



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

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

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

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

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

TRUTH RELATIVISM IN METAETHICS

Patrick Denning

PhD Philosophy

The University of Edinburgh

2021

ABSTRACT

Metaethical relativism is the view that whether a moral claim is true depends on the standards endorsed by an individual or society. This view is attractive because it allows one to hold that moral claims can be true or false in an ordinary correspondence sense, without being committed to the view that moral claims state objective facts. But what could it mean to say that whether a moral claim is true depends on an individual or society's standards? How could a single claim be true or false in an ordinary sense, and yet be true for some people while false for others?

This thesis defends a certain answer to this question: metaethical truth relativism. Metaethical relativists have historically endorsed a different view: content relativism, which is the view that the semantic contents of moral claims vary across contexts of use. However, semanticists working outside metaethics have recently shown that an alternative exists, namely truth relativism: the view that the truth values of moral contents vary across contexts. Metaethicists have been slow to investigate the significance of this discovery. This is surprising, because truth relativism appears to avoid some of the most influential objections to content relativist views. Thus, this development in semantics appears to offer an exciting new direction for research on relativism in metaethics.

The scope of this thesis is not narrowly semantic, however. Questions in moral semantics often interconnect with questions in substantive metaethics, and the question of whether ethical thought and talk is semantically relativistic is no exception. Therefore, while the core of the thesis focuses on moral semantics, it also contains a fair amount of substantive metaethical material, located primarily in the first and final chapters.

The plan of the thesis is as follows. Chapters 1 & 2 begin with a discussion of the most historically influential argument for metaethical relativism: the argument from disagreement. Chapter 1 discusses an epistemic version of this argument, but concludes that this version provides only limited evidence for relativism. However, chapter 2 introduces a different version of the argument, with a more semantic focus, (which appeals to the notion of faultless

disagreement) and begins to make the case that this argument provides evidence not only for metaethical relativism broadly construed, but for truth relativism in particular.

The broad strokes of this argument are that there can be moral disagreements which are faultless (in a sense which rules out objectivist views) and furthermore that only truth relativism (as opposed to content relativism) can vindicate the sense that these are genuine disagreements, rather than mere differences of opinion. Chapter 2 presents empirical evidence for the claim that putative moral disagreements can be *faultless*. Chapters 3 and 4 then take a deep dive into the topic of *disagreement*. Chapter 3 develops a positive account of disagreement which is available only to truth relativists. And chapter 4 advances detailed arguments against contextualist accounts of disagreement. Chapters 3 and 4 together thus defend in detail the claim that truth relativism can, while content relativism cannot, account for the sense that there is genuine disagreement in the relevant cases. Chapters 2 through 4, therefore, form a single long argument for truth relativism. They show that there is a compelling argument for relativism broadly construed, and that deep analysis of the issues surrounding this argument favour truth relativism over content relativism.

Chapter 5 zooms back out to the broader metaethical landscape, discussing two remaining objections to relativism. The first objection is that relativism has objectionably subjectivistic implications, and the second is that moral standards cannot helpfully be understood as determined by societies (rather than individuals). In the course of responding to these objections, natural opportunities to flesh out the substantive metaethical commitments of the theory will arise. And this will set us up to conclude with a brief comparison of truth relativism with some of the other main options in the space of substantive metaethical theory.

LAY ABSTRACT

Moral relativism is the view that whether a moral claim is true is not an objective matter (like claims in mathematics or the natural sciences) but rather is relative to the standards endorsed by a society (in something like the way claims about what is legal or rude are). Although moral relativism is popular among non-philosophers, it has historically been unpopular among philosophers, since it faces objections which many philosophers have regarded as decisive. This thesis develops a new version of moral relativism, based on recent advances in semantic theory. The thesis argues that this version of relativism is not vulnerable to the objections in question, and should be taken more seriously by philosophers.

CONTENTS

<u>Introduction</u>	(6)
<u>Chapter 1</u>	(18)
<u>Chapter 2</u>	(51)
<u>Chapter 3</u>	(76)
<u>Chapter 4</u>	(107)
<u>Chapter 5</u>	(133)
<u>Conclusion</u>	(169)
<u>References</u>	(174)

INTRODUCTION

ABSTRACT: This introduction presents the background theoretical space within which the thesis is located. It explains why there is a focus on semantics in the thesis, and presents the relevant semantic, and more broadly metaethical notions. The introduction concludes with a brief plan of the rest of the thesis.

The discussion of this thesis takes place against the background of an ongoing debate, with an associated literature. As with any such debate and literature, a distinctive vocabulary and set of concepts are used. Since this thesis is framed in terms of this vocabulary and this set of concepts, it is necessary at the beginning to lay out how I understand these concepts and this vocabulary.

The central focus of this thesis, furthermore, is a certain metaethical view, known as moral relativism. Moral relativists hold that moral claims are not true or false objectively, but are rather made true or false by the moral standards of those who make such claims. This looks like a wide-ranging claim about the fundamental nature of moral truth. But debates over moral relativism typically focus on issues in natural language *semantics*, a quite technical field straddling philosophy of language and empirical linguistics. When one first encounters these debates, this semantical focus may seem puzzling. So I want to begin this thesis by not only laying out my understanding of the various notions involved in these debates, but also by explaining why the latter so often focus on the semantics of moral language.

1. Semantics for moral relativists

The core thesis of moral relativism is that moral claims are made true by moral standards, and therefore that a particular moral claim can be true relative to one set of standards, but false relative to another.

Core Claim: a particular moral claim can be true relative to one set of standards and false relative to another.

But this commitment comes with a specifically semantic challenge. For from the perspective of semantics, it is puzzling how one and the same moral claim be true when one person makes it, and false when another person does.

To see why this is puzzling, first consider that we individuate claims by their meanings. Thus, when we talk about a given moral claim being true for me and false for you, we talk about something with a unique meaning being true for me and false for you. For example, consider the following anecdote recounted by the ancient historian Herodotus:

Darius, after he had got the kingdom, called into his presence certain Greeks who were at hand, and asked – “What he should pay them to eat the bodies of their fathers when they died?” To which they answered, that there was no sum that would tempt them to do such a thing. He then sent for certain Indians, of the race called Callatians, men who eat their fathers, and asked them, while the Greeks stood by, and knew by the help of an interpreter all that was said – “What he should give them to burn the bodies of their fathers at their decease?” The Indians exclaimed aloud, and bade him forbear such language.¹

Intuitively, we want to say that there is a unique claim at issue between the Greeks and Callatians. If, then, we individuate claims by their meanings, it follows that the Greeks’ and Callatians’ words *mean the same things*, even if (supposing we are relativists) we also want to say that both parties are correct, given their different moral standards.

But now, here is why this is hard to understand, from the point of view of semantics. The dominant tradition in natural language semantics construes the meaning of a sentence as its *truth conditions*.² But if we accept this, then we must accept that if two claims have the same meaning, it follows that they have the same truth conditions. But if we accept *that*, then it

¹ (Hdt.)

² (Segal & Larson, 1995)

seems that we must accept that the Greeks' claims about morality have the same truth conditions as the Indians' claims. Consequently, if we understand truth conditions straightforwardly as states the world could be in, then we will be led inexorably to the conclusion that the Greeks' and Indians' claims cannot both be correct. After all, the world surely is in just one determinate state when they make their respective utterances. Therefore, assuming there is a unique claim at stake (with a unique meaning), the truth conditions of this claim are either determinately satisfied or determinately not satisfied. And therefore, we seem compelled to conclude that what we wanted to say in the beginning, that both parties' claims are correct, cannot be true.

Contrast this semantic challenge with relativism's appeal in other areas of meta-ethical interest, namely metaphysics, psychology, all-in practical normativity, and epistemology. Moral relativism seems to ask no more of metaphysics than that there are facts about our standards, to make our moral claims true. It seems to ask no more of our psychology than that we can be motivated by our own standards. It seems to ask no more of practical normativity than that we have reasons to conform to our own standards. And it seems to ask no more of epistemology than that we can know what our own standards entail. In all these respects, moral relativism seems to offer solutions to many of the central problems of metaethics. But before we can hope to secure these metaphysical, psychological, all-in normative, and epistemological payouts, we need first to solve the problem just mentioned. This is why semantics so often takes centre stage in discussions of moral relativism.³

2. Content relativism

Now a standard kind of answer to this semantic problem has been to point out that the argument which we used to generate the problem must prove too much. After all, it is utterly uncontroversial that there are some sentences which can be correctly asserted and denied by different people. For example, I can correctly say that I am PD and you can equally correctly say that you are not PD, and accepting this obviously does not require that we reject truth

³ This is not to glibly assume that moral relativism does solve all of these problems in other areas, only that it looks like a promising way to go.

conditional semantics, or that we insist that the state of the world cannot be constant across our utterances, or that we accept any kind of logical contradiction. Nevertheless, there is a clear sense in which the sentence “I am PD” has the same meaning when I say it as when you do. The argument given above, therefore, must have gone wrong somewhere.

What we can say is that the argument equivocates by sliding between two kinds of meaning. As we learn from Kaplan, context sensitive expressions have meanings both in the form of *character* and in the form of *content*. The sentence “I am PD” has a unique, context insensitive character. It is in this sense that the sentence has the same meaning, regardless of who utters it. However, the sentence’s content is context-sensitive. That is, the sentence has different contents when used by different people. Thus, the content of “I am PD” when I say it is roughly equivalent with the content of the sentence that PD is PD. In contrast, the content of the same sentence when Aristotle says it is roughly equivalent with the content of the sentence that Aristotle is PD. All of this is compatible with truth conditional semantics because it is the content, not the character, that is understood in terms of truth conditions.

This has seemed to many a promising solution to the semantic problem for moral relativism.⁴ After all, it explains how a particular claim can have the same meaning, in one sense, when different people use it, while still being true in the mouths of some, and false in the mouths of others, all consistent with truth conditional semantics. And since it is quite uncontroversial that *some* claims are like this, there seems no principled objection to arguing that moral claims are like this. Call the view that this is the correct semantic framework for moral language: content relativism.

3. Truth relativism

This view faces its own challenges, however. Although we will save a discussion of the details of these challenges for later, the fundamental problem is that content relativism implies that moral thought and talk is fairly transparently about something much more mundane, namely our own standards. On one prominent content relativist view, for example, *all we are saying*

⁴ (Wong, 1984) (Harman, 1975).

(at least at the level of semantic content) when we say that something is good is that it is conducive to our ends.⁵ And for various reasons (including, prominently, concerns around disagreement) philosophers have found this view unsatisfactory.

In response to this dissatisfaction, recent years have seen the development of an alternative, form of relativism. According to these new-style *truth* relativists, moral claims are made true by our standards, but this is not because they are *about* our standards. Rather, on this view, moral claims are *about* morality, they have contents (propositions) which are distinctly moral, it is just that these contents are made true by standards (just as, they might remind us, ordinary propositions are made true by states of the world).

There are two broad models for explaining how this can work in semantic terms. One which is relatively more conservative, in the sense that it departs relatively less from the Kaplanian picture, and one which is relatively more ambitious.

4. Conservative truth relativism

The more conservative approach⁶ starts with Kaplan's⁷ framework for context-sensitivity. As we've seen, the basic elements of this framework are the notions of character and content. The character of a sentence determines the sentence's content, relative to a context of use, and the content determines the sentence's truth-value, relative to a circumstance of assessment.

Content relativism locates the relativity of truth to standards that we are looking for at the level of character, by saying that the characters of moral sentences are such as to determine different contents at different contexts of use. In this respect, it is an orthodox approach, because it hangs on to the Fregean idea that semantic content determines truth value⁸ – so that if two people believe the same thing (at the same world) what they believe must have the same truth value. Content relativism allows for relativity without violating this principle,

⁵ (Finlay, 2014).

⁶ Kolbel (2002) (2004a) (2015a) (2015b) is a pioneer of this approach.

⁷ (Kaplan, 1989).

⁸ (Speaks, 2019).

because it says that people who have superficially similar beliefs may not really believe the same thing. Instead, according to content relativism, one person may believe one thing, and another person believe something else, and this is what allows one person's belief to be true and another's to be false.

What the conservative new-style relativist starts off by pointing out, though, is that there is already a kind of relativity at the level of content, even in the standard Kaplanian picture, for Kaplan holds that a given content can be evaluated as true or false only relative to a possible world.

To illustrate, recall that in Kaplan's framework the character of a sentence determines its content relative to a context of use, which can be modelled as a collection of parameters, such as a time, place, world, and speaker. The context of use is one kind of *index*, a set of parameters to which a semantic function can be relativised. In this case, the character – the function that takes sentences and spits out propositions – is relativised to the context of use, so that a given sentence is paired with a given content only relative to a context of use. But Kaplan's theory also involves a second index, which Kaplan calls a circumstance of evaluation. This index takes contents and spits out truth values, relative to a possible world. So on Kaplan's view, a proposition's truth value is relative to a possible world, in the sense that a possible world is a parameter of the index which determines the truth value that gets assigned to the proposition by the semantic theory.

What the conservative new-style relativist wants to do, then, is keep Kaplan's character/content framework, and simply add a new parameter to the index which previously determined a possible world only. Instead of being true relative to a possible world, on this view, a proposition is true relative to a (possible world, moral standard) pair. Brogaard (2008) (2012) and Kolbel (2004b) have defended versions of this view for moral discourse.

5. Ambitious truth relativism

Some truth relativists however (notably MacFarlane (2014)) think that this form of relativism does not go far enough. To understand why this is so, we need to bring in some considerations in speech act theory, as follows. Conservative truth relativism entails that two contradictory moral *assertions* can both be correct. This is one sense in which it vindicates the Core Claim. However, difficulties arise for conservative truth relativism when we consider, in addition to assertion, the speech act of retraction. Suppose, for example, a reformed meat eater, now vegetarian, recalls some pro-meat-eating statement they made many years ago, say, “killing cows for meat is permissible”. It seems that they could correctly say “the assertion I made all those years ago was incorrect, *I take it back/retract it*”. Conservative truth relativism has difficulty accounting for this.

This is because conservative truth relativism, by importing the Kaplanian framework wholesale, countenances proposition truth only relative to a single context, the context of use. It is thus difficult to see how it can account for retraction, because retraction is made correct by considerations involving *two* contexts, the context of use and the context of assessment. That is, a retraction is correct, plausibly, if the content of the targeted assertion is false as used at the earlier context and assessed at the later context.

This may be non-obvious. It may seem that retraction is required just in case the content asserted is false at the context of the retraction. But this misses the cross-contextual role of world-states and moral standards in retraction. A retraction is, plausibly, justified if the content is false relative to the *world-state* of the *context of assertion* and the *standards* of the *context of retraction*. To see this, imagine that, in the future, plant-diseases wipe out all strains of edible crops. In that case, vegetarian standards might permit meat-eating. But even in that scenario, it will seem that our reformed vegetarian would still want to say “the assertion I made all those years ago was incorrect, I take it back”. After all, their earlier statement was made on the grounds of values they now reject. However, the content [killing cows for meat is permissible] is true relative to their current context. Thus, to get the result that retraction would still be correct (in the plant-disease scenario), we need a view which is able to assess a content for truth as used at a world-state and assessed at a standard. Similar considerations, as we will see in chapter 3, apply to the speech act of *rejection* (which also requires this kind

of 2-D approach to content (more than merely sentence) truth). As we will see, appeal to the notion of rejection plays an important role in the argument of that chapter.

This argument is one motivation for MacFarlane's ambitious relativism. This view differs from the Kaplanian framework inasmuch as while the former construes content in terms of truth at a world and standard, both of which are determined by the context of use, the latter construes content in terms of truth at a world determined by a context of use and a standard determined by a context of assessment (where the key point is that the context of assessment can float free, within the semantic machinery, from the context of use).

Both notions of truth, crucially, are technical ones, designed to model, in abstract terms, that which a speaker aims to communicate through typical uses of declarative sentences (namely, a truth condition). Thus, neither view is required to deny the coherence of ordinary (non-relativised) use of 'true' simpliciter. Rather, both views can offer relativistic characterisations of the semantics of the ordinary truth predicate (and indeed, this move is forced, given that the ordinary truth predicate can be applied to contents which, on this view, have their extensions only relatively). Thus, both views can hold that assertions of the form 'p is true' are correct only if the relevant content is true at the asserter's (i.e. the alethic evaluator's) context. Thus, according to these views, it is possible for one person, A, to make a moral claim, and for two other people, B and C, to say "that is true" and "that is false" respectively, and for both to be correct. This is a second sense in which both views vindicate the Core Claim.

6. Metaethical theories

These considerations are not intended as an argument for either truth relativism in general, or its ambitious form. They are intended merely to explain why this thesis will involve doing some semantics, and to sketch the basic semantic frameworks which will be discussed. The aim of the thesis is, ultimately, to defend ambitious truth relativism over contextualism. But the arguments are yet to come.

Although we will end up doing a fair bit of philosophical semantics, the scope of the thesis is intended to be broadly metaethical. To evaluate truth relativism in metaethics, we need to investigate not only whether a truth-relativist semantics is superior to a content-relativist semantics, but also whether metaethical relativism, broadly construed, is attractive. Both questions are discussed in the thesis.

Before we get started, we therefore also need to locate relativism in the broader space of metaethical theory. Start with moral *realism*. Moral realists hold, first, that moral judgements are beliefs which represent the world (in a non-minimal sense of ‘belief’ and ‘represent’). Second, realists hold that some substantive moral beliefs are true (in a non-minimal sense of ‘true’). And third, realists hold that at least fundamental moral beliefs are made true by (non-minimally) “mind-independent” facts (i.e. facts that do not involve mental states). Intuitively, realism can be thought of as the view that whether something is wrong is a fact that is “out there” in the world, waiting to be discovered, in much the way that facts about rocks, numbers and societies are. Also, intuitively, realism can be thought of as *opposed* to the view that whether something is wrong “depends on us”, in the way that whether something is rude or tasty or funny, plausibly, does.⁹

Different views can be approached by thinking about which components of the realist picture they reject. Expressivist opponents (who have the least in common with realists) get off at the first juncture, denying that moral judgements are beliefs which represent the world (and therefore, a fortiori, denying all the other claims as well).¹⁰ Error theorists get off at the second juncture, accepting that moral judgements are (in a robust sense) beliefs, but denying that any non-trivial moral beliefs are true (and therefore, a fortiori, denying that these beliefs are made true by mind-independent facts).¹¹ Finally non-realist cognitivists (who have the most in common with realists) get off at the third juncture, accepting that moral judgements are beliefs and that some non-trivial moral beliefs are true, but denying that the facts that make these beliefs true are mind-independent.¹²

⁹ (Enoch, 2011) (Parfit, 2011) (Brink, 1988).

¹⁰ (Blackburn, 1984) (Gibbard, 2003) (Schroeder, 2008).

¹¹ (Mackie, 1977) (Joyce, 2001) (Olson, 2014).

¹² (Street, 2016) (Korsgaard, 1996).

We have said that our focus in this thesis is moral *relativism*. Moral relativists reject moral realism, as thus defined. However, the reason why they do so is somewhat obscured by this, standard, definition of realism. The obscurity arises because there is a division between forms of non-realist cognitivism. Some versions of this view hold that there is, as it were, a “single true morality” while others hold that there are “many true moralities”. Both of these views, notice, are consistent with the distinguishing thesis of non-realist cognitivism, which is the combination of views that moral judgements are beliefs (in a robust sense), that some moral beliefs are non-trivially true, and that the facts that make these beliefs true, in general, are *mind-dependent*. For example, Kantian constructivists¹³ hold that rational agents *as such* have reasons to act morally (*de re*), so that moral beliefs are made true by facts about what rational agents *qua* agents have reason to do. On this view, there is a perfectly respectable sense in which morality is *objective* – any action with a determinate moral status will have this status for all agents and from all perspectives (so if ϕ -ing is wrong, it is wrong for any and all agents, regardless of the agent’s beliefs, desires, cultural background, etc.). This view is not a form of *realism*, as standardly construed, because it makes moral facts constitutively *mind-dependent*. But since this view is one of the views that relativists want to deny, on the grounds that (by their lights) it attributes an implausible degree of objectivity to morality, we are going to want a notion of objectivity which includes not only realism, but also views of this type. I will stipulatively use the terms ‘objectivity’ and ‘objective’ for this purpose: morality is objective, in this sense, if either realism or a non-realist cognitivist view of this type is true. Otherwise, morality is not objective. Crucially, morality is not objective if a form of non-realist cognitivism which holds that there are *multiple* “true moralities” is true. I will use the term ‘relativism’ *simpliciter* to pick out this class of view (non-realist cognitivism, which denies error theory), for this characterisation does not distinguish between truth and content relativism at the level of semantics. Furthermore, I will use ‘realism’ to refer to the substantive metaethical view of that name, and ‘invariantism’ to refer to the truth-invariant semantics with which it is associated.

¹³ (Korsgaard, *ibid.*)

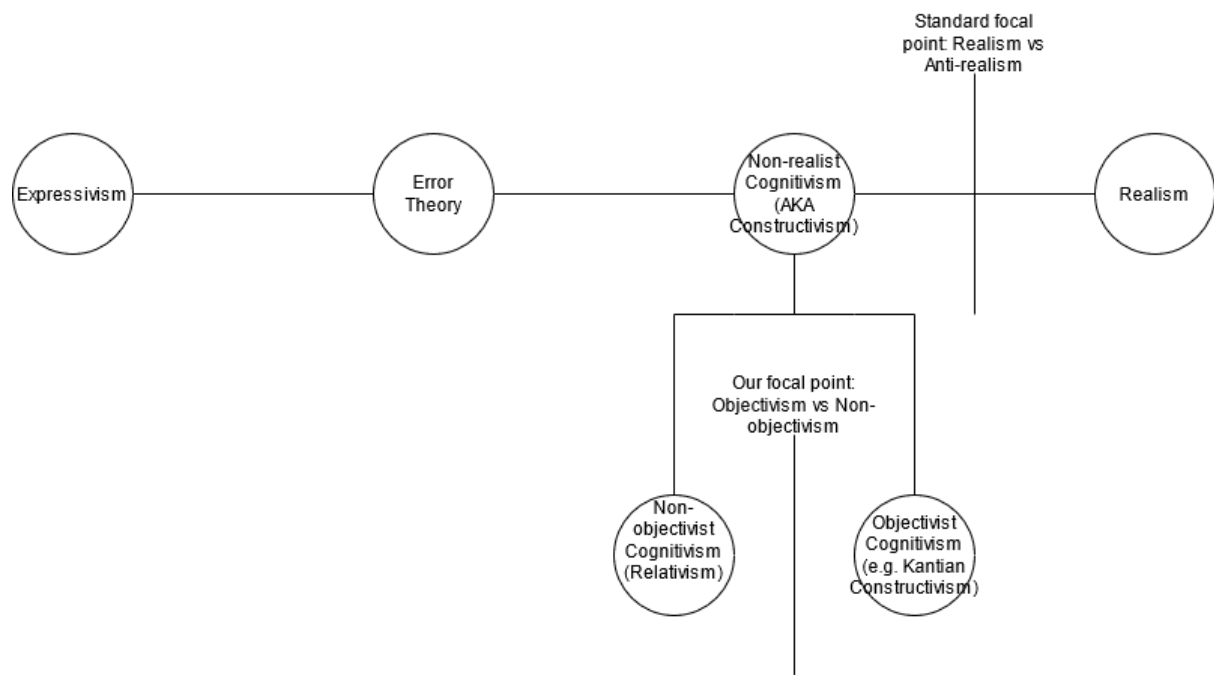


Fig 1. The logical space.

5. Plan of the thesis

Having defined terms, let me offer a brief plan of the rest of the thesis.

Chapters 1 & 2 begin with a discussion of the most historically influential argument for metaethical relativism: the argument from disagreement. Chapter 1 discusses an epistemic version of this argument, but concludes that this version provides only limited evidence for relativism. However, chapter 2 introduces a different version of the argument, with a more semantic focus, (which appeals to the notion of faultless disagreement) and begins to make the case that this argument provides evidence not only for metaethical relativism broadly construed, but for truth relativism in particular.

The broad strokes of this argument are that there can be moral disagreements which are faultless (in a sense which rules out objectivist views) and furthermore that only truth relativism (as opposed to content relativism) can vindicate the sense that these are genuine disagreements, rather than mere differences of opinion. Chapter 2 presents empirical evidence for the claim that putative moral disagreements can be *faultless*. Chapters 3 and 4 then take a deep dive into the topic of *disagreement*. Chapter 3 develops a positive account of disagreement

which is available only to truth relativists. And chapter 4 advances detailed arguments against contextualist accounts of disagreement. Chapters 3 and 4 together thus defend in detail the claim that truth relativism can, while content relativism cannot, account for the sense that there is genuine disagreement in the relevant cases. Chapters 2 through 4, therefore, form a single long argument for truth relativism. They show that there is a compelling argument for relativism broadly construed, and that deep analysis of the issues surrounding this argument favour truth relativism over content relativism.

Chapter 5 zooms back out to the broader metaethical landscape, discussing two remaining objections to relativism. The first objection is that relativism has objectionably subjectivistic implications, and the second is that moral standards cannot helpfully be understood as determined by societies (rather than individuals). In the course of responding to these objections, natural opportunities to flesh out the substantive metaethical commitments of the theory will arise. And this will set us up to conclude by briefly comparing truth relativism with some of the other main options in the space of substantive metaethical theory.

CHAPTER 1

ABSTRACT: Moral relativism is often motivated by arguments from disagreement. In this thesis we will discuss two such arguments. First, in this chapter, we will examine epistemic versions of the argument. After a detailed review, we will conclude that this version of the argument provides at best limited evidence for moral relativism. In the next chapter we will turn to the second version of the argument, namely the argument from so-called faultless disagreement.

Historically, the most influential argument for moral relativism has been one which focuses on moral disagreement.¹⁴ Although there is an inchoate sense, in the minds of many, that moral disagreement somehow provides evidence for relativism (or at least, that it undermines objectivism) this sense is not always made precise. It is a familiar fact¹⁵ that disagreement *per se* does nothing to support relativism. And it is often thus assumed (on both sides of the debate, and not without reason) that the onus is on the proponent of the argument to provide a precise specification of what it is exactly about moral disagreement, in particular, that allegedly supports moral relativism. This has, perhaps unsurprisingly, led to a proliferation of views about how precisely the argument should be formulated.

¹⁴ (Mackie, 1977: 36-38) is an influential statement of the argument, Westermarck (1906-08) (1932) is an important early advocate of the view that the argument supports relativism in particular, and see also Sextus Empiricus (reprinted in (Gowans, 2000)) for an ancient version. A comprehensive recent study is (Tersman, 2006). Empirically informed versions of the argument include: (Brandt, 1954) (Machery et al., 2005) (Doris & Plakias, 2008) (Abarbanell & Hauser, 2010) (Fraser & Hauser, 2010). Armchair versions include: (Bennigson, 1996) (Tolhurst, 1987) (Rowland, 2017) (Risberg & Tersman, 2019) (Wright, 1995) (Rawls, 1974) (Loeb, 1998) (Seipel, 2020). Responses to the argument include (Brink, 1984) (MacGrath, 2010) (Huemer, 2005) (Shafer-Landau, 2003) (Enoch, 2009) (Sturgeon, 1994) (Moody-Adams, 1997) (Clarke-Doane, 2014).

¹⁵ For a helpful explanation of this point see (McGrath, 2007).

Nevertheless, there is a broad line of thinking which has a claim to be the most prominent precisification of the worry, which we might call the *epistemic* argument from disagreement (or, as it is sometimes known, the argument from “rationally irresolvable” disagreement). It is this line of argument which we will explore in this chapter. But first, it may help to say which versions of the argument we will *not* be exploring.

First, some versions of the argument appeal not only to claims about moral disagreement, but also to substantive claims about the nature of moral *judgement* (such as that moral judgements necessarily motivate, or that they essentially involve emotion). I will not discuss arguments of this kind, here, because they shift the focus of the discussion to debates in moral psychology, which fall outside the scope of the thesis.¹⁶

Second, recent years have seen the development of a new form of argument, focusing on so-called “faultless disagreement”. These arguments, which are more semantically minded, are normally discussed in the context of aesthetics and matters of taste (Kolbel, 2004a) (Lasersohn, 2005) (Egan, 2014). There is obvious scope, however, for applying them to ethics. I discuss these arguments in detail in chapter 2. But for now, I set them aside.

Third, some arguments appeal to a putative tension between facts about moral disagreement and non-relativist accounts of moral concepts and semantics. Most famously, Horgan & Timmons (1991) argue against naturalistic moral realism by appeal to intuitions about disagreement in “moral twin-earth” scenarios.¹⁷ I do not discuss these arguments in this thesis, because their focus on debates about reference would, again, take us too far afield of our main topic of interest.

The plan of the chapter is as follows. Part 1 discusses Wright’s (1992) version of the argument, which focuses on a claim about the *mere possibility* of a certain kind of disagreement. Although Wright’s argument is interesting and intuitively compelling, I will argue that when it is unpacked in detail, it fails to undermine moral objectivism. I therefore turn, in part 2, to assess

¹⁶ (Harman, 1975) and (Prinz, 2007) advance arguments of this kind.

¹⁷ (Ridge, 2014) and (Bjornsson, 2012) also advance arguments of this kind.

a modification of Wright's argument which focuses on actual, rather than merely possible, disagreement. I argue that while this argument is more promising, it also has its shortcomings. Centrally, it appeals to empirical claims which, I argue, we do not have decisive reason to accept. I ultimately conclude that epistemic versions of the argument provide at most limited evidence against the objectivity of morality. However, also in the conclusion, I explain how the argument to be discussed in chapter 2 can be understood as a development of Wright's core idea.

A final proviso: although this argument has often been associated with relativism, it is not clear that, even if sound, it would provide evidence for relativism in particular, as opposed to evidence for non-objectivism in general. This is because the argument aims primarily to undermine moral objectivism, and a further step is obviously required to move from the rejection of objectivism to the acceptance of any particular non-objectivist theory. Since I will not ultimately place great weight on this version of the argument, I will not explore the interesting question of whether it favours any particular anti-realist theory over another. Rather, I will focus merely on whether the argument succeeds in undermining realism.

PART 1 – MERELY POSSIBLE MORAL DISAGREEMENT

The idea that some epistemic feature of moral disagreement undermines moral objectivity is a natural one, and was discussed in ancient times. Sextus Empiricus, for example, provides a nice statement of the intuitive worry:

Fire, which heats by nature, appears heating to everyone; and snow, which chills by nature, appears chilling to everyone: indeed, everything which affects us by nature affects in the same way everyone... But none of the so-called good things affects everyone as good, as we shall suggest. Nothing, therefore, is by nature good.¹⁸

Sextus goes on to give a list of moral disagreements, many of them cross-cultural. It is this proto-anthropological record which he takes to show that nothing "affects everyone as good".

¹⁸ (Gowans, 2000)

Although not one of Sextus' examples, Herodotus' anecdote (as mentioned in the introduction) is relevant:

Darius, after he had got the kingdom, called into his presence certain Greeks who were at hand, and asked – "What he should pay them to eat the bodies of their fathers when they died?" To which they answered, that there was no sum that would tempt them to do such a thing. He then sent for certain Indians, of the race called Callatians, men who eat their fathers, and asked them, while the Greeks stood by, and knew by the help of an interpreter all that was said – "What he should give them to burn the bodies of their fathers at their decease?" The Indians exclaimed aloud, and bade him forbear such language.¹⁹

Herodotus and Sextus took it that the existence of such disagreement undermines moral objectivism. And this is, again, a natural thought. There does seem to be something suspiciously *different* going on when we disagree about certain moral questions, as opposed to what goes on when we disagree about whether fire is hot. And the fact that people disagree so often about moral questions, but never seem to disagree about whether fire is hot, does seem, intuitively, to tell against the objectivity of these moral questions.

Yet we must be cautious. Clearly, not *all* disagreements undermine objectivity. After all, it is not *really* true that *everything* that "affects us by nature affects in the same way everyone". Even if no one has disagreed about whether fire is hot or snow is cold, people have disagreed about many things plausibly regarded as objective (for example, the age of the earth). These disagreements, surely, do not give us reason to doubt the objectivity of these matters. Thus, while the basic idea underlying the argument is a natural one, we must take care. Not all natural ideas are correct.

Of course, philosophy has moved on since Herodotus and Sextus, and more sophisticated versions of the argument are available. One of the most compelling, and one which seems to capture Sextus' basic idea, is Wright's (1992).

¹⁹ (Hdt.)

1. Wright on cognitive command

Wright's argument revolves around the idea of "cognitive command". For Wright, cognitive command is a domain-neutral way of framing debates around whether realism, in any particular domain, is true. Domains which satisfy the cognitive command constraint, for Wright, should be understood realistically. Domains which do not, should not. But what is cognitive command?

Begin with the pretheoretical notion of realism as representation, the notion of a discourse which functions to provide a map of an independent reality. And now, consider a paradigm case of representation, taking a photograph. When one uses a camera to take a photograph of a scene, one operates a device whose function is the production of representations of an independent reality (independent, that is, of the camera, its design and operation, the purposes we have for it, the role it plays in our society, and so on). But notice that, precisely because this is what cameras do, we expect photographs (taken from the same scene, in the same conditions, and from the same vantage point) to represent things *in the same way*. If two cameras produce *different* photographs, then it seems just platitudinous that they must either have been trained on a different scene (under different conditions, or from a different vantage point), or that one or both must have malfunctioned. And this seems intimately related to the facts about *what cameras do*, namely that they function to produce representations of a camera-independent reality.

This is suggestive of a claim about representation more broadly, namely that when our *beliefs*, and *language* function to represent an independent reality, we can come to think (and sincerely speak) in incompatible ways *only if* we are given divergent input (for example, if we have different evidence), or we have somehow malfunctioned (by, for example, failing to respond rationally to our evidence). In Wright's terminology, we can disagree only if we suffer from some form of *cognitive shortcoming*.

Wright provides the following nice statement of his idea:

[Cognitive command] is an attempt to begin to crystallise a very basic idea we have about objectivity: that where we deal in a purely cognitive way with objective matters, the opinions which we form are in so sense optional or variable as a matter of permissible idiosyncrasy, but are *commanded* of us – that there will be a robust sense in which a particular point of view *ought* to be held, and a failure to hold which can be understood only as a rational/cognitive failure.

Although somewhat imprecise, this claim seems somehow important and true. The objectivity of a certain domain of inquiry is, surely, connected somehow to the permissibility of the norms governing beliefs which concern that domain of inquiry. If belief about some subject matter was, on some fundamental level, a matter of permissible idiosyncrasy, it would be hard to see how the matters that such beliefs concern are not also, on some level, dependent on idiosyncratic features of believers.

Key to precisifying this claim is to notice that it is *normative*. In this way, it avoids a kind of circularity. For there is a trivial claim in the nearby vicinity of what Wright is saying, namely that if a domain serves to map an objective reality, then two people who disagree cannot both believe something *true*. Since this claim is formulated in *alethic* (rather than normative) terms, it comes close simply to rephrasing the claim that a domain is objective (hence the circularity). But, notice, Wright is *not* making *this* claim. Rather he is making a *normative* claim, about the *correctness* of beliefs, rather than the *truth* of beliefs. Thus, Wright's argument is non-circular, because it moves from a *normative* premise to a conclusion about objectivity.

Noticing this invites the question of what *kind* of normativity Wright has in mind. And this is a crucial question, because different interpretations of the normative clauses in Wright's argument lead to different versions of the argument. As an interpretative matter, it's plausible that Wright has *epistemic* normativity in mind. In particular, the normative claim he wants to make seems to be this:

Cognitive Command Constraint: A domain of inquiry satisfies the cognitive command constraint iff it is apriori that any disagreement formulated in the vocabulary of the

domain must involve a cognitive shortcoming, unless the disagreement involves vagueness, or run of the mill epistemic indeterminacy.²⁰

If this is right, then it sets up a general form of argument against realism, in any domain. In particular, we can say that for any domain D, if D satisfies cognitive command then realism about D is true, otherwise it is false. So if we can show, for any particular domain, that the cognitive command constraint is not satisfied, this will amount to showing, if Wright is correct, that realism about that domain is false.

This extends even more straightforwardly to objectivist non-realism (see introduction). For on this view, the single true moral code is supposed to be constructed by pure reasoning. It is therefore straightforward that moral disagreements involve some error in reasoning on at least one side (abstracting, again, from vagueness or epistemic indeterminacy).

If we now focus in on ethics, we can formulate an argument against the view that morality is objective. According to Wright, the objectivity of *any* domain of inquiry requires that the Cognitive Command Constraint is satisfied by that domain. Thus, if we can show that ethics does *not* satisfy the Cognitive Command Constraint, and if Wright is correct, then this will show that morality is not objective.

But before we can assess this argument, we need to first unpack some of the details in the Cognitive Command Constraint. In particular, we need to unpack the requirement of *apriority*, the provision for *vagueness*, the provision for run of the mill *epistemic indeterminacy*, and the notion of a *cognitive shortcoming*.

Section 2 Details of Wright's argument

First, why the apriority requirement? Why require that it be *apriori* that disagreement involves a cognitive shortcoming? Wright has in mind the following worry. Suppose, by happy coincidence, all thinkers were constituted in such a way that shortcoming-free disagreement

²⁰ This is a paraphrase, but I believe it captures Wright's formulation.

about some domain never occurs. Suppose this were to be true of comedy, or aesthetics, or matters of taste. For example, suppose an evil genius were to tamper with every thinker's psychology in such a way that every thinker was disposed to love the taste of marmite. Then, there would be no potential for shortcoming-free disagreement about whether marmite is delicious. For plausibly, someone who loves marmite cannot correctly deny that marmite is delicious. Nevertheless, and this is Wright's concern, there would remain, in this scenario, a legitimate worry about the objectivity of deliciousness discourse. For it would still be *conceivable* that there could be disagreement between those who love and hate marmite, which could be free from cognitive shortcoming. And the conceivability of this hypothetical scenario would, in Wright's view, undermine realism about matters of taste.

Moving on to vagueness and epistemic indeterminacy. Wright first accepts that realism is compatible with some degree of vagueness. It may be a vague matter whether Pluto is a planet, for example, but this is perfectly compatible with pretheoretical realism about astronomy. But vagueness can, plausibly, generate disagreements which are free from cognitive shortcoming. If two people disagree about whether some borderline tall person is tall, for example, it need not be the case that either suffers from any cognitive shortcoming.²¹ Therefore, there are at least some cases of shortcoming-free disagreement, compatible with realism, namely cases involving vagueness.

Similar considerations apply to run of the mill epistemic indeterminacy. It's plausible that there are cases where a given evidence set underdetermines rational belief. For example, suppose that a thinker's evidence makes a proposition, *p*, 70% probable. Then, it seems, he or she should believe *p*. But suppose the evidence makes *p* 69% probable, or suppose it makes *p* 68% probable, etc. At some point, the determinate rational requirement to believe shades into

²¹ It could be argued that shortcoming free thinkers should *suspend judgement* about vague matters. However, even if this sometimes true, it is not true of higher order vagueness. One is not rationally required to suspend judgement about a borderline case of a borderline case, since, by hypothesis, it is impossible to know whether the borderline case is borderline, and it is therefore impossible to know that one is rationally required to suspend judgement. Furthermore, one is never rationally required to ϕ if one cannot know that one is rationally required to ϕ . Therefore, one is not rationally required to suspend judgement in cases of higher-order vagueness. And therefore, there are at least some cases (those of higher order vagueness) where vagueness generates shortcoming-free disagreement.

indeterminacy. Thus, it seems plausible that there could be a shortcoming-free disagreement between a person who accepts p on the grounds of a given evidence set, and a different person who suspends judgement based on the same evidence.²² Wright wants to account for these cases, by bracketing out epistemic indeterminacy.

Finally, there is the notion of a cognitive shortcoming. This is where the real action lies, for this is the central normative notion underlying the cognitive command constraint. How does Wright understand it?

As we noticed above, it's important that this is a normative notion, not an alethic one. As we also noticed, this raises the question of what kind of normativity is relevant. And it's fairly clear that Wright understands the relevant normative notions *epistemically*. The idea seems to be that the Cognitive Command Constraint is satisfied just in case it is apriori that all disagreements involve some *epistemic fault*. Furthermore, it's fairly clear that Wright understands this notion in a broad way, so that it includes not only irrationality, but also defects in evidence, and any other shortcoming that can be reasonably be understood as epistemic in character (such as inattention to evidence, fatigue, limitations in imagination or intellectual creativity, and so on).

Epistemic shortcoming: an individual suffers from an epistemic shortcoming if they are irrational, if they lack access to evidence, or are in any other way less than ideal, from an epistemic point of view.

Section 3. Why Wright's argument is initially plausible

We are now abreast of details of the claims that Wright is committed to. We can now ask what motivates these claims. In particular, why think that the cognitive command constraint is *true*? We have seen that it is, in some respects, intuitively plausible. But can we say anything more

²² Again, this is at least true for cases of higher-order epistemic indeterminacy.

precise? It would be nice, in particular, if we could identify a *deductive* link between the possibility of epistemic-shortcoming-free disagreement and pretheoretical realism.

Suppose some proposition, p , is knowable and objective. Then, there is a clear sense in which anyone who considers p but fails to know it suffers from an epistemic shortcoming. Furthermore, if two people disagree about p , and p is objective, then at least one of them ipso facto fails to know that p (given the factivity of knowledge). Therefore, if p is objective and knowable, it must be the case that at least one party to any disagreement about whether p suffers from an epistemic shortcoming. Here is our deductive link.

Notice the role of the assumption that p is knowable, here. Wright accepts that a realist could resist his argument by claiming that some moral propositions cannot be known. Then, perhaps all that an epistemic shortcoming-free disagreement over p would show is that p is unknowable. However, Wright thinks that moral disagreements are necessarily knowable (in other words, that morality is *epistemically constrained*). Even if we do not accept this in general, we can say that if epistemic shortcoming-free disagreement over p seems possible, then either p is not objective, or p is not knowable. Thus, if it is plausible that some particular moral p is knowable, epistemically faultless disagreement over this p would provide some reason to reject realism.²³

Now that we have a more precise interpretation of Wright's argument, we can ask whether it seems sound. As we saw, the basic idea which underlies it seems plausible. But as we cautioned initially, apparently plausible ideas do not always turn out to be correct. Analysis of the details of the arguments is needed. And in fact, as I will now argue, when we try to be more precise about the notion of an epistemic shortcoming, we will find that there is no way of sharpening up this notion which is entirely happy for Wright's argument. The epistemic reading of Wright's Cognitive Command Constraint, we will therefore find, fails to provide a sound argument against moral objectivism.

Section 4 – Why Wright's argument fails

²³ (Tersman, 2006) interprets Wright in this way.

First, a proviso. As we noted above, Wright seems to have in mind (and his argument seems to require) a highly general notion of an epistemic shortcoming, which includes not only the notions of rationality and possession of evidence, but also things like proper *attention* to evidence (for if someone fails to notice some piece of their evidence, it is not obvious that they are irrational if they fail to base their beliefs on this evidence), full conceptual competence (for if someone fails to grasp relevant concepts, then they make an epistemically relevant mistake, but are not straightforwardly irrational), and a number of other factors. Here, for simplicity's sake, I will set these other factors to one side, and focus just on the notions of evidence possession and rationality. In what follows, I will take it as read that the cases in question involve no shortcomings of these other kinds (that all parties are properly attentive to their evidence, have considered all possibilities, etc.).

4.1 First problem – evidentialism

Now, an immediate problem is that according to a popular view, evidentialism, what it is rational for an individual to believe is *determined* by this individual's evidence. The problem here is that, recall, Wright presents the notion of a shortcoming-free disagreement as involving cases where parties have *the same* evidence (remember the analogy of a representational device, such as a camera – the idea was that different output must be the result of either divergent input (different evidence) or malfunction). But if evidentialism is right, then it's *trivial* that if two people have *the same* evidence, they cannot rationally disagree. And this is in itself plausible. Think about what it would look like in, for instance, the moral case for two people to have exactly the same evidence. It would, plausibly, involve them having all the same moral intuitions, and all the same background moral beliefs. But it would then be hard to see how two such people could coherently disagree. Thus, immediately, there is a worry that the Cognitive Command Constraint is trivially satisfied, for any discourse whatsoever. For it seems that, trivially, any two individuals who have *exactly the same* evidence cannot rationally disagree. Consider, for example, taste discourse. If two people have *exactly the same* evidence with respect to matters of taste, then, plausibly, they have, *inter alia*, all the same

gustatory and olfactory sensations, and all the same states of liking or disliking these sensations. It is hard to see how two such people could coherently disagree.

In light of this consideration, it seems clear that we should not formulate the notion of an epistemically shortcoming-free disagreement as requiring that two parties have *the same* evidence. Rather, we want to say something like that they have *equally good* evidence, or that neither's evidence is *defective*. This allows the constraint to range across domains where, intuitively, different evidence sets can be equally good (as one might think is the case in matters of taste).

4.2 Second problem – dilemma

With this proviso in place, we now need to precisify the notion of an epistemic shortcoming by fleshing out the notions of an individual's being fully rational, and an individual's having no defects in her evidence. But here, and this is the fundamental problem, Wright's argument faces a dilemma. At the most general level, there are two broad ways we could construe these notions. On a broadly *internalistic* picture, rationality consists in responding appropriately to one's evidence, and one's evidence consists in the things one is capable of responding to, paradigmatically – one's nonfactive mental states (such as one's beliefs, experiences, and intuitions). On the other hand, assuming a broadly *externalistic* approach, rationality consists in being reliably connected up with the facts, and this involves having belief forming processes which are reliable in the sense of being truth-conducive. This approach also sits most happily with the view that the notion of *evidence* is somehow factive. For example, an externalist might hold that one's evidence is the set of things one knows.²⁴

Now the problem (and this is why this dichotomy amounts to a dilemma) is that neither of these two approaches is entirely happy for Wright's argument. Let me explain why.

4.3 Internalist horn of the dilemma

²⁴ C.f. (Kelly, 2014).

First, on the internalist conception, the problem is that shortcoming-free disagreement is *too easy* to come by. For on the internalist conception, there simply is no sense to the notion that one person's evidence can be *better* than another's²⁵. This is because *all* it takes for one to have evidence to the effect that p is for one to have a mental state that supports p. And this makes it almost trivially possible that two thinkers could have equally good evidence for incompatible propositions. Thus, for example, consider a disagreement between a thinker who is radically deceived by a cartesian evil demon, and a thinker who is not so deceived. On the internalist conception (assuming both thinkers have comprehensive sets of beliefs about the external world) there is no sense in which the deceived thinker's evidence is inferior to the non-deceived thinker's evidence. Furthermore, on the internalist conception, both are rational just to the extent that they respond appropriately to their evidence. Thus, their disagreement will be shortcoming-free (they will have equally good evidence, and they will both be fully rational²⁶), but they will systematically disagree about the external world. And this undermines the intuitive force of Wright's claim, which, recall, was that it is apriori that objectivity of a discourse exerts a tight *discipline* or *command* on thinkers' beliefs. For here is a paradigmatically objective domain where, as we can now see, shortcoming-free disagreement is fairly obviously possible.²⁷

Now, it might be objected that this example concerns a domain which is epistemically *unconstrained*. It could then be argued that this objection misses the mark, for Wright holds that Cognitive Command applies only to epistemically constrained domains. But there are two problems with this response. First, on the internalistic conception, it seems that shortcoming-free disagreement will be straightforwardly possible, *regardless* of the domain in question. So pick *any* domain which is both plausibly epistemically constrained and plausibly objective. For example, take prudential normativity. It seems that we can construct a Cartesian

²⁵ Admittedly, *comprehensiveness* could be viewed as a superiority-determining feature of evidence sets, on this view. But this point does nothing to help the Wrightian internalist here.

²⁶ They can also, surely, be free of all *other* defects, at least inasmuch as these defects are acceptable to an internalist (e.g. they can surely be equally attentive to their evidence).

²⁷ It might be objected that the thinker who is deceived fails to know that he or she is deceived, and that this is an epistemic shortcoming. But notice that *on the internalist conception*, this cannot count as an epistemic shortcoming, because, by hypothesis, the deception is indiscernible to the thinker who is deceived.

Demon scenario for this domain. That is, it seems that there is at least one possible world where there is a thinker who is radically deceived about what would make their life go well, and also a thinker who is not so deceived. Thus again, assuming prudential normativity is objective, the intuitive connection between the possibility of shortcoming-free disagreement and objectivity is severed. Now, it might be objected that prudential discourse is *not* objective. But then, one wonders if there are *any* epistemically constrained domains that *are* objective. And if not, then it looks like the epistemic constraint is doing all the work, and the considerations around disagreement themselves add nothing to the case against objectivism.

Perhaps one good candidate for an epistemically constrained domain, which is objective, and where, on the internalist conception, shortcoming-free disagreement seems impossible, might be mathematics (at least in certain sub-domains, such as arithmetic). For it's hard to imagine a case where someone is free from epistemic shortcomings and is radically deceived about mathematics. Therefore, there is a position in logical space for the view that the Cognitive Command Constraint is non-trivial. However, this position would involve holding that *every* epistemically constrained domain, other than mathematics, is not objective (e.g. including prudential and epistemic rationality). And this is a view that many moral anti-realists would hesitate to endorse.

In any case, here is the second problem with this response. Recall that the argument was that if a given *p* is knowable and objective, shortcoming-free disagreement about that *p* is impossible, because if a given *p* is knowable and objective, then any individual who fails to know this *p* suffers from an epistemic shortcoming. But, on the internalistic conception, reflection on Cartesian Demon cases calls these claims into question. For consider the claim that if a *p* is knowable and objective, then there is a clear sense in which any individual who considers this *p* but fails to know it suffers from an epistemic shortcoming. If we have in mind an internalistic notion of an epistemic shortcoming, this claim looks dubious. For on an internalistic notion it does not seem that the demon-deceivee suffers from an epistemic shortcoming. Nevertheless, it is controversial whether the mere possibility of demon-

deception, with respect to some propositions, undermines our knowledge of these propositions.²⁸

This is of course not to say that skeptical arguments based on such scenarios are clearly mistaken. However, the conclusion we have reached is that Wright's argument turns fundamentally on a similar claim – that individuals who have radically different (internalistically construed) evidence sets are possible. It therefore seems open to the realist to frame such possibilities as ones involving radical deception, along the lines of standard radically skeptical scenarios. And the only special application such possibilities seem to have in ethics is then that ethics is claimed to be epistemically constrained. Disagreement, per se, seems to be playing a limited role here. Furthermore, the argument would also seem to commit the ethical non-objectivist to a wide-ranging and radical skepticism, since all domains of inquiry seem to admit the possibility of such scenarios. This is a price which many metaethical non-objectivists would be unwilling to pay.²⁹

Thus, the internalistic option does not look entirely happy for the Wrightian anti-realist. Again, in summary, the internalist conception makes it *too easy* to conceive of disagreements which are shortcoming-free in the relevant sense. And as a result, the argument becomes difficult to distinguish from the kinds of concerns which motivate generalised radical skepticism.

4.4 Externalist horn of the dilemma

This conclusion might motivate a proponent of Wright's argument to opt for the broadly *externalist* option. But this leads to the opposite problem. Given the factivity inherent in this conception, it now becomes *too hard* to conceive of shortcoming-free disagreements at least in a theoretically neutral way.

²⁸ (Kelly, 2014).

²⁹ (Enoch, 2009) defends a similar response.

Admittedly, on this conception, it's plausible that if a *p* is knowable and objective, then any agent who considers this *p* but fails to know it suffers from an epistemic shortcoming. So the connection between objectivity and shortcoming-free disagreement is on solid ground. However, the new problem is that if a domain is objective, then shortcoming-free disagreement will be *trivially* impossible, on the externalist conception. And there will therefore be a serious concern that, if we take an externalist interpretation, Wright's argument begs the question against the moral objectivist.

To see this, let's focus in on the following question: do false background *moral* beliefs count as defects in one's evidence for one's particular moral beliefs?³⁰ On the externalist conception, the answer should be: yes – a false (background) moral belief is *ipso facto* a defect in one's evidence set.³¹ But if this is right, shortcoming-free disagreement seems inconceivable, given objectivity. For if two people disagree over some moral *p*, then (given the principle-governed nature of moral discourse) it seems that they must either disagree over some further background *ps*: *p*₁...*p*_{*n*}, or they must be incoherent. This is because if they have *exactly the same* background moral beliefs, but disagree about some particular moral question, it seems that one party must be at least subtly incoherent (as we have already noticed). But if they have *different* background moral beliefs, then *ipso facto* (given objectivity) one of them has some false background moral beliefs, and thus suffers from a cognitive shortcoming. Thus, either way, the disagreement is not shortcoming-free, if it concerns an objective domain.

A proponent of Wright's argument could object, here, by appealing to cases of what we might call disagreements-in-weak-principle. For example, suppose two people disagree about whether punishment is ever intrinsically good. Person A, a utilitarian, thinks that punishment is never intrinsically good. Person B, a weak retributivist, thinks that punishment is sometimes intrinsically good *but only marginally so*. Suppose the weight that B assigns to the value of punishment is so tiny that her retributivist intuitions are always swamped by considerations that the utilitarian would also accept. And thus, A and B end up agreeing about all actual

³⁰ (Tersman, 2006) and (Enoch, 2009) emphasizes this question.

³¹Strictly speaking, one would say that a false belief doesn't count as part of one's evidence at all, so that one is irrational inasmuch as one's beliefs are sensitive to it. Nothing important hangs on the choice of words.

cases of the form “such and such individual should receive punishment of such and such severity”. It might be objected, then, that the weak retributivist and the utilitarian disagree about a particular matter of principle, *without* there being any divergence in any of their *other* beliefs, and *without* there being any incoherence on either side. Rather, it seems possible that in this case they have *all the same* background beliefs, but also that their belief-systems are equally and perfectly coherent. Thus, this kind of case (cases of disagreement-in-weak-principle) seems to be a counterexample to the claim that, on the externalist conception, all cases of disagreement must involve either a false background belief or a failure of coherence.

32

However, notice that cases like these turn on absences of *consideration* (as mentioned above, the Cognitive Command Constraint requires that the individual considers whether p). For in the weak retributivist case, there will be *hypothetical* disagreements over statements of the form “such and such individual should receive punishment of such and such severity”. After all, we can imagine cases where there are *no* utilitarian considerations in favour of punishing a wrongdoer. And in such a case, the weak retributivist would be required (on pain of incoherence) to disagree with the utilitarian. Of course, actual thinkers will not normally consider such cases. And if neither side considers such cases, neither side will have *any beliefs at all* about them – thus avoiding incoherence. However, as we noticed above, there is a clear sense in which failure to consider a relevant proposition counts as an epistemic shortcoming. Thus, cases of weak-disagreement-in-principle *do* involve epistemic shortcomings, and therefore provide no solution to the externalist horn of the dilemma.

Returning to the main thread, recall that our focus was on the question of whether false moral beliefs count as epistemic shortcomings. We have found that when we reflect on this question, we run into a problem for the Wrightian externalist. The problem is essentially a dialectical one, namely that, on the externalistic conception, the Wrightian argument becomes question-begging against the realist. This is because, on the externalist conception, epistemic-shortcoming-free disagreement is trivially ruled out by objectivity.

³² Thanks to Guy Fletcher for suggesting this objection.

A different way of framing what is essentially the same point is to note that the externalistic conception explains epistemic concepts in *alethic* terms. And thus, on the externalistic conception, the argument fails to avoid the problem of *circularity*, mentioned above. This is because the question of whether shortcoming-free disagreement is possible will, on this conception, be too tightly connected with the notions of truth, and thus of objectivity for us to get any *independent* purchase on this question. So even if we come to the argument suspending belief about objectivity and truth, we will then be simply unable to say whether epistemic shortcoming-free disagreement is possible, because on the externalistic conception, questions about the latter collapse immediately into questions about the former. Alternatively, if we come to the argument either accepting or rejecting objectivism, then our views about the possibility of shortcoming-free disagreement will simply fall out of our prior commitments, in a dialectically unacceptable way.³³

4.5 Actual disagreement

I see only one way for a proponent of Wright's argument to proceed here, and this is to move to the claim that there *actually are* epistemic-shortcoming-free moral disagreements. This would allow them to grasp the internalistic horn of the dilemma, for it's more plausible that *actual* instances of shortcoming-free disagreement, construed internalistically, would undermine the possibility of knowledge, given objectivity.

To see this, consider two different scenarios involving (internalistically) epistemic-shortcoming-free disagreement. In the first scenario, consider a world where half of people perceive *shape* properties just as we do, and therefore perceive some objects as being five-sided. Suppose, in this scenario, that the other half of people are specifically incapable of perceiving the property of five-sidedness (suppose they perceive all five-sided objects as either four-sided or six-sided). Suppose this applies not only to visual perception, but also to all other ways in which sidedness can be ascertained. Now, it's quite plausible that the denizens of *this world* get their knowledge undermined by this disagreement. For half of them

³³ (Tersman, 2006) (Enoch, 2009) and (Shafer-Landau, 2003) all offer a version of this response to the Wright-style argument.

see shapes one way, and another half see shapes another way, and there is no perspective-neutral way of rationally resolving the disagreement. However, as we have seen, it's controversial whether the *mere possibility* of a world like this undermines our actual knowledge of five-sidedness.

In the second scenario, however, suppose that this same world is not merely possible, *but actual*. In other words, imagine that in the *actual world*, there is a roughly 50-50 split between people who believe that there are five-sided things, and people who believe that there are no five-sided things. Furthermore, imagine that these disagreements are (internalistically) shortcoming-free, and that everyone knows this. In *that case* (in the case of *actual*, widespread, shortcoming-free disagreement) it *is* plausible that our actual knowledge would be undermined. After all, assuming that shape discourse is objective, such systematic disagreement would provide straightforward evidence of unreliability of beliefs about shape. And if the disagreement is epistemic-shortcoming-free (even if only in the internalistic sense), then this evidence cannot be *defeated*, even in principle.

Transpose, now, to the moral domain. Many have argued that there actually is widespread, shortcoming-free disagreement about morality. If this is right, then the actual scenario, with respect to our moral beliefs, is relevantly similar to the *second* five-sidedness scenario. And therefore, it would follow that we *actually* lack moral knowledge (given the assumption that morality is objective). Again, the key point is that this "actualising" move avoids the internalistic horn of the dilemma discussed above. For it is not at all trivial that there actually is widespread (internalistically) shortcoming-free disagreement (where, we saw, it was trivial that internalistically shortcoming-free disagreement is possible). And furthermore, it *is* plausible that if there actually are a lot of (internalistically) epistemically shortcoming-free disagreements then this fact would undermine our knowledge, given objectivity. Nor does this claim collapse into radical skepticism (whereas, we saw, the version of the argument which appeals to a claim about the mere possibility of this kind of disagreement does, plausibly, entail a radical and domain-general skepticism).

In part 2, I pursue this version of the argument. The key question is whether the anti-realist can show that there actually is widespread (internalistically) shortcoming-free disagreement. Note that we have not yet discussed the question of whether morality is epistemically constrained, and I here set this question aside. For there is a route to an argument against objectivism, either way. Even if morality is not epistemically constrained, the claim about disagreement would still entail that we lack knowledge of many things we pretheoretically take ourselves to know. For example, if it turns out that the ancient romans were epistemically shortcoming-free, in the relevant sense, vis a vis their beliefs about the permissibility of gladiatorial combat, then it would follow that if morality is objective, we do not know that gladiatorial combat is morally impermissible. However, it's extremely plausible, pretheoretically, that we know that gladiatorial combat is impermissible. Thus, if there is widespread epistemically shortcoming-free disagreement about this matter, this fact would undermine objectivism, even if morality is not in general epistemically constrained.

PART 2 – ACTUAL MORAL DISAGREEMENT

So is there epistemically shortcoming-free moral disagreement? And if there is, how much is there? Is our actual situation comparable to the hypothetical situation (involve disagreement about five-sidedness) described above? Before we try and answer this question, let's try and sharpen up what we mean by it. What kind of actual disagreement, exactly, are we looking for?

5. Clarifications

First, discussion of this question often focuses on disagreement which would persist in *wide reflective equilibrium*. Briefly, reflective equilibrium is a process of taking stock of one's moral beliefs and intuitions, filtering out those which are untrustworthy, and attempting to bring one's moral beliefs and intuitions into coherence with each other, and with one's non-moral knowledge.³⁴ The notion of wide reflective equilibrium seems a good precisification, for our purposes, of the notion of freedom from (internalistically construed) cognitive shortcoming.

³⁴ For discussion see (DePaul, 1986) (Rawls, 1951) (Daniels, 1979) (Cath, 2016).

Thus, the first thing to say is that we're looking for disagreement which would persist in wide reflective equilibrium. Notice that this is *not* to say that we're looking for disagreement between individuals who already are in wide reflective equilibrium. Rather, we're looking for disagreements which would *persist*, even if both parties were to reach states of wide reflective equilibrium, from their actual starting points. If there are disagreements of this form, then this would, plausibly, undermine knowledge of the topics of these disagreements, given objectivity. (Compare the five-sidedness case, even if various parties to the disagreement suffered from certain shortcomings, this would not solve the skeptical worry, if the disagreement would persist after idealisation away from these shortcomings).

However, not just *any* disagreement of this form would suffice to undermine objectivity, for two reasons. First, it matters how *many* people there are on each side of the disagreement. If the vast majority of people, throughout history, have taken one side of an issue, and only a tiny minority have taken the other side, then it will not be plausible that this disagreement undermines knowledge, given objectivity, even if it is shortcoming-free in the relevant sense. This, plausibly, is why disagreement about colour, between colour blind people and normally sighted people, does not clearly undermine colour knowledge, on the assumption that colour properties are objective. The fact that colour blind people are in a significant minority partially explains why it seems natural to regard such people as suffering from a cognitive shortcoming, rather than merely different. Similar considerations apply to psychopaths. The fact that there are relatively few of these people makes it reasonable to regard them as morally defective, rather than merely different, so that the moral disagreements we have with them do not undermine our moral knowledge (assuming objectivity) even though these disagreements may persist in wide reflective equilibrium. Of course, there won't be a bright line beyond which we can say that there are enough people on both sides of the disagreement. But there should be enough people that it will not be plausible to regard these people as suffering from some straightforward defect.

The second point we need to take into account is that it matters what the *scope* of shortcoming-free disagreement is, in terms of *which* moral propositions are subject to such disagreement. If

shortcoming-free disagreement turns out only to affect propositions that are plausibly not known, or such that it is plausible (even on an objectivist view) that they lack determinate truth conditions, this would be less clearly problematic for the objectivist. For, first, inasmuch as it is less pretheoretically plausible that we know these propositions, objectivists could bite the bullet and accept that we lack knowledge. And second, if a proposition is indeterminate, then it will be unsurprising that we lack knowledge of this proposition. Consider these examples. On the one hand, suppose that the most we can show is that disagreement about matters such as whether the repugnant conclusion³⁵ is true, or exactly what percentage of their income affluent westerners should donate to charities, or whether the doctrine of double effect is true, etc. would persist in wide reflective equilibrium. This, plausibly, would do little to undermine objectivism, because it's not obvious that we have knowledge of such matters.³⁶ On the other hand, if it could be shown that disagreement between modern westerners and ancient Romans about whether gladiatorial games are permissible would persist in wide reflective equilibrium, this would be problematic for the objectivist. This is because it's pretheoretically plausible that we know that gladiatorial games are not permissible, and any view which fails to explain why this is true is thereby at a disadvantage.

Now that we know what kind of disagreement we are interested in, how can we get purchase on whether this kind of disagreement exists? Proponents of arguments of this kind take two broad approaches. The first approach essentially involves arguing, from the armchair, that certain familiar moral disagreements are unlikely to be resolved in wide reflective equilibrium, based on our general knowledge about these disagreements. The second is experimental, and involves conducting research specifically designed to investigate whether particular moral disagreements depend on any factor which would be eliminated in wide reflective equilibrium.

³⁵ (Parfit, 1984)

³⁶ The general idea here is that if shortcoming-free disagreement involved only the kinds of thorny questions that have long puzzled moral philosophers, this should be acceptable to the moral objectivist. For the objectivist could then view ethics as no more problematic, from an epistemic point of view, than philosophy more generally.

This is sometimes framed as an attempt to show that none of a broad range of so-called “defusing explanations” obtains. One way an explanation could “defuse” the threat to objectivism would be by showing that the relevant disagreement would not persist in wide reflective equilibrium. For example, if the realist could show that the disagreement counterfactually depends on a non-moral disagreement, then it would follow that this disagreement would not persist in wide reflective equilibrium. But this is not the only way in which an explanation could defuse the threat posed by a given disagreement. For if it could be shown that the disagreement involves participants *talking past* one another or, as mentioned above, that the subject of the disagreement is indeterminate, this would defuse the threat to objectivism without showing, *per se*, that the disagreement would not persist in reflective equilibrium.

Although this list is incomplete, here are some of the most important defusing explanations:

Incoherence: if the disagreement involves incoherence on the part of one of the parties, this could mean that the disagreement would be resolved in wide reflective equilibrium. For people in wide reflective equilibrium hold sets of beliefs which are fully coherent. This may seem a weak requirement. But the power of coherence reasoning should not be underestimated. Consider, for example, Peter Singer’s famous (1972) argument in favour of charitable giving. This argument appeals only to claims which Singer takes to be already accepted by his audience, and yet the conclusion of the argument is radically revisionary of ordinary moral beliefs. Importantly, *strength* of belief plays a role, here, in dictating the proper response to incoherence. Singer’s argument turns on the assumption that the premises he advances are not only accepted, but *strongly* held by his audience. Bearing this in mind, there is potential for a defusing explanation with wide scope. If realists could argue that all or almost all people, across times and cultures, share certain fundamental moral beliefs, and that proper reasoning from these fundamental moral beliefs exerts fairly tight control over the beliefs which could be held in wide reflective equilibrium, then this would entail that almost all moral disagreements would be resolved in wide reflective equilibrium, because they involve some (potentially subtle) incoherence. This strategy may nicely fit with an approach

to moral psychology which draws inspiration from a Chomskyian conception of language, according to which the structure of language is constrained by a universal grammar.³⁷

Other irrationality: incoherence is one form of irrationality, but there are other forms (such as bias, parochialism, wishful thinking, and other motivated reasoning). One way in which a plausible defusing explanation could be built up out of these notions is by pointing out that moral disagreements often involve conflicts among competing interests. For example, it may be that moral disagreements across social classes (over the proper rate of taxation, or who should own the means of production) would dissolve if neither class was guilty of motivated reasoning. Equally, it might be that moral disagreements between groups in direct conflict (such as Israelis and Palestinians) would dissolve if neither side was guilty of motivated reasoning.

Non-moral disagreement: many moral disagreements seem to depend on non-moral disagreements. For example, moral condemnation of homosexuality in Europe is apparently largely a consequence of the influence of Christianity. Thus, disagreements over gay rights may in some sense depend on non-moral disagreements about what God has or has not prohibited. Cannibalism and human sacrifice have also been justified on religious grounds. One can also imagine more familiar examples: for example, disagreements over whether the death penalty is just may depend on disagreements over whether it is an effective deterrent. Since false non-moral beliefs would be eliminated in reflective equilibrium, it may be that moral disagreements which seem to depend on non-moral disagreements would not persist in reflective equilibrium.

Talking-past: some apparent moral disagreements may involve application of shared moral principles in different contexts. For example, senilicide is practised in some communities which live in hostile environments. It is possible that these communities view senilicide as a necessary evil, and justify it on utilitarian grounds, given the danger that may be posed to the group by expenditure of resources on keeping older and weaker individuals alive. If that's the

³⁷ For useful discussion of coherence and objectivism, see (Sayre-McCord, 1996). For discussion of fundamental moral agreement, see Sauer (2019).

case, then there may be no real disagreement between these societies and our own. For we might agree that senilicide would be justified in these circumstances. More subtle variations on this explanation involve failures of translation. Some cultures frame their normative discourse in terms of concepts that are not obviously translatable into our moral concepts. For example, if a Polynesian thinks that something is *tapu*,³⁸ but we think it's morally permissible, it's not *obvious* that there is a disagreement, because it's not obvious that the Polynesian concept of *tapu* contains our concept of moral impermissibility (although it may do). Similar problems may arise when we attempt to translate the *non*-moral beliefs of other cultures. For other cultures' beliefs about non-moral states of affairs may involve subtle culturally encoded meanings, which attempts at translation may fail to capture, and this may mean that we fail to notice background non-moral disagreements. (This is the problem of so-called "situational meaning", see (Moody-Adams, 1997)). A related problem is that it may not always be possible to tease apart moral and non-moral disagreements. For example, does disagreement about abortion depend on disagreement about when it is permissible to kill a person or on disagreement about which individuals count as persons? To understand whether this is the case in a cross-cultural context requires quite a lot of understanding about a given society's overall world-view. And this understanding may not always be easy to come by.

If this battery of defusing explanations seems powerful, note that there is no obvious entailment from the premise that some moral disagreement *involves* some factor to the conclusion that the moral disagreement *counterfactually depends* on this factor. Consider the case of non-moral disagreement. As research on so-called "moral dumbfounding" has shown, it is not necessarily the case that people abandon their moral beliefs when they abandon the non-moral beliefs which they take to justify these moral beliefs. Rather, non-moral justifications often play the role of post-hoc rationalisation. In the experiments on dumbfounding, many subjects declined to abandon their belief that incest is wrong even after it is pointed out that the incestuous couple would not have children, that no one else would find out, and so on.³⁹

³⁸ See Joyce (2001).

³⁹ (Dwyer, 2009)

Cases of apparent defusing explanations could work like this. For example, even if rich and poor people cite disagreement about non-moral, economic matters (e.g. the Laffer curve) to explain their incompatible beliefs about what the tax rate should be, it's nevertheless possible that these economic beliefs play a post-hoc, rationalising role, and that the relevant moral disagreement would remain, even after the relevant non-moral disagreements were eliminated. Similar considerations apply to many of the other possible defusing explanations. Thus, in order to offer a genuine defusing explanation, it needs to be shown not only that disagreements involve defusing factors, but also that these disagreements counterfactually depend on these factors. And in many cases, this will be difficult to do.⁴⁰

5.1 Non-Experimental

Let's first consider the non-experimental approach to this question. This, again, essentially involves arguing, from the armchair, that certain familiar moral disagreements are unlikely to be resolved in wide reflective equilibrium, based on our general knowledge about these disagreements. Take, for example, the disagreement between us and the ancient romans, over whether gladiatorial games are permissible. What could defuse this disagreement? It could be that the disagreement involves some aspect of ancient roman religion. But it's not obvious what aspect this might be. Admittedly, the games were ultimately outlawed for religious reasons. But this appears to be because Christianity (which became dominant in Rome) positively forbade gladiatorial games, rather than because the games lost support from the old religion. Nor is it obvious that any other non-moral beliefs could explain the disagreement. It's possible that the ancient Romans failed to recognise an incoherence between some of their basic moral beliefs and their belief that the games were permissible, so that they could have been convinced by some philosophical argument. But it's hard to see how this could be shown conclusively. It's also possible that given that Roman society employed a distinctive set of normative concepts, which are not obviously translatable into our moral concepts, there is no genuine disagreement between us and the Romans. Again though, it's hard to see how this could be shown.

⁴⁰ (McGrath, 2010) makes this point.

This historical example therefore seems inconclusive. Instead, consider an interesting and much discussed contemporary case, namely that of the Yanomamo people of South America. In a famous (1967) study, aspects of Yanomamo society were documented by the anthropologist Napoleon Chagnon. At this time, the Yanomamo were one of the few remaining relatively uncontacted traditional societies. According to Chagnon, salient features of Yanomamo culture are violence and male chauvinism.⁴¹

If Chagnon's account is right, there is much moral disagreement between us and the Yanomamo. Although there is also much non-moral disagreement between us and the Yanomamo, it's not obvious how this could explain the relevant moral disagreement. Nor is it plausible that, for many of the things that we and the Yanomamo disagree about, these things are indeterminate or unknown to us. Chagnon's own explanation of why Yanomamo culture is this way is, essentially, that given the social environment in which the Yanomamo live, dispositions towards violence are adaptive. There is controversy about this explanation. But we may grant it for the sake of argument. The important point is that this would not count as a defusing explanation, because it's not particularly plausible that Yanomamo people endorse some higher level principle which says something like "dispositions toward violent behaviour are admirable in environments where this behaviour is adaptive, but not otherwise". Rather, even if the Yanomamo's environment is a *cause* of their particular moral beliefs, it's not plausible that facts about their environment play any explicit role in their belief system. More plausibly, they simply view tendencies towards violence as fundamentally admirable. And would likely continue to do so even if they lived in an environment where these traits were no longer adaptive.

It seems, then, that it is somewhat plausible that disagreement between westerners and the Yanomamo would persist in wide reflective equilibrium. However, it's hard to establish this claim conclusively. There is always room for a realist to argue that the disagreement turns on

⁴¹ Chagnon's account is not uncontroversial among anthropologists, nor is it perfectly comfortable for modern sensibilities vis a vis non-western cultures. But this does not obviously undermine its accuracy. And a detailed discussion of whether Chagnon's account is descriptively accurate would go beyond the scope of this thesis.

subtle incoherence, or that it turns on culturally encoded and hard to detect non-moral disagreement. It is hard to see how it could be shown that this is not the case.

5.2 *Experimental*

Perhaps, then, we can do better by approaching the problem experimentally. A number of experimentally minded philosophers have attempted to focus in on particular disagreements, and gather more robust data, in order to help us make a less speculative assessment. The most impressive example of this kind of research is Nisbett & Cohen's (1996) work on the "southern culture of honour" and Abarbanell and Hauser's (2010) work on Mayan attitudes towards the act-omission distinction. First, Nisbett & Cohen studied differences in attitudes towards violence as a response to personal insults in the American south, relative to the American north-east. Using a number of experimental methods, they found that non-hispanic white southerners are much more likely than non-hispanic white northerners to think it morally justified to respond to insults with violence, and to think that those who fail to do so are deficient in masculinity. Nisbett & Cohen hypothesise that this disagreement is ultimately explained by the fact that southern non-hispanic whites are predominantly descended from herding societies, whereas northerners are predominantly descended from farming societies, and that dispositions to respond to personal insults with violence are adaptive in herding cultures (essentially, because livestock is relatively easier to steal, so that successful herders will cultivate reputations for violence which deter potential rustlers). Nisbett & Cohen also argue that it is unlikely that this disagreement can be explained using any of the standard defusing explanations. Here, again, it's useful to notice that it is probably *not* the case that both northerners and southerners accept some principle of the form "respond violently to insults if you find yourself in a social situation which makes this adaptive, but not otherwise". What is more likely is that moral intuitions were, via some historical process, internalised by the two cultures, and that members of these cultures now possess vestigial and non-instrumental moral attitudes, such that defending one's honour is seen as non-instrumentally good by southerners, but not by northerners.

In another study, Abarbanell and Hauser investigated whether rural and lesser educated Mayan subjects make the familiar act/omission distinction, using interviews based on trolley-problem style moral dilemmas. Intriguingly, Abarbanell and Hauser found that unlike Westerners and urbanised, educated Mayans, rural less-educated Mayans fail to make this distinction (for example, they fail to judge that killing is worse than letting die). Furthermore, Abarbanell and Hauser were careful to attempt to account for some of the defusing explanations we have considered. In particular, they used native-speaking research assistants to check translations of questionnaires into the Mayans native language, and to select the most appropriate translations of relevant moral words. They also took care to examine Mayan attitudes towards relevant causal factors. In particular, they found that Mayans did not fail to make the act/omission distinction in virtue of non-moral beliefs (e.g. about causation).

What to make of these experiments? It does seem that they provide some evidence that the disagreements in question would persist in wide reflective equilibrium. However, it is again hard to see how this evidence could be decisive. In the former case, realists could always argue that the relevant intuitions might be amenable to reasoning in light of non-moral facts. For example, perhaps southerners would be amenable to the argument that the culture of honour produces greater overall harms, or might feel that their intuitions are “debunked”, on reflection, by the historical-genealogical facts Nisbett and Cohen cite. In the latter case, even if (what is not obvious) Abarbanell & Hauser succeed in showing that the disagreement cannot be defused, it also seems plausible that moral questions about the act/omission distinction are imperfectly determinate, or at least that we do not know the answers to these questions. So in both cases, it’s still reasonably plausible that the disagreements could be explained in ways congenial to objectivism.

Although it is somewhat plausible that at least some of these disagreements (or relevantly similar ones, these are just a few examples) *would* persist in reflective equilibrium, it seems hard to produce an argument to that effect which is likely to move a realist. Bearing in mind additional vagaries about the *scope* of shortcoming-free disagreement which would be required in order to pose a problem for objectivism, the case against objectivism, based on present empirical evidence, seems inconclusive.

6. *Can we bypass the empirical questions?*

An interesting way to continue the argument, in spite of this apparent empirical impasse, is this. Richard Rowland (2017), a proponent of the kind of argument we have been discussing here, argues that even if there is not decisive reason to think that there are moral disagreements which would persist in wide reflective equilibrium, objectivism is nevertheless undermined by the mere fact that this is a salient possibility. This is because, as Rowland points out, many experts have believed that moral disagreement would persist in wide reflective equilibrium, and if many experts have believed some proposition, then this is itself a non-trivial positive reason to believe this proposition. But moreover, according to Rowland, if there is a non-trivial reason to believe a proposition which *entails* that some theory should be rejected, then there is non-trivial reason to believe that this theory should be rejected. Thus, Rowland claims, we have non-trivial reason to reject moral objectivism, even if we lack decisive reason to accept that widespread moral disagreement would persist in wide reflective equilibrium, simply in virtue of the fact that many experts have *believed* that widespread moral disagreement would persist in wide reflective equilibrium.

Sarah McGrath (2010) makes a similar point. While she notes the difficulty of showing that actual moral disagreements would persist in wide reflective equilibrium, McGrath argues that it is not clear that these points should comfort the objectivist. According to McGrath, if the persistence of this disagreement would undermine objectivism, then there is a sense in which uncertainty about the persistence of disagreement should lead us to uncertainty about objectivism. As McGrath puts the point “in effect, the credence that it is reasonable to give to the speculative claim [that disagreement would be resolved in wide reflective equilibrium] sets an upper bound to the credence that it is reasonable to give to moral realism”. Thus, on McGrath’s view, if it is true that persistence of disagreement in reflective equilibrium would undermine objectivism, then if we should be agnostic about whether disagreement would persist, we should be agnostic about objectivism.

We should assess these arguments independently, as they are subtly different. Begin with Rowland's. Grant that the opinions of the relevant experts give us *some* reason to think that moral disagreement would persist in ideal wide reflective equilibrium. And grant that this gives us some reason to reduce confidence in objectivism. How weighty is this reason?

Relative to the overall balance for and against objectivism, it seems that the answer is: not very. For notice that this kind of reason can be found for *any* substantive philosophical thesis. After all, it's a commonplace that for any substantive philosophical thesis, *p*, one can find a non-negligible number of experts who deny *p*. This is a simple consequence of the paucity of consensus in the discipline. But this means that there is a reason of the relevant kind to reduce confidence in *every* substantive philosophical claim. Indeed, the mere fact that there are many expert metaethicists, who are also anti-realists, means that there are already reasons of the relevant kind to reduce confidence in objectivism. So if all that the argument from disagreement adds to the case against objectivism is one additional reason of this kind, then its dialectical significance is relatively minor.

McGrath defends her argument in a different way. She invites us to consider an analogy with libertarianism about free will. As she points out, some libertarians are committed to thinking that the natural world is, at some level, indeterministic. Plausibly, this means that if we should be agnostic about whether the natural world is indeterministic, then we should be *at most* agnostic about whether libertarianism is true. Equally, according to McGrath, if objectivism requires convergence in ideal conditions, and we should be agnostic about whether there would be such convergence, we should be at most agnostic about whether objectivism is true.

Note, though, that *both* libertarianism and hard determinism come with empirical commitments. After all, many hard determinists are committed to the view that the natural world is deterministic. But of course, if we should be agnostic about whether there is indeterminacy in nature, then we should also be agnostic about whether nature is deterministic. So it seems that, by McGrath's reasoning, we should be agnostic about *both* libertarianism and determinism, if we should be agnostic about whether there is indeterminacy. Likewise, *both* moral objectivism and moral anti-objectivism make predictions

about moral disagreement. Thus, it seems that by McGrath's reasoning, if we should be agnostic about whether moral disagreement would persist in wide reflective equilibrium, we should be agnostic about *both* objectivism and anti-objectivism.

This seems too strong, though. Perhaps what is more plausible is that the argument from disagreement merely *fails to discriminate* between the two views. This might require us to suspend judgement, if there were no other sources of evidence. But to the extent that there are other arguments which might support one view over another, the upshot seems to simply be that the argument from disagreement does not help adjudicate between the two views. Thus, the conditional does not seem to hold: it does not seem to be true that if it is rational to suspend judgement about whether actual moral disagreement would persist in wide reflective equilibrium, then it is rational to suspend judgement about whether moral objectivism is true.

8. Conclusion

The aim of this chapter has been to assess the implications of epistemic versions of the argument from disagreement. What we have found is that the implications are limited. We saw that epistemic claims about *merely possible* disagreement fail to undermine objectivