is disposed to endorse a suitable (i.e., constitutive) interpretation of the rationality of hinges, then, there seems to be a way to prevent the radical collapse of the rational validity threatened by the Problem.

Now, should we be disposed to endorse it? Here's a reason why we shouldn't: if hinges are mandated by the structure of our epistemic practices, would it not be possible for there to be systems similar to ours in structure, but different in content? After all, it is a fact of *any* game that they include, in addition to some kind of prop (even imaginary), particular rules that define the boundaries of actions within that game. According to this proposal, everything that is needed in order to stipulate the rationality of hinge propositions, is a sufficiently stable system of 'rational' evaluation. For any such system, like for any game, it will be possible to identify a particular set of rules that governs it.

For instance, Imagine a community of people with a system of rational evaluation that works very broadly like ours —very roughly, say, their members would form and reject beliefs on the basis of reasons, and measure their reliability as informants by attributing to each other knowledge (or lack thereof). Surely, since the rational structure would be the same, both our and their system would be equally rationally valid. But would their beliefs necessarily be the same as ours? And even if they did, couldn't they have relied on other methods of forming these same beliefs? In fact, there seems to be no reason to think this ought to be the case. But if there isn't, and if our belief system is rationally valid just as much as theirs, then isn't this view opening the doors to a particularly ruinous form of epistemic relativism?

Maybe. But let's first get clear on what the relativist charge consists of more exactly. I can see two main ways in which epistemic practices with the same overall structure can differ from each other with respect to their content: they can differ with respect to the content of *the beliefs* that are justified within each system, and the content of the *basic*[138] *methods* employed to justify them[139]. For instance, two different epistemic systems may be

[137] Wouldn't this undermine the "non-epistemicity" of the general framework reading? Yes and no. It would, insofar as hinges, according to this view, are *in some sense* rational. For this reason, Coliva takes her own view to be ultimately *epistemic*. However, the grounds on which we have established the non-epistemicity of hinges is on their rational unresponsiveness —i.e., their being not rational in a standard sense. Since the kind of rationality that this reading ascribes to hinges does not impinge that notion of rationality, I don't see why it should undermine the general strategy they employ to block the closure-based sceptical challenge.

[138] According to Coliva, a basic method of justification is a method that "is at the core of all human life given the kind of creatures we are" (2015, 128). More specifically, this will be the case when "it does not presuppose other instances of itself and is necessary for other epistemic practices" (141).

[139] For simplicity, I subsume 'epistemic principles or rules' (á la Boghossian 2006) under "basic methods".

constituted by the same set of empirical beliefs (that the Earth revolves around the sun, cats don't grow on trees, a cloudless sky looks blue and so on), and yet, members of one system may have formed most of their beliefs via observation and testimony, and members of the other by tarot reading or by consulting sacred texts or the stars. On the other hand, we can imagine systems where, although beliefs are formed through the same methods (for simplicity, although they need not), the beliefs thus formed, and the hinges on which they rely, are radically different from ours. For instance, a group with our hinges would believe that there is a cat when they see a cat in front of them. In the same circumstance, members of a group accepting sceptical hinges (i.e., that we are radically deceived) would instead believe that there is only the appearance of a cat, and members of a group that believes that the world is a manifestation of the Eternal Red Tortoise would believe that what they saw was a cat-looking manifestation of said tortoise.

Now, the problem is that a view that allows epistemic systems of these sorts to be rationally valid would run into the problem of allowing (a) justification for beliefs we wouldn't otherwise take to be justified (i.e., that there is just a mere appearance of a cat, rather than a real cat, when we see one); and (b) methods for justifying beliefs we would intuitively take to be inappropriate (i.e., tarot reading) to be appropriate. A proponent of the framework reading that wants to resist the charge of relativism, then, must explain two things: first, why it is not possible, according to their view, for the methods we use to justify our beliefs to justify different and incompatible beliefs; second, why it is not possible for beliefs to be justified by different and unacceptable methods.

One way of doing the latter would be to argue that there is only one set of methods that can be legitimately used to rationalise beliefs, and it's ours. How so? An obvious way could be to say that any rational enquirer, like us, ultimately forms their beliefs on the basis of the deliverance of their senses. Coliva suggests something along these lines when she says that "if we think about creatures like human beings, or of creatures who are relevantly similar to human beings, it is hard to see how else they could form beliefs about material objects if not through the deliverances of their senses" (2015, 144)[140].

This may be a good reason, but it needs some refining, since it seems to give too much importance to perceptual justification. Even granted that testimonial justification *could somehow* be ultimately reduced to perceptual justification (and this itself is a not trivial point to be established, and a highly controversial one[141]), it is not clear how one would go about doing that with inference. So maybe a more cautious claim would be to say that all (humanoid) rational inquirers base their beliefs either on their senses or their rational faculties alone.

[140] This point is similar in spirit to Boghossian's absolutism (2006) (and more in particular, with his endorsement of what he calls the 'universality principle', according to which "There is a uniquely correct set of epistemic principles, which all rational agents are bound by"). The question about whether a proponent of this strategy can endorse absolutism, however, is what's at stake here.

[141] Reductionism (very roughly, the view that one's belief is justified via testimony only if one has other positive reasons in favour of the testimony —say, about the reliability of the testifier) and its denial (Anti-Reductionism) in the epistemology of testimony are hotly debated positions, and neither can be unreflectively taken to be true. For a useful introduction to the debate, see Nick Leonard's 2023 SEP entry on the "Epistemological Problems of Testimony".

But is that so? Take for instance astrology, and imagine that there are some groups of people whose everyday beliefs are grounded, for the most part, on information obtained by consulting celestial bodies[142]. Their method may well be not very reliable, and their beliefs may ultimately turn out to be unjustified. Still, their rational system, we can imagine, is roughly structured like ours: they, like us, search for reasons to support their beliefs, and are able to give them upon being asked. The only difference is that while we base our beliefs about the weather by looking at the sky, say, they do so by looking at the position of the planets. Why wouldn't it be possible to even conceive of some such community?

The framework theorists could say that astrology is not a *basic* method for justifying a belief[143]. Someone who relied on it must have used *their eyes* to consult the planets, and possibly *infer* from what they saw following basic rules of logic. And the same would go for other funky practices as well, like tarot reading and the like. This, however, won't be entirely satisfactory. For it doesn't follow from the fact that some justificatory practices often function only if coupled with others, that they aren't epistemically basic. Testimony is made possible by the fact that we have eyes and ears to read and listen, but it would be quite of a stretch to draw from this the conclusion that testimony can be reduced to perception —not in any straightforward way at least[144]. Similarly, I have to use my eyes to infer, based on my doctor's average consultation time and the people I see around me, that I will have to wait at least another hour before being received. And yet, in this case, we say that my belief is justified via inference, not by perception, nor that for this reason perception is basic and inference isn't. The fact that looking (and, in some cases, inferring) is often a step in the process that leads to justifying a belief does not mean that it is perception (or inference) *alone* that ultimately justifies the belief. When a person from that remote community comes to believe that they are going to have a bad Monday because their Jupiter is in Pluto, a truly nefarious occurrence, their belief is not justified by their having seen (or read) that it is so —or at least, not only, for they could have seen or read about Jupiter and still not formed the belief. What makes the belief justified for them is having learned about the position of Jupiter *coupled with* their taking astrology to be an adequate means for rationally forming their beliefs.

In fact, that different methods *can* be used (for the purpose of this argument, it doesn't matter how legitimately) to justify a belief is not just a possibility, but a reality. In his famous discussion of Galileo and Bellarmine disagreement on 'world systems', Rorty makes this point indubitable. The position he defends, as well as the very way in which Bellarmine thinks and acts, and his general image of the world, notes Rorty, reveals his conviction that the Bible is a genuine source of evidence, and that forming beliefs based on the scripture is a reliable way of forming justified beliefs and obtain secure knowledge of the world. Whether Bellarmine's belief is ultimately justified or not, the Bible does constitute for him (as well as for the majority of people at that time, and a non-negligible part of the world today too) a legitimate method for justifying his belief. What else, other than acting and thinking in accordance with the truth of a norm is required for one to

[142] The plausibility of scenarios of this sort are typically reinforced by appealing to cases discussed in anthropological studies. For instance, some such recurring example in the philosophical literature is that of the Azande tribe, whose members' belief system is argued to differ in radical ways from those of Western communities.

[143] This is precisely what Coliva attempts to do in her (2015, §4.4).

[144] see footnote 16.

accept it? This case, like many similar ones, shows that very often it is not just perception and inference (if at all) but commitments to other methods as well (such as Bellarmine's commitment to the canonical interpretation of the Bible) that are used to justify the beliefs of members of those epistemic communities where such methods are taken as legitimate epistemic practices.

Even so, this will at best prove that our epistemic practices (more specifically, perception and inference) are not always sufficient for justifying beliefs. Still, a point could be made that, since perception is often a key ingredient for belief justification, members of this faraway community would not end up forming beliefs that are substantially different from ours, and so the relativist threat would not be a serious one.

But this doesn't seem to be right either. First of all, there doesn't seem to be any in principle reason why a stars-consulting community couldn't form the funkiest of the beliefs, especially when it comes to fundamental matters (like the nature of the world, its constitution, and so on). After all, without even stretching our imagination too wide, it is possible to find communities that have held (and still do) very odd beliefs on very imaginative grounds[145]. Second, and more important, the necessity of perception and inference is tied to an unjustifiably chauvinistic conception of rational inquiry. For why think that all rational inquirers *ought to* be like us and justify beliefs the way we do? Wouldn't it be possible to at least *conceive of* alternative epistemic practices? One example could be that of a God endowed with a divine intuition which allows her to grasp truths about her environment without the use of the senses. But even without the recourse to divine powers, we could imagine alien creatures (who may as well be, all in all, ultimately not that different from us) who enjoy a felicitous attunement with their environment, so much so that facts that are relevant for them in a particular circumstance would cause them to form (defeasibly) true beliefs about them. What could a proponent of the framework reading do to block this line of thought? Sure, it may not be too likely that there *is* some race of alien creatures with such powers. But this isn't enough to block this relativist threat.

So much then for epistemic practices intended as *the methods* by which we justify our beliefs. Now, what about the possibility that the same methods justify different and incompatible beliefs? Could (say) our perceptual experience justify different and incompatible beliefs from ours? More precisely: could, say, my perceptual experience as of a cup justify not only ordinary beliefs about the presence of that cup but also, say, the sceptical belief about the mere appearance of a cup? Presumably, the answer to this question will depend on one's preferred view about justification. Internalist views will generally (although not necessarily) be more permissive than externalist ones with respect to the content of the beliefs that can be justified. Take beliefs justified by perception. One option could be to pick and defend a strongly externalist view of perception that takes it to directly put us in contact with the world, so that the possibility of odd sceptical beliefs about virtual hands being justified would be ruled out. This might seem quite *ad hoc*, but theoretically feasible (provided that the framework theorist's commitment to an undercutting solution could be maintained).

[145] To mention one for all: according to the Prince Philip Movement, a cargo cult of the Yaohnanen tribe, Prince Philip, the husband of Queen Elizabeth II, is the reincarnation of the ancient son of a mountain spirit.

A problem with this, however, is that, according to this view, hinge commitments play a central role with respect to the justificatory status of individual beliefs. Recall that, independently of one's preferred internalist or externalist constraints on justification, framework theorists take hinges as the condition of possibility of justification. Referring to perceptual justification in particular, for example, Coliva writes that:

> [p]erceptual [...] warrants depend for their obtainment on two ingredients: an experience with a given phenomenal and representational content *together with the assumption of some very general proposition*, such as 'There is an external world', 'My sense organs are generally reliable" [...] and possibly other ones'. (Coliva 2013, 249) (my emphasis).

The idea then is that, irrespectively of the particular view (in this case, of perception) adopted by the framework theorist, what can be justified will ultimately depend on the hinges one is committed to. For imagine you have an experience as of a cup in front of you. For this experience to provide a justification (a perceptual warrant) for the belief that there is a cup, it must be the case that you also are assuming that there is an external world out there made out of, among other things, that very cup you are seeing —and possibly that you aren't radically deceived too —or so the thought goes. If you *didn't* assume these very general propositions (if you were committed to the thought that, say, perception is a grand illusion) then you would *not* have justification for that belief but, perhaps, *another* —the one, say, that there is a very elaborated virtual reconstruction of a cup in your head. If so, then, what would prevent members of epistemic communities grounded on radically different hinges from justifying radically different beliefs and incompatible with ours?

Maybe proponents of this view could take a hard line and say that *there could not be* different hinges[146]. The structure of our rational evaluation is such that it guarantees not only that there are some propositions that must be assumed in order for the rational system to be in place, but that such propositions are necessarily the same *we* also assume —*our* hinges, say, that there is an external world, and that we are not radically deceived. More precisely, the thought here is that the assumption of *our* hinge commitments is the precondition for *any* meaningful reason-giving practice. That is, our rational system is not only valid, but also *unique*. Following this thought, then, one could say that denying our hinges, like the sceptic does, is a self-defeating move, since no meaningful move can be made within the system that denies the very conditions of its possibility. In other words, it would just be impossible. But why think so? The key intuition of this strategy is that the structure of our rational practices mandates that *some* propositions must be taken for granted. From this, it would indeed be tempting to infer that the propositions that are to be taken for granted must necessarily coincide with our fundamental commitments (i.e., that there is an external world, that we are not radically deceived and so on). But why *should* it be the case? Crispin Wright, pointing out a potential flaw of his own view, put this point very clearly:

[146] This is a similar route to the one taken by Sankey (1997) and (2010), Boghossian (2006), and Rachel (1999).

> if the most favourable light that can be cast on our acceptance of a material world, or other minds, consists in argument that [...] any system of rational objective thought has to incorporate some conception of the kind of stuff that inhabits other cognitive localities—then we seem to have no claim to the objective correctness of the most fundamental categories of substance that we actually employ. More, there will be no obstacle in principle to the idea of alternative, equally valid ways of conceiving the substance of the world [...]. What are the barriers to an entitlement to wood spirits, ectoplasm, gods and a plethora of existing but non-actual spatio-temporally unrelated concrete possible worlds?

The transcendental consideration that we engage in the practice of giving and asking for reasons concerns the *structure* of epistemic evaluations —what Wittgenstein calls "the *logic* of our investigations"[147]. Precisely because it concerns its logic, however, this proposal is neutral with respect to the content of the system thus constituted. The observation that hinges play, with respect to the rational system of evaluation, the same role that rules play in games doesn't tell us anything about the content of the rules and the kinds of moves that are thus made possible.

In fact, there seems to be nothing about systems of rational evaluation resting on radically different hinges that would contradict the 'logic' of our epistemic practices. We *can* conceive of some such system based on the sceptical hinge commitment that, say, there is no external world and we are radically deceived —indeed, we can imagine a community, call them Sceptics, where its members, when presented with a chair-like object, do not form beliefs about chair-looking physical objects, but rather about, say, mere appearance of a chair (or about a 'virtual' chair). Like us, Sceptics would go about in their everyday epistemic businesses forming beliefs about things around them (with the difference that they would take themselves to be surrounded by virtual objects) and, upon being questioned, they would give the same perceptual or testimonial reason we would also give.
But let's grant that, for some reason, it would indeed be self-defeating to assume some such Sceptical hinge commitments. Even so, and more compellingly, we can still conceive of systems based on non-Sceptical hinge commitments that are nonetheless so radically different from our own to raise serious relativistic worries. For instance, we can imagine a community that takes reality to be a manifestation of a spiritual creature, what they call the Eternal Red Tortoise. Unlike members of a belief system based on commitments similar to ours (who, upon seeing a chair-like object, will form beliefs about chairs) members of one such epistemic community will form beliefs about, say, chair-looking parts of the eternal Red Tortoise. Is it impossible to imagine this within the constraints of this view? I don't see why. On the contrary, the framework theorist doesn't seem to be in a position to explain how such beliefs could not be justified. If so, then, and more in general, the framework theorist's attempt to block the Problem by responding to the PSK challenge along foundationalist lines leads inevitably to endorsing some kind of epistemic relativism.

[147] Coliva's move is inspired by the following often quoted passage of Wittgenstein's On Certainty: "That is to say, it belongs to the *logic of our scientific investigations* that certain things are indeed not doubted." (OC 341) (emphasis added).

If the relativist threat cannot be dodged, though, why dodge? Just as externalists take the sceptic's fault to lie in their assumption of an internalist epistemology, it is open to proponents of this strategy to suggest that in order to overcome scepticism we need to overcome its assumption of an absolutist (in the sense of non-relativistic) epistemology. Even granted that there was a good and honest formulation of epistemic relativism[148], though, it is not at all clear that it would correspond to the one this view mandates commitment to. For one thing, based on what we have just said, it appears clear that taking hinges to be constitutive of epistemic rationality offers no guarantee against the possibility of radically different and incompatible rational systems. If that's true, lots of beliefs would be granted the same rational standing as ours, not only those formed in epistemically dubious ways (like, for instance, by consulting the stars) but also, and more worryingly, beliefs with potentially harmful content (like those formed in racist, fascist or sexist communities). Now, I don't doubt that those who are already sympathetic to epistemic relativism may not find in the framework reading's commitment to relativism a reason against it. On the other hand, I am sure that, even among relativism-friendly epistemologists, very few will want to accept the particularly radical form of epistemic relativism the framework theory would licence. Before embarking in its defence, then, I suspect that framework theorists will prefer exploring other ways to solve the Problem.

III. SURRENDER TO THE ANGST

Recall the Problem: on the one hand, we have the Pyrrhonian Demand that, in order to be rational, a belief must be supported by another belief. On the other hand, we have the framework theorist's commitment to a *sui generis* foundationalism according to which a belief's chain of support ultimately bottoms down on a set of fixed propositions that are not rationally responsive. The Problem then asks: how can we preserve the rationality of our belief system if its validity rests on the rationality of its grounds?

In the previous section, I've looked at an attempt to steer clear from the Problem by questioning the Pyrrhonian Demand and argue that there are other ways in which beliefs (and, in particular, hinges) can be said to enjoy some kind of rational standing (i.e., by being constitutive of rationality). But that move appears to lend itself to a characterisation of rationality that is relativistic in a bad way. So what else can the framework theorist do to avoid the Problem?

Another option could be to endorse the Pyrrhonian Demand, admit that, ultimately, our rational system *is* groundless, but at the same time resist the temptation to conclude from

[148] Not all forms of relativism have the obvious problematic consequences of the extremist variety discussed here. For defences of relatively more palatable relativisms (pun intended) see Kusch (2002), MacFarlane (2003) or (but this may be more controversial) Rorty (1979) (for useful mappings of relativism see also Boghossian 2006 and Coliva and Baghramian 2020).

this that our everyday empirical beliefs are not justified[149]. After all, the idea goes, if we could preserve the rationality of (nearly) all empirical beliefs *individually*, the arationality of our belief system *as a whole* wouldn't be that much of a Problem. For instance, imagine it was possible to distinguish between a rational evaluation of an *deflationary* variety, which takes the rationality of the system to be tightly connected to the rationality of the beliefs that are a part of it, on the one hand, and of a *inflationary* variety on the other, where the validity of a system is taken 'as a whole', 'floating freely' from the actual rational standing of each individual belief. If this was possible, one could argue, then the rationality of each and everyone of my ordinary, empirical beliefs *could* be saved in spite of the ultimate groundlessness of the system as a whole.

Indeed, hinge epistemologists' commitment to offer a unitary solution to the problem of scepticism, whereby hinges are seen both as grounds (of the system of rational evaluation as a whole) and limit (of the things that can be known), puts some pressure on the adoption of a rational evaluation of the former, deflationary, kind. When this is assumed, however, there is little one can do (*pace* Coliva) to stop the arationality of hinges to corrode the validity of the whole system and, with it, all the individual beliefs that form part of it. Arguably, this is precisely where the Problem comes from: once that (from a deflationary stance) the rational validity of each and every belief is taken to flow upwards from the validity of their ultimate ground, the rational unresponsiveness of hinges must necessarily undermine the rational standing of the entire belief system that rests on them, and with it each and every individual belief that is part of it[150].

But things may not be this way. On an inflationary view, for instance, the arationality of hinges would undermine the validity of the system as a whole, yes, but this wouldn't necessarily affect the rational standing of each and every individual belief, since their rational status and that of the system would be independent from each other —or so the thought would go. According to Pritchard, this result is, in fact, a natural consequence of the framework theorist's commitment to Locality. Because hinges are exempt from epistemic appraisal, any attempt to ground ordinary empirical beliefs on them is doomed to failure more or less in the way the sceptic's attempt to infer the lack of groundedness of our everyday empirical beliefs from the lack of groundedness of the denial of sceptical hypotheses is also, from a hinge theoretical point of view, fundamentally misguided. This is

[149] One may worry that this move will raise relativistic worries again. For if we concede that the rational system is, as a whole, groundless, couldn't there be radically divergent hinge systems, with radically divergent and equally rational beliefs attached to them? Maybe, although I am not sure that proponents of this view would not have anything to say in response to this worry. In any case, I will not be interested in this problem here. Partly, this is because this line of criticism would not differ enough from the one just pursued to justify a separate treatment. Most importantly, however, this is because I believe this particular variation of the framework reading raises new and interesting problems of its own.

[150] This is a form of what has come to be known as the "leaching problem" (Wright 2004), and that underwrites The Problem. For Wright (although not in Wright's terms) leaching can be contained thanks to the hinge epistemologist's commitment to Locality, which grants that there are special cases where the closure principle can't be applied —i.e., when entailment to propositions that aren't in the market for knowledge is involved. In a nutshell, this response consists in a renounce to drawing the rational force of our belief system from the rationality of its grounds. The obvious problem with Wright's proposed solution is that it does not explain what other source of rationality is left for the bulk of our empirical beliefs. This is precisely the problem that both the view discussed and the one I will address in this section aim to solve.

what commitment to an image of rational evaluation as essentially *local* implies —i.e., it implies that the rational validity of our empirical beliefs does not flow upwards all the way from the very bottom of the system as a whole (i.e., from hinges) but, say, from other, more 'local' sources of support[151].

If this is correct, and an inflationary view of rationality is indeed compatible with the framework reading, it could be possible to solve the Problem by rejecting the connection between the lack of groundedness of our belief system and that of our individual beliefs and argue that, although hinges' arationality does ultimately undermine the rational validity of the system as a whole, *provided one had good reasons for them individually*, each and every individual belief could still be rationally grounded.

But this is easier said than done. For one, this is because, when it suggests to trade a deflationary rational evaluation for an inflationary one, the framework reading loses much of its initial appeal —which, if you recall, rested on its promise to offer a treatment of scepticism capable of alleviating both sides of the problem at once. This was made possible by the commitment of the framework reading to a unique image of the structure of our rational evaluation as one where the ultimate source of validity of the system and its boundaries coincide in one convenient place: hinge propositions. In making a distinction between deflationary *vs* inflationary rational support, however, this strategy narrows down the anti-sceptical power of hinges, in that it relieves them from their role as the source of rational validity for individual empirical beliefs.

But maybe the scope of hinge epistemology need not be that grandiose after all. Let's grant that. Are the framework theorist's chances of getting it right at least with respect to this narrower goal any better now? I don't think so. Here's briefly why: according to inflationism about rational evaluation, at a local level —i.e., when considering the rational standing of some everyday belief— the rational standing of a belief (about, say, the fact that the moon is full) must be independent of the ultimate ground on which it stands (according to the framework theorist, the hinge that, say, we are not radically deceived in our perceptual judgements). In other words, it must be only the local perceptual support (say, that I see that the moon is full) that rationalises the belief. However, as per classical sceptical scenarios, we can imagine a situation where an epistemic subject is in the exactly same internal state as ours but, unluckily, finds themselves in an epistemically 'bad' environment —they could be a brain in a vat, for instance. Arguably, then, all their perception can justify is, at best, beliefs about appearance —such as that it looks as though the moon is full. However, if this is plausible, whichever rational support we (in the good case) have for our belief is also available to the subject in the bad case (since, by stipulation, the two subjects are internally identical). But if the evidence in the two cases is the same, and it can at most justify beliefs about appearances, how can we claim to have evidence for the ordinary belief that there *is* a full moon today (and not the mere appearance of one)? More exactly, this point can be made by relying on the following:

[151] Here is a passage where Pritchard attempts to motivate just this thought: "[t]ake my putative rationally grounded knowledge that the car I drive is dark blue. That this is so", he claims, "is entirely compatible with my being radically and fundamentally in error in my beliefs" (2015, 98).

Underdetermination Principle (UP): If one cannot have favouring support[152] for P over Q (where P and Q describe incompatible alternatives), then one is not justified in believing that P.

In this case, the UP can be employed to show that, since the evidence locally available to a subject in the good scenario (us) is the same as that available to a subject in the bad scenario, and the latter can only justify appearance beliefs, then there is no evidence available to us that can favour our empirical belief about the moon over, say, the belief that there is only the appearance of a full moon. For each and every individual empirical belief, then, radical EWSK would be reinstated once again:

External-World Sceptical Argument from Underdetermination
(P1) We do not have favouring reason for P over Q
(P2) If one cannot have favouring support for P over Q (where P and Q describe incompatible alternatives), then one is not justified in believing P
★ I am not justified in believing P

In response, it could be pointed out that the soundness of the argument threads on an internalist conception of local support. Consider the following:

Sameness of Evidence Lemma (SEL) A subject's perceptual evidence (*e*) is compatible with the obtaining of sceptical scenarios (SH) where *e* is false.

The plausibility of SEL rests on a broadly internalist notion of evidence, whereby evidence supervenes exclusively on facts internal to the subject (namely, on the subject's mental states). If this is true, it would suffice to replace such notion of evidence with an externalist one to get rid of SEL, and with it the problem that triggers the underdetermination principle. In fact, so long as we have externalistically conceived reasons, since these aren't available to the subject in the bad scenario, our reasons *do* favour our belief, and this is so irrespectively of whether one's experiences in the two cases are subjectively indistinguishable —or so the thought would go.

Notice however that not any kind of externalism will do (if any), as the framework theorist is under some pressure to accept *at least some* internalist constraint on the notion of justification. In part, this is because not doing so may swamp the plausibility of the framework reading itself, given that externalism offers its own way out of scepticism, and a much more straightforward one. But mainly it is because buying into a purely externalist shortcut would undermine a strong motivation for the view itself. If you recall, this motivation rests on the fact that the framework theorists promised to offer an undercutting response to scepticism capable of bringing internalism home safe and, possibly, of carving

[152] The notion of "favouring support" was introduced by Duncan Pritchard (2012) in opposition to that of "discriminating support", which is possessed when one has the relevant 'discriminating capacities'. Crucially, *favouring* support for proposition *p* may be available to one even if one ultimately lacks the capacity to discriminate whether *p*. Roughly, this is because, even in the absence of this capacity, one may still have more evidence for *p* than against it. The notion of favouring support is useful in that it provides a "minimal" threshold that must be met for one to be able to have support for a proposition (*p*) over some incompatible alternative (*q*).

out a sceptic-free conception of *internalist* rational support (no less). In light of this, a minimal internalistic requirement for this strategy could be phrased as follows:

> *Subjective Condition Demand* (SCD): In order for X to count as that which contributes to the rationalisation of a belief B for subject S, it ought to be the case that X's role (as that which contributes to the rationalisation of B) depends in some important way on S's subjective perspective,

where some X will 'depend in some important way' on S's subjective perspective if, for instance, and roughly, X is reflectively accessible to S or if it supervenes on facts internal to S's mind (broad access internalist or mentalist lines). With this requirement in place, we can then formulate a UP that embeds this SCD as follows:

> UP_{SCD} If one cannot have internalistically conceived favouring support for P over Q (where P and Q describe incompatible alternatives), then one is not justified (in an internalistic way) in believing that P[153].

Naturally, UP_{SCD} licences the construction of a sceptical challenge similar to the one above, and one that the framework theorists cannot ignore, given their commitment to SCD. Notice however how SCD (and consequently UP_{SCD} as well) is compatible with the idea that some other, non-internalist reason may play a role in supporting our ordinary beliefs. That's precisely what makes it a minimal requirement —i.e., so long as it is possible to reserve *some* space for internalistically conceived reasons, the SCD is happy to concede, that'll do.

This may be good news, for it seems to open the door to the possibility of combining together externalist and internalist strategies for epistemic justification. Imagine for instance we could have externalist support that was also rationally grounded: being externalist, it would elude SEL by granting favouring support to epistemic agents in the good case, since it wouldn't be available to those in the bad case. Being rationally grounded, on the other hand, it would guarantee that this favouring support is also cashed out in internalist-friendly terms, thus offering the kind of response to the sceptic that the framework theorist is committed to giving. In his (2015), Duncan Pritchard motivates precisely one such idea by embedding it in his disjunctivist treatment of underdetermination-based radical scepticism. According to Epistemological Disjunctivism (ED), agents in the good and bad cases do not possess the same evidence because perception in the good case provides them with *reflectively accessible factive support* —namely, a kind of externalistically conceived support that the perceiver can also access reflectively.

But how does this work more precisely? Recall that the SCD asks that our individual empirical beliefs are supported also by internalistically conceived reasons. For this strategy to work, then, it must be the case that the support enjoyed by my empirical belief (say, the belief that *p*: "there is a tree in front of me") isn't provided only by local external facts (like,

[153] This formulation of the UP is due to Pritchard (2016, 29)

say, that "I see *p*") but also by some other internalist condition. One straightforward way of satisfying this requirement (and the one Pritchard proposes) is to cash out this internalistically conceived condition in simple access internalist terms. Roughly, according to access internalism one is justified in believing *p* if one can have reflective access to one's basis for believing that *p*. Overall, then, according to this strategy, in paradigmatic cases of perception we can have local support of our empirical beliefs because: (1) we have a factive reason for them —namely, the fact that we *see that p*, which entails the truth of *p*—, (2) we have reflective access to this factive reason that satisfies the internalist condition.

Combining these requirements in a coherent disjunctivist view of epistemic justification has proven to be exceptionally difficult[154]. Important problems have been raised concerning, for instance, the basing relationship between the empirical belief and its factive support, as well as the precise nature of the access relationship. Partly due to these problems, ED is now an unpopular view among epistemologists. In what follows, however, I want to draw attention to a different problem from the ones that are normally raised to specifically target the plausibility of ED: I want to argue that even if ED *were* capable of offering the kind of hybrid support that is needed to satisfy SCD and avoid SEL, it would not be able, as part of the broader anti-sceptical programme proposed by the framework theorist, to avoid the view's relapse into scepticism.

To see this, let's first take a closer look at the access requirement in (2). A natural way to conceive of reflective access is in terms of a subject's awareness of an object (read: a fact, a reason) *as* something that is relevant to or contributes to the justification of the relevant belief. The reason for this is clear, and I take it is part of what motivates this strand of internalism about epistemic justification in the first place —roughly, the thought that reasons are precisely those things that *we can conceive of as that which justifies our beliefs*[155]. Following Bergmann (2006), let's call this version of access internalism *strong access internalism* (or SAI). According to a strong access internalist version of ED, then, claim (2) ought to be understood as maintaining that, when we have reflective access to a factive reason for *p*, we have reflective access to this reason *as a basis* for our belief that *p*.

The problem with strong access internalism is that it is vulnerable to generating familiar Pyrrhonian worries. How so? By the light of SAI, in order to count as a reason for my believing that *p*, the factive support that "I see that *p*", must be coupled with my awareness of it *as a basis* for my belief that *p*. But naturally, my awareness of this factive support is itself susceptible to being evaluated with respect to its justification status[156]. For how can I be sure (say) that this awareness is genuine, and I am not confused? Or: can I rule out the possibility that I wasn't utterly irrational or insane in conceiving of that support as a basis for my belief? Crucially, answers to this question will have to meet the standard imposed by

[154] Some critics of epistemic disjunctivism include Fratantonio (2019), Zalabardo (2015), Ashton (2015), Ranalli (2014), Dennis (2014) and Ghijsen (2015).

[155] Here I use 'conceive of' rather liberally, as broadly synonymical to 'be aware of', and encompassing both conceptual / nonconceptual and doxastic / non-doxastic connotations. For a more nuanced discussion, see Bergmann (2006).

[156] Note that this is independent of doxastic or non-doxastic conceptions of 'awareness'. After all, not only of doxastic states can we legitimately demand that they be justified. For a more detailed defence of the problems non-doxastic versions of strong access internalism run into (as well as a distinction between actual and potential varieties), see Bergmann (2006).

the Pyrrhonian Demand —i.e., not be arbitrary, circular, or leading to an infinite regress. Can proponents of this strategy satisfy the demand?

I think we should give a negative answer to this question. For what would a good ground look like in this case? On the one hand, proponents of this strategy cannot appeal to hinges, for this would establish a direct grounding relationship between individual empirical beliefs and hinges, which would undermine the inflationist commitment that opened up the way for this solution in the first place. On the other hand, however, nothing short of hinges would do, or a more straightforward way out of the Pyrrhonian worry would be available that would undermine the core appeal of the framework theory —or of any hinge epistemology really.

So maybe the proponent of this strategy may want to opt for a weaker form of access internalism, one that need not commit to conceiving (whether conceptually or not) of the factive support as being in some way relevant to the justification of the belief. *Weak access internalism*[157] (WAI) rejects the claim that one's awareness of that which justifies a belief ought necessarily to involve a conception of it *as* that which does that job. So for a proponent of this strategy that opts for WAI, although one may have some conception of their factive support for their belief that *p*, they are not aware of such support *as* something which contributes to the justification status of that belief.

The obvious problem for framework theorists that choose this strategy, however, is that the thought motivating SAI constitutes a main advantage of internalist views of justification compared to externalist ones, and the crucial motivation for imposing an awareness requirement in the first place (recall the SCD). For weak access internalists, someone who bases their belief that *p* on good grounds, since they are not aware of such ground as a basis for that belief, would be in the exact same subjective state concerning the status of their belief as someone for whom the justifiedness of their belief was a mere accident. Crucially, however, it is precisely because of their ability to account for the intuitive difference between these two cases that internalist accounts are usually taken to offer an advantage over competing externalist accounts.

A main selling point of the framework reading is that it promises to offer a solution to scepticism that doesn't sacrifice key internalist insights that put internalist theories of epistemic justification in a position of advantage over externalist ones. By adopting a weak version of access internalism, proponents of this strategy give up precisely one of these benefits, thus undermining what motivated the main appeal of the hinge project in the first place. Otherwise one may wonder: if it wasn't already an explanatory goal of this view to make good of the advantages of internalist views of epistemic justification, why bother with this project in the first place, given that externalism already offers a solution to scepticism, and a much more straightforward one?

In brief, framework theorists who choose to rely on the virtues of epistemological disjunctivism in order to steer clear of the underdetermination-based sceptical worry without giving up the minimal internalist requirement are confronted with two infelicitous alternatives: endorse SAI and face the same Pyrrhonian puzzle they attempted to escape, or opt for WAI, and undermine what motivated the choice of their strategy in the first place.

[157] Moser (1985, 1989) and Fumerton (1995)

As a last resort, one may wonder whether proponents of this strategy could not be tempted to propose to embrace Pyrrhonian scepticism. Clearly, this would require some qualification, since a simpleminded commitment to Pyrrhonian scepticism can hardly count as a way to defend this strategy. Here's a way in which they could offer to motivate this move: to begin with, one could notice that Pyrrhonian scepticism isn't so bad after all. When properly understood, it could be argued, Pyrrhonian scepticism is not a sceptical argument *per se*, although it can be phrased as such. Instead, it can also be understood more as a *stance*, an attitude we may assume when examining the rational status of our beliefs, the doubting attitude that resists both commitment to and rejection of any proposition —a sort of reasoned suspension of judgement.

Understood this way, an important feature of this attitude (this stance) is that it is not always appropriate to assume it. For instance, in our ordinary exchanges, we don't. We don't normally go around asking why people believe what they do —if someone tells us they met our common friend down the road, heading to the park, we believe them, as we should. Sure, we may question whether they actually saw what they saw, perhaps because we have defeaters —that is, reasons against their testimony, like the fact that we know that our common friend went to Aberdeen (of all places!). Still, our questioning in this case is very different from the much deeper-cutting Pyrrhonian questioning stance, which isn't directed at the perceptual testimony but rather at perception itself as a method for justifying our beliefs. But the point is that we simply *don't do this in an everyday setting*. Indeed, that's what makes them everyday settings. On this reading of the Pyrrhonian stance (due mainly to Fogelin (1994), and endorsed by Pritchard (2011)): "ordinary belief is excluded from the [Pyrrhonian] skeptical challenge because it doesn't aspire to have the rational basis which the Pyrrhonian skeptical techniques are targeted at undermining" (Pritchard 2011, 8).

Even if it is true that my belief that p, which is inferentially supported by the reflectively accessible fact that I see that p, is *not* ultimately grounded, then, this does not undermine the validity of the inferential support *within the ordinary context in which I have reflective access to it*. And this is because, according to this line of argument, in that ordinary context, the Pyrrhonian worry is not legitimately brought up. So, insofar as what one is interested in is just ordinary beliefs and their grounds (as proponents of this strategy are, given their commitment to inflationism about rational evaluation) then one does have rationally grounded knowledge of them.

From a dialectical point of view, the implausibility of this move is, I think, obvious. In a nutshell, proponents of this strategy are asking us to avoid Pyrrhonian scepticism by surrendering to a version of it. First, they have offered us to pay the huge price of committing the validity of our rational system as a whole to the pit of Pyrrhonian scepticism (and to buying an unpopular view of the structure of justification, ED) in order to save the justificatory status of our individual beliefs; to do that, however, they then ask us to adopt a Pyrrnohian stance.

Even if one wanted to grant the dialectical plausibility of this move, it seems difficult to get rid of the general feeling that, once so much is conceded to scepticism, a view doesn't have the right to present itself as offering a *solution* to scepticism. Even if this strategy did offer a solution to closure-based sceptical worries about the boundaries of our knowledge, I wonder who, if anyone, would be willing to sacrifice the validity of our system of rational evaluation for it.

But endorsing the Pyrrhonian stance doesn't only pose dialectical worries. Instead, commiting to the Pyrrhonian stance forces a dilemma to proponents of this strategy that has no happy solution. To see this, notice that, unlike standard philosophical views, this stance doesn't necessarily involve commitment to a particular structure of rational evaluation, or about its ultimate validity —indeed, it has been described as consisting in a universal suspension of judgement that resists acceptance as well as rejection of *any* proposition. It is more like a doubting stance that, applied piecemeal, *can*, whenever adopted, call into question the grounds of our beliefs individually. More precisely, according to Pritchard:

> [the Pyrrhonian stance] consists of skeptical techniques (modes) which can be applied, piecemeal, to a wide range of beliefs, but it does not [...] attempt to call a large body of our beliefs into question en masse. (Pritchard 2011, 7)

In fact, it is precisely because of these characteristics that this attitude is inappropriate in the context of everyday beliefs —because we do not adopt it in our everyday life. The problem however is that, whether or not proponents of this strategy do endorse some such reading of Pyrrhonian scepticism, they are nonetheless committed to a particular Pyrrhonian *position* —that is, a philosophical position that occupies a particular place with respect to the Pyrrhonian challenge. More exactly, they endorse a *sui generis* sort of foundationalism that takes hinge propositions to be their arational grounds. This, I noted, leads to the endorsement of the Agrippean mode of hypothesis. Which means that, according to the framework theory (or at least, the version supported by proponents of this strategy), rational evaluation *does* ultimately reach a bottom, which *has* particular features. And it is in virtue of their commitment to this position, and independently of their adoption of the Pyrrhonian stance, that individual empirical beliefs are ultimately not rationally grounded. For it is a logical consequence of the rational unresponsiveness of hinges that whatever is grounded on them (each and every individual belief) will not be (inferentially) justified. And so long as a belief is ultimately resting on insecure ground it is *not* rationally grounded.

So the Pyrrhonian *stance* and this *position* are in stark contrast with each other: one cannot consistently endorse both. One cannot hold a commitment to a particular form of rational evaluation while at the same time be neutral about it. One cannot hold that hinges are the arational ground of our belief system and at the same time suspend judgement about it. Endorsing one comes at the cost of giving up the other. Which one should the framework theorist give up then? The Pyrrhonian stance might relieve the framework theorist from the challenge posed by the Problem but, since it would come at the cost of abandoning any

claim about the rationality of hinges (or hinges in general), it would potentially reopen the closure-based challenge. So that won't do. Sticking to the Pyrrhonian position would alleviate the latter issue, but it would still leave open the Problem, since the resources of the Pyrrhonian stance would not be available. In fact, this doesn't seem a dilemma framework theorists could ever come clear about.

CONCLUDING REMARKS

The framework reading commits to a peculiar picture of the structure of rationality, one where the grounds and the boundaries of rational investigation happen to coincide, and are expressed by a set of extra-rational assumptions: hinges. What motivates the main appeal of this revisionary picture is not only the fact that it purports to offer a unitary solution to scepticism in both its two main incarnations —as a challenge to rationally grounded knowledge of the external world, and as a challenge to the validity of our epistemic practices; but also, that it aims to do this without sacrificing fundamental internalistic insights about the nature of justification.

But if I am right, the framework reading is itself susceptible to very serious sceptical worries, arising precisely from the peculiar nature of hinge propositions. I have argued that there are two main ways in which framework theorists can overcome this threat: to extend the notion of rationality to include hinge propositions, or to surrender to the idea that the system of our rational evaluation is, as a whole, groundless. The first strategy may be able to successfully salvage an internalistic conception of justification, but it does so at the cost of committing to an implausible restructuring of the notion of rational evaluation. The other strategy, on the other hand, may be able to provide some solace from one formulation of the sceptical threat, but only if it renounces to preserve internalist insights about epistemic justification. Since this was a major advantage of the framework theory compared to competing solutions of the sceptical problem, however, this strategy undermines a crucial motivation for the view itself.

Bibliography

Adam, A. (1998) *Artificial Knowing: Gender and the Thinking Machine*, Routledge, ISBN 9780415129633

Alcoff, L. M. (2007), "Epistemologies of Ignorance: Three types", in S. Sullivan and N. Tuana (eds.), *Race and Epistemologies of Ignorance*, 39–58. Albany, NY: State University of New York Press.

Alcoff, L. M. (2010), "Epistemic Identities", *Episteme*, 7 (2):128-137.

Alexander, J., Betz, D., Gonnerman, C. & Waterman, J. P. (2018). "Framing how we think about disagreement". *Philosophical Studies* 175: 2539-66.

Alston, W. (1988), "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2: 257–99

Anderson, E. (1995), "The Democratic University: The Role of Justice in the Production of Knowledge" *Social Philosophy and Policy* VOL. 12 (2):186-219

Anderson, E. (1995), "Knowledge, Human Interests, and Objectivity in Feminist Epistemology", *Philosophical Topics*, VOL. 23 No.2.

Anderson, E. (1999), "What Is The Point of Equality" *Ethics*, 109: 287–337.

Anderson, E. (2012) "Epistemic Justice as a Virtue of Social Institutions", *Social Epistemology*, Vol 26, No. 2, p. 163-173

Anderson, E. (2012), "Epistemic Justice As A Virtue Of Social Institutions", *Social Epistemology*, 26 (2):163-173.

Angwin, J. & Mattu, S. & Larson J. & Kirchner, L. (23rd of May 2016). "Machine Bias" *ProPublica*, online source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ashton, N. (2015) "Undercutting Underdetermination-Based Scepticism". *Theoria*, 81 (4), pp. 333-354. .

Backes, M. (2021) "Can groups be genuine believers? The Argument From Interpretationism." *Synthese* 199, 10311–10329.

Bagenstos, S. R. (2006). "The Structural Turn And The Limits Of Antidiscrimination Law" *California Law Review*, 94(1), 1–47.

Bailey, A. (2017) "Tracking Privilege-Preserving Epistemic Pushback in Feminist and Critical Race Philosophy Classes", *Hypatia*, 32: 876–92.

Beddor, B. (2015). "Process Reliabilism's Troubles With Defeat". *The Philosophical Quarterly*, 65(259): 145–59.

Bender, E. M., Gebru, T., et al. (2021), "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. March. pp. 610–623,

Bergmann, M. (2009). "Rational Disagreement After Full Disclosure". *Episteme* 6: 336-53.

Bird, A. (2019), "Group Belief". In *The Routledge Handbook of Social Epistemology* (eds. Miranda Fricker, Peter Graham, David Henderson, and Nikolaj Pedersen).

Bird, A. (2010) "Social Knowing: The Social Sense of 'Scientific Knowledge'". *Philosophical Perspectives*, 24: 23–56.

Bogardus, T. (2009). "A vindication of the equal-weight view". *Episteme* 6: 324-35.

Boghossian, P. (2006), *Fear of Knowledge: Against Relativism and Constructivism*, Clarendon Press. Pennsylvania State University.

Brandom, R. (1994), *Making it Explicit*. Cambridge, Mass.: Harvard University Press.

Brandom, R. (1995), "Knowledge and the Social Articulation of the Space of Reason", in *Philosophy and Phenomenological Research*, Dec., 1995, Vol. 55, No. 4

Bratman, M. E. (1984), "Two Faces of Intention". *The Philosophical Review*, Vol. 93, No. 3, pp. 375-405

Bratman, M. E. (2006). "Dynamics of Sociality". In *Midwest Studies in Philosophy*. 30(1), 1–15. doi:10.1111/j.1475-4975.2006.00125.x.

Bratman, M. E. (2009), "Shared Agency". In Chrysostomos Mantzavinos (ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*. Cambridge University Press. pp. 41--59.

Bratman, M. E., (1999), *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

Broncano-Berrocal, F., and Simion, M. (2020). "Disagreement and Epistemic Improvement" Manuscript.

Brook, A. (2005) *Cognition and the Brain: The Philosophy and Neuroscience Movement*, ed. Brook & Akins. Cambridge University Press.

Browne, S. (2015), *Dark Matters*, Duke University Press.

Carnap, R. (1950) "Empiricism, Semantics, and Ontology," in Paul Moser and Arnold Nat, *Human Knowledge* Oxford University Press. (2003).

Cave, S. & Dihal, K. (2020), "The Whiteness of AI", *Philosophy & Technology* 33, pp. 685–703.

Chisholm, R.M. (1977). *Theory of Knowledge*, 2nd edition, Englewood Cliffs: Prentice-Hall.

Chisholm, R.M. (1973). *The Problem of the Criterion*. Milwaukee, WI: Marquette University Press.

Chrisman, M. (2008): "Ought to Believe", *Journal of Philosophy*, 105, pp. 346–370.

Chrisman, M. (2018) "Epistemic Normativity and Cognitive Agency" *Nous*, 52:3, pp. 508–529.

Chrisman, M. (2020), "Believing As We Ought And The Democratic Route To Knowledge" in *The Ethics of Belief and Beyond: Understanding Mental Normativity*, Sebastian Schmidt and Gerhard Ernst ed., Routledge, NY.

Chrisman, M. (2022), *Belief, Agency, and Knowledge*: Essays on Epistemic Normativity, Oxford University Press, UK.

Christensen, D. (2007). "Epistemology of disagreement: The good news", *The Philosophical Review* 116: 187-217.

Christensen, D. (2009). "Disagreement as evidence: The epistemology of controversy", *Philosophy Compass* 4: 756-67.

Christensen, D. (2013). "Epistemic modesty defended", In D. Christensen & J, Lackey (eds.), *The epistemology of disagreement: New essays*. Oxford University Press.

Clifford, W. (1877), "The Ethics of Belief", in *Contemporary Review*, 29: pp. 290–309.

Coliva, A & Baghramian, M. (2020). *Relativism*, Routledge London.

Coliva, A. (2013), "Moderatism, Transmission Failures, Closure and Humean Skepticism" in Dodd, Dylan, and Elia Zardini (eds), *Scepticism and Perceptual Justification* OUP.

Coliva, A. (2015), *Extended Epistemology*, Palgrave Mcmillan

Coliva, A. (2016), "Which Hinge Epistemology?" in *International Journal for The Study of Skepticism*, 6 (2016) 79-96

Coliva, A. (2020), "Skepticism Unhinged", in *Belgrade Philosophical Annual* Vol. 33

Conant, J. (1998). "Wittgenstein on Meaning and Use," *Philosophical Investigations* 21: 222–250.

Conee, E. & Feldman, R. (1985), "Evidentialism." *Philosophical Studies*.

Conee, E. (2009). "Peerage", *Episteme* 6: 313-23.

Conee, E. (2010). "Rational disagreement defended". In R. Feldman & T. A. Warfield (eds.), *Disagreement*. Oxford University Press.

Conee, E. and Feldman, R, (2001) "Internalism Defended." in *American Philosophical Quarterly*, 38(1): 1–18;

Dennis, P. (2014). "Criteria for Indefeasible Knowledge: John McDowell and 'Epistemological Disjunctivism'", *Synthese* 191(17), 4099–113.

Dotson, K., (2011), "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–257. .

Du Bois, W. E. B. (1989). *The Souls of Black Folk*. Orig. ed. 1903. New York: Penguin.

Dular, N. (2021), "Mansplaining as Epistemic Injustice", *Feminist Philosophical Quarterly*, Vol. 7, No. 1, article 1.

Durkheim, E (1982), *The Rules of Sociological Method*, W. D. Halls (trans.), Free Press, New York

Eckenweiler, L. (12th June 2019). "Seeking Asylum: Epistemic Injustice and Humanitarian Testimonies", in *Justice in Global Health Emergencies & Humanitarian Crises* (The University of Edinburgh): (accessed Oct 2022)

Elga, A. (2007). "Reflection and disagreement". *Noûs* 4: 478-502.

Elga, A. 2010. "How to disagree about how to disagree". In T. Warfield & R. Feldman (eds.), *Disagreement*. Oxford University Press.

Epstein, B. (2019), "What are social groups? Their metaphysics and how to classify them". *Synthese* 196, 4899–4932 (2019).

Fair Trials (2021). "Automating Injustice" online article sourced on: https://www.fairtrials.org/articles/publications/automating-injustice/

Feldman, R. (2000), "The Ethics of Belief", *Philosophy and Phenomenological Research*, Vol. 60, No. 3, pp. 667-695

Feldman, R. (2005). "Respecting the evidence". *Philosophical Perspectives* 19: 95-119.

Feldman, R. (2007). "Reasonable religious disagreements". In L. Antony (ed.), *Philosophers without gods: meditations on atheism and the secular*. Oxford University Press: 194-214.

Feldman, R. (2008), "Modest Deontologism in Epistemology" *Synthese*, 161:339–355.

Feldman, R. (2009). "Evidentialism, higher-order evidence, and disagreement". *Episteme* 6: 294-312.

Feldman, R. (2014). "Evidence of evidence is evidence". In J. Matheson and R. Vitz. *The ethics of belief*. Oxford University Press: 284-99.

Flores, C. & Woodard, E. (2023), "Epistemic Norms on Evidence-Gathering" in *Philosophical Studies*. 162: 165-81

Forsyth, D. R., (2010). *Group Dynamics* (5 ed.). Belmont, CA: Wadsworth, Cengage Learning.

Fratantonio, G. (2021). "Reflective Access, Closure, and Epistemological Disjunctivism". *Episteme*, 18(4), 555-575. doi:10.1017/epi.2019.26

Fricker, M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford: Oxford University Press.

Friedman (2020), "The Epistemic and The Zetetic", *The Philosophical Review* 129 (4):501-536.

Friedman, J. (2013). "Suspended judgement". *Philosophical Studies*, 162: 165-81.

Frye, M. (1983). *The Politics of Reality: Essays in Feminist Theory*. Crossing Press, US

Gandy, O.H. (1998). *Communication and Race: A Structural Perspective*. Edward Arnold and Oxford University Press.

Ghijsen, H. (2015). "The Basis Problem for Epistemological Disjunctivism Revisited", *Erkenntnis* 80(6), 1147–56.

Giere, R. (2002) "Scientific cognition as distributed cognition", in Peter Carruthers, Stephen P. Stich & Michael Siegal (eds.), *The Cognitive Basis of Science*. Cambridge University Press. pp. 285.

Gilbert, M. (1987) "Modelling collective belief" *Synthese* 73: 185–204.

Gilbert, M. (1993). "Agreements, Coercion, and Obligation." *Ethics* 103: 679–706.

Gilbert, M. (1994). "Remarks on Collective Belief," in Frederick F. Schmitt (ed.), *Socializing Epistemology: The Social Dimensions of Knowledge*. Lanham, MD: Rowman & Littlefield, 235–55.

Gilbert, M., & Pilchman, D. (2014). "Belief, Acceptance, and What Happens in Groups: Some Methodological Considerations". In J. Lackey *Essays in Collective Epistemology*. Oxford University Press.

Goldberg, S. (2017), "Should Have Known", *Synthese* Vol 194, p. 2863–2894

Goldberg, S. (2018), *To The Best of Our Knowledge: Social Expectations and Epistemic Normativity*, OUP, Oxford

Goldberg, S. C. (2017). "A Proposed Research Program for Social Epistemology". In P. J. Reider (Ed.), *Social Epistemology and Epistemic Agency: Decentralizing Epistemic Agency* (pp. 3-20). London: Rowman & Littlefield Publishers.

Goldberg, S. C., (2020) *Conversational Pressure: Normativity in Speech Exchanges*. Online edn, Oxford Academic, 20 Aug. 2020), , accessed 21 Aug 2023.

Goldman, A. (1979). "What is justified belief?" In: G. S. Pappas (ed.) Justification and Knowledge. Dordrecht: Reidel: 1–25; reprinted in A. I. Goldman (2012). *Reliabilism and Contemporary Epistemology*. New York: Oxford University Press: 29–49.

Goldman, A. I. & Olsson, E. J. (2009). "Reliabilism and the value of Knowledge". In A. Haddock, A. Millar & D. Pritchard (eds.). *Epistemic Value*. Oxford University Press: 19-41.

Goldman, A. I. (2014), "Social Process Reliabilism: Solving Justification Problems in Collective Epistemology", in Lackey 2014: 11–41. doi:10.1093/acprof:oso/9780199665792.003.0002

Graham, P. and Lyons, J. (2021). "The structure of defeat: Pollock's evidentialism, Lackey's distinction, and prospects for reliabilism". In: Simion and Brown (ed.) *Reasons, Justification and Defeat*. Oxford: Oxford University Press.

Habermas, J. (1968), *Knowledge and Human Interest*. Beacon Press Boston

Hakli, R. (2007), "On the Possibility of Group Knowledge without Belief", *Social Epistemology*, 21:3, 249-266, doi: 10.1080/02691720701685581

Hao, K (2018). "What is machine learning?" *MIT Technology Review*, sourced online at https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/

Hao, K (2020). "We read the paper that forced Timnit Gebru out of Google. Here's what it says." *MIT Technology Review*, sourced online at https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

Haslanger, S. (2007), "'But Mom, Crop-Tops Are Cute!' Social Knowledge, Social Structure and Ideology Critique" in *Philosophical Issues*, Vol. 17, pp. 70-91.

Hawthorne, J. & Srinivasan, A. (2013). "Disagreement without transparency: Some bleak thoughts". In D. Christensen and J. Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press.

Ho, L. H. (2008). *A Philosophy of Evidence Law*. Oxford University Press

Hoffman, A. L. (2019) "Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse", *Information, Communication & Society*. Vol. 22, Issue 7: Data Justice.

Hookway, C. (2010), "Some Varieties of Epistemic Injustice: Reflection on Fricker", *Episteme*, 7 (2):151-163.

Hornsby, J and Langton R. (1998), "Free Speech and Illocution", *Legal Theory*, 4 (1):21-37.

Huang, Linus Ta-Lun, Hsiang-Yun Chen, Ying-Tung Lin, Tsung-Ren Huang, and Tzu-Wei Hung (2022) "Ameliorating Algorithmic Bias, or Why Explainable AI Needs Feminist Philosophy." *Feminist Philosophy Quarterly*, 8 (3/4).

Huebner, B (2013) *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*. Oxford University Press, USA.

Huges (2023), "Epistemic Feedback Loops (Or: How Not to Get Evidence)" in *Philosophical Studies*, 106 (2):368-393 (2021)

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT press.

Ichikawa, J. J. (2022), "You Ought to Have Known: Positive Epistemic Norms in a Knowledge-First Framework" *Synthese*, 200 (5):1-23

James, W. (1896): "The Will to Believe", in: *The Will to Believe: And Other Essays in Popular Philosophy and Human Immortality*, Dover: Dover Publishing 1956, 1–31.

Johnson, R.C. (2020), "Mansplaining and Illocutionary Force", *Feminist Philosophy Quarterly*, Vol. 6, No. 4, article 3.

Jones, L. K. (2020). "Twitter wants you to know that you're still SOL if you get a death threat — unless you're President Donald Trump" access online (Oct 2022) at https://medium.com/@agua.carbonica/twitter-wants-you-to-know-that-youre-still-sol-if-you-get-a-death-threat-unless-you-re-a5cce316b706

Kappel, K. (2019). Bottom up justification, asymmetric epistemic push, and the fragility of higher order justification. *Episteme*, 16 (2):119-138

Keller, E. F. (1985). *Reflections on Gender and Science*. Yale University Press

Kelly, T. (2005). "The epistemic significance of disagreement". In T. Gendler & J. Hawthorne (eds.), *Oxford Studies in Epistemology*, Vol. 1. Oxford University Press: 167-196.

Kelly, T. (2010). "Peer disagreement and higher-order evidence". In R. Feldman & T. A. Warfield (eds.), *Disagreement*. Oxford University Press : 111-74.

Kelly, T. (2013). "How to be an epistemic permissivist". In M. Steup & J. Turri (eds.), *Contemporary debates in epistemology*. Blackwell.

Kelly, T. (2014). "Evidence", in *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/evidence/

Kelp, C. (2022), "Defeat and Proficiencies" in *Philosophical Issues*, Volume 32, Issue 1.

Kelp, C. (2023), *The Nature and Normativity of Defeat*. Cambridge: Cambridge University Press.

Kimmerer, R. W. (2013). *Braiding Sweetgrass*, Milkweed Editions.

King, N. L. (2012). "Disagreement: what's the problem? or: a good peer is hard to find". *Philosophy and Phenomenological Research* 85: 249-72.

Kornblith, H. (1993), "Epistemic Normativity" in *Synthese*, Vol. 94: 357-376.

Kornblith, H. (2001) "Epistemic Obligations and the Possibility of Internalism" in *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, Abrol Fairweather & Linda Zagzebski ed. Oxford University Press, NY.

Kornblith, H. (2010). "Belief in the face of controversy". In R. Feldman & T. A. Warfield (eds.), *Disagreement*. Oxford University Press.

Kornblith, H. (2013). "Is philosophical knowledge possible?" In D. E. Machuca (ed.), *Disagreement and skepticism*. Routledge.

Krieger, L. H. (1995). "The content of our categories: A cognitive bias approach to discrimination and equal employment opportunity" *Stanford Law Review*, 47(6), 1161–1248.

Kusch, M. (2002), *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford: Oxford University Press.

Lackey, J (2020) *The Epistemology of Groups*, Oxford University Press, Oxford.

Lackey, J. (2007), "Norms of Assertion", *Noûs*, Vol. 41, No. 4, pp. 594-626.

Lackey, J. (2008). "What should we do when we disagree?" In T. S. Gendler & J. Hawthorne (eds.), *Oxford studies in epistemology*. Oxford University Press: 274-93.

Lackey, J. (2010). "A justificationist view of disagreement's epistemic significance". In A. Haddock, A. Millar, & D. Pritchard (eds.), *Social epistemology*. Oxford University Press.

Lackey, J. (2014), "Socially Extended Knowledge", *Philosophical Issues* Vol 24 p. 282–298.

Lackey, J. (2016) "What is Justified Group Belief", *The Philosophical Review* Vol 125 p. 341–396.

Lackey, J. (2016). "What Is Justified Group Belief?", *The Philosophical Review*, 125(3), 341–396.

Lackey, J. (2018), "Group Assertion". *Erkenntnis*, 83, 21–42. https://doi.org/10.1007/s10670-016-9870-2.

Lackey, J. (2021) "Epistemic Duties Regarding Others" in *Epistemic Duties: New Arguments, New Angles*. Kevin McCain and Scott Stapleford ed.

Langton, R. (1993), "Speech Acts and Unspeakable Acts" *Philosophy & Public Affairs*, Vol. 22, No. 4, pp. 293-330.

Lasonen-Aarnio, M. (2013). "Disagreement and evidential attenuation". *Noûs* 47: 767-94.

Ledford, H. (2019), "Millions of black people affected by racial bias in health-care algorithms" *Nature*. Oct 29, 2019. (accessed Sept 29, 2021)

Levin, J. (2018) "Functionalism", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), accessed online at https://plato.stanford.edu/entries/functionalism/

Lewis, D. (1983). "Mad Pain and Martian Pain". *Philosophical Papers*, Volume I: Oxford University Press.

List, C. (2005). "Group Knowledge and Group Rationality: A Judgment Aggregation Perspective." *Episteme* 2: 25–38.

List, C. (2014). "Three kinds of collective attitudes", *Erkenntnis*, 79 (9 Supp). pp. 1601-1622. ISSN 0165-0106 DOI: 10.1007/s10670-014-9631-z.

List, C. and Pettit P. (2004). "Aggregating Sets of Judgments: Two Impossibility Results Compared." *Synthese* 140: 207–35

List, C., & Pettit, P. (2016). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press. Oxford

Loar, B. (1981), *Mind and meaning*, Cambridge: Cambridge University Press.

Longino, H. E. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton University Press.

Longino, H. E. (1991) "Multiplying Subjects and the Diffusion of Power", *The Journal of Philosophy*, Vol. 88, No. 11.

Longino, H. E. (2021) "What's Social About Social Epistemology" *The Journal of Philosophy*, Vol. 119, No. 4.

Lonzi, C. (1970), *Sputiamo Su Hegel*. Editoriale grafica, Italy.

Luzzi, F. (2016), "Testimonial Injustice Without Credibility Deficit (or Excess)", *Thought*, Vol 5, No. 3.

MacFarlane, J. (2003), "Future Contingents and Relative Truth". *The Philosophical Quarterly* 53 (212): 321–336. doi:10.1111/1467-9213.00315.

Machuca, D. E (2011), "Pyrrhonism in Ancient, Modern, and Contemporary Philosophy", in *The New Synthese Historical Library*.

Magnus, P. D., (2007), "Distributed Cognition and the Task of Science" in *Social Studies of Science*, 37 (2):297--310

Major, B., O'Brien, L. T. (2005). "The Social Psychology of Stigma". *Annual Review of Psychology*. 56 (1): 393–421.

Manne, K. (2020), *Entitled: How Male Privilege Hurts Women*, Crown: New York.

Martín, A. (2021) 'What Is White Ignorance' *The Philosophical Quarterly* Vol. 71, No. 4

Matheson, J. (2015). *The epistemology of disagreement*. Palgrave McMillan.

McDowell, J. (1995), "Knowledge and the Internal" in *Philosophy and Phenomenological Research*, Vol. 55, No. 4 (Dec., 1995), pp. 877-893.

McHugh, Connor (2012): "Epistemic Deontology and Voluntariness", *Erkenntnis* 77, 65–94.

McKinnon, Rachel. 2016. "Epistemic Injustice." *Philosophy Compass*, 11 (8): 437–446. .

Medina, J. (2013) *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. New York, NY: Oxford University Press.

Mikkola, M. (2019), *'Does Pornography Silence Women?', Pornography: A Philosophical Introduction* New York, online edn, Oxford Academic,

Miller, J. (2020), "Is An Algorithm Less Racist Than A Loan Officer?" in *The New York Times*. June 5, 2020. Accessed online (Oct 29th, 2021) https://www.nytimes.com/2020/09/18/business/digital-mortgages.html

Mills, C. (2007) 'White Ignorance', *Race and Epistemologies of Ignorance*, 247: 26–31.

Mills, C. W. (1997), *The Racial Contract*. Ithaca, NY: Cornell University Press.

Mills, C. W. (2017), *Black Rights/White Wrongs: The Critique of Racial Liberalism*. New York: Oxford University Press. Oxford Scholarship Online, 2017. doi: 10.1093/acprof:oso/9780190245412.001.0001.

Miragoli, M. (2020). "Group Belief Functionalism", Manuscript.

Moss, S. (2015), "Time-Slice Epistemology and Action under Indeterminacy." *Oxford Studies in Epistemology* 5: pp. 172–194.

Moyal-Sharrock, D. (2004). *Understanding Wittgenstein's On Certainty*. Basingstoke: Palgrave Macmillan.

Moyal-Sharrock, M. and W. H. Brenner (eds.). (2005). *Readings of On Certainty*. Basingstoke: Palgrave Macmillan.

Nelson, M. (2010), 'We have No Positive Epistemic Duties', *Mind* 119: pp. 83–102

Noble, S. (2018), *Algorithms of Oppression*, NYU Press.

Nottleman, N. (2021), "Against Normative Defeat", *Mind* Vol. 130 pp. 1183-1204

Nozick, R. (1981). *Philosophical Explanations*, Oxford: Oxford University Press.

Palermos, O. & Pritchard, D. "The Distribution of Epistemic Agency", in *Social Epistemology and Epistemic Agency*: *De-Centralizing Epistemic Agency*, (ed.) P. Reider, (Rowman & Littlefield, 2016).

Peels, R. (2023), *Ignorance: a philosophical study*. Oxford University Press: New York, US.

Pettigrew, R., "Believing is said of groups in many ways (and so it should be said of them in none)". Manuscript

Pettit, P. (2003) "Groups with Minds of Their Own," in Frederick Schmitt (ed.), *Socializing Metaphysics*. New York: Rowman & Littlefield, 167–193.

Pohlhaus, G. Jr (2020) "Epistemic Agency Under Oppression", *Philosophical Papers*, 49:2, 233-251, DOI: 10.1080/05568641.2020.1780149

Pohlhaus, G., Jr (2012) 'Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance', *Hypatia*, 27: 715–35.

Pohlhaus, G., Jr (2021) 'Epistemic Oppression, Ignorance, and Resistance', in Kim Q. Hall, and Ásta (eds), *The Oxford Handbook of Feminist Philosophy*.

Prichard, H.A., (1950). *Knowledge and Perception*, Oxford: Clarendon Press.

Pritchard, D. (2011). "Wittgensteinian Pyrrhonism," in *Pyrrhonism in Ancient, Modern, and Contemporary Philosophy*, (ed.) D. Machuca, 193–202, Dordrecht: Springer.

Pritchard, D. (2012) *Epistemological Disjunctivism*, OUP, Oxford.

Pritchard, D. (2015) *Epistemic Angst*, Princeton University Press.

Pritchard, D. (2021), 'Ignorance and Inquiry', *American Philosophical Quarterly* 58: 111-23.

Quijano, A. (1999), "Coloniality and Modernity/Rationality". In Goran Therborn, ed. *Globalizations and Modernities*. Stockholm: FRN

Quinton, A. (1975) "Social-Objects", *Proceedings of the Aristotelian Society* 75, 1-27.

Rachels, J. (1999) "The Challenge of Cultural Relativism" in *The Elements of Moral Philosophy* (ch.2), eds. McGraw-Hill.

Rafanelli, Lucia M. (2022), "Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy" in *Big Data & Society* January–June: 1–5

Ranalli, C. (2014). "Luck, Propositional Perception, and the Entailment Thesis." *Synthese*, 191(6), 1223–47.

Rawls, J. (1971), *A theory of justice*. Cambridge, MA: Harvard University Press.

Rawls, J. (1993), *Political Liberalism*. New York: Columbia University

Reicher, S. (1982). "The Determination of Collective Behaviour." Pp. 41–83 in H. Tajfel (ed.), *Social identity and intergroup relations*. Cambridge: Cambridge University Press.

Reicher, S. (1987) "The Psychology of Crowd Dynamics", *Blackwell Handbook of Social Psychology: Group Processes*. ed. Michael A. Hogg & R. Scott Tindale. Blackwell Publishers Inc. Malden, Mass.

Rorty, R. (1979), *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.

Rosen, G. (2004), 'Skepticism About Moral Responsibility', *Philosophical Perspectives* 18 Ethics: 295–313.

Roser, M., Ritchie, H., Ospina-Ortiz, E. (2015) "Internet" *Our World in Data*, (accessed Sept 29, 2021). https://ourworldindata.org/internet

Sankey, H. (2010). "Witchcraft, Relativism and the Problem of the Criterion". *Erkenntnis* (1975-), 72(1), 1–16.

Santos, B. de Sousa (2013). *Epistemologies of the South: Justice against Epistemicide*, Routledge, London & New York.

Santos, B. de Sousa (2019). *The End of the Cognitive Empire: La afirmación de las epistemologías del Sur*. Madrid: Trotta. ISBN 978-84-9879-780-0

Schmitt, F. F. (1994), "The Justification of Group Beliefs", in *Socializing Epistemology: The Social Dimensions of Knowledge*, ed. Schmitt Frederick F., 257–87. Lanham, MD: Rowman and Littlefield.

Schwitzgebel, E. (2023) "Belief", in *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2023/entries/belief/>.

Sellars, W (1956), "Empiricism and the Philosophy of Mind." In *Minnesota Studies in the Philosophy of Science*, vol. 1, edited by Herbert Feigl and Michael Scriven, pp. 253–329. Minneapolis, MN: University of Minnesota Press.

Sellars, W (1969), "Language as Thought and as Communication." *Philosophy and Phenomenological Research* 24 (4): pp. 506–527.

Siegel, S. (2012). "Cognitive penetrability and perceptual justification". *Noûs*, 46(2), 201–222.

Simion, M (forthcoming), "Defeat", in *The Blackwell Companion to Epistemology*, Third Edition: Kurt Sylvan, Matthias Steup, Ernest Sosa, and Jonathan Dancy (eds.).

Simion, M. & Kelp, C., 'What Is Normative Defeat'. Manuscript

Simion, M. (2019a). "Epistemic Norms and Epistemic Functions". Manuscript.

Simion, M. (2019b). "Disagreement, Knowledge First". Manuscript.

Simion, M. (2019c). "Knowledge-First Functionalism", *Philosophical Issues*, 29 (1):254-267.

Simion, M. (2021), "Resistance to Evidence and the Duty to Believe." *Philosophy and Phenomenological Research*. Vol. 1, 14.

Simion, M., Carter, A., and Kelp, C. (2022), "On behalf of knowledge first collective epistemology". In Silva, P. And Oliveira, L. Eds. *Propositional and Doxastic Justification*. London: Routledge.

Simion, M., Kelp, C. (2023), "Trustworthy artificial intelligence" In *Asian Journal of Philosophy* 2, 8 (2023).

Simpson, R. M. (2013). "Epistemic peerhood and the epistemology of disagreement". *Philosophical Studies* 164: 561-77.

Smith, R. (2012). "Segmenting an Audience into the Own, the Wise, and Normals: A Latent Class Analysis of Stigma-Related Categories". Communication Research Reports (29 ed.). 29 (4): 257–65. doi:10.1080/08824096.2012.704599.

Snow, J. (2018, February 14). "'We're in a diversity crisis': Cofounder of Black in AI on what's poisoning algorithms in our lives". *MIT Technology Review*. Sourced online (Sept 29, 2021) at https://www.technologyreview.com/2018/02/14/145462/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/#:~:text=There%20is%20a%20bias%20to,of%20people%20in%20the%20world.

Solnit, R. (2014). *Men Explain Things to Me*. Moe's Books, Berkeley, California.

Sosa, E. (2000). "Skepticism and Contextualism," *Philosophical Issues* 10, 1–18.

Spelman, E. (2007) "Managing Ignorance", in S. Sullivan and N. Tuana (eds.), *Race and Epistemologies of Ignorance*, 119–31. Albany, NY: State University of New York Press.

Spewak, D. (2023). "Perlocutionary Silencing: A Linguistic Harm That Prevents Discursive Influence" *Hypatia*, 1-19. doi:10.1017/hyp.2023.2

Spivak, G. C. (1999) *A critique of postcolonial reason: Toward a history of the vanishing present*. Cambridge, Mass.: Harvard University Press.

Srinivasan, A. (2015), "Normativity Without Cartesian Privilege" in *Philosophical Issues*, 25, doi: 10.1111/phis.12059.

Steup, M. and Neta, R. (2020). "Epistemology", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2020/entries/epistemology/>.

Strawson, P. (1985). *Skepticism and Naturalism. Some Varieties*. London: Methuen.

Strohmaier, D. (2019). Two theories of group agency. *Philosophical Studies*. 177 (7):1901-1918

Sullivan, S. & Tuana, N. (2007) "Race and Epistemologies of Ignorance" in SUNY series, *Philosophy and Race*: Ronald Bernasconi and T. Denean Sharpley-Whiting editors. State University of New York Press, Albany.

Tajfel, H., and Turner, J. C., (1979). "An integrative theory of intergroup conflict". In W.G. Austin & S. Worchel (eds.), *The social psychology of intergroup relations*. pp. 33–47. Monterey, CA: Brooks/Cole

Tanesini, A. (2016), "'Calm Down, Dear': Intellectual Arrogance, Silencing and Ignorance", *Aristotelian Society Supplementary Volume*, Volume 90, Issue 1, Pages 71–92.

Thorstad, D. (2021) "Inquiry and The Epistemic", *Philosophical Studies* 178 (9):2913-2928.

Tollefsen, D (2015). *Groups as Agents*, Polity Press.

Townley, C. (2003), "Trust and the Curse of Cassandra (An Exploration of the Value of Trust)" *Philosophy in the Contemporary World*, Volume 10 Number 2: 105-111.

Townley, C. (2006), "Toward A Revaluation Of Ignorance", *Hypatia*, vol. 21, no. 3

Tuomela, R. (1992). "Group Beliefs." *Synthese*, 91: 285–318

Tuomela, R. (1993). "Corporate Intention and Corporate Action." *Analyse und Kritik*, 15: 11–21

Tuomela, R. (1995). *The Importance of Us*. Stanford, CA: Stanford University Press.

Tuomela, R. (2007). *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press. Oxford

Ullmann, S. (2021), "Google Translate is sexist. What it needs is a little gender-sensitivity training" *The Conversation*. Apr 05, 2021. Accessed (Sept 29, 2021) at https://scroll.in/article/991275/google-translate-is-sexist-and-it-needs-a-little-gender-sensitivity-training#:~:text=Instead%20of%20this%20laborious%20gender,gender%2Dsensitivity%20training%20at%20work.

Weatherson, B. (2013). "Disagreements, philosophical and otherwise". In D. Christensen & J. Lackey (eds)., *The epistemology of disagreement*: New essays. Oxford University Press.

Weisberg, M. (2009). "Three kinds of idealization". *Journal of Philosophy* 104: 639-59.

Willard-Kyle, C. (2023) "The Knowledge Norm of Inquiry" *The Journal of Philosophy*. 120 (11):615-640

Williams, B. (1970): "Deciding to Believe", in his *Problems of the Self*, Cambridge: Cambridge University Press.

Williams, M. (1991). *Unnatural Doubts*. Oxford: Blackwell. Oxford.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press. Oxford

Wilson, R.A. (2005) "Collective memory, group minds, and the extended mind thesis. Cognitive Processing" 6, 227–236 (2005). *Cogn Process*, DOI 10.1007/s10339-005-0012-z.

Woomer, L. (2019) "Agential Insensitivity and Socially Supported Ignorance", *Episteme*, 16: 73–91.

Wray, K. B. (2001). "Collective belief and acceptance". *Synthese*, 129, 319–333. .

Wray, K. B. (2007). "Who has Scientific Knowledge?". *Social Epistemology*, Vol 21, pgg. 337-347.

Wrenn, C (2007) "Why There Are No Epistemic Duties." *Dialogue*, 46: pp. 115–36.

Wright, C. (1985). "Facts and Certainty," *Proceedings of the British Academy* 71: 429–472.

Wright, C. (2004). "Warrant for Nothing (and Foundations for Free)?" *Aristotelian Society Supplementary Volume*, 78: 167–212.

Young, I.M. (1990), *Justice and the Politics of Difference*, Princeton, N.J.: Princeton University Press.

Young, I.M. (1995), "Mothers, Citizenship, and Independence: A Critique of Pure Family Values", *Ethics* 105: 535–56.

Young, I.M. (2011), *Responsibility for Justice*. Oxford University Press, USA.

Zalabardo, J. (2015). "Epistemic Disjunctivism and the Evidential Problem." *Analysis* 75(4), 615–27.

van Inwagen, P. (1996). "It is wrong, everywhere, always, for anyone, to believe anything upon insufficient evidence". In J. Jordan & D. Howard-Snyder (eds.). *Faith, freedom and rationality*. Rowman and Littlefield: 137-54.

van Inwagen, P. (2010). "We're right. they're wrong". In T. Warfield & R. Feldman (eds.), *Disagreement*. Oxford University Press.

van den Hoven, E. (2019). "Automated hermeneutical injustice", sourced online at: https://www.cohubicol.com/blog/automated-hermeneutical-injustice/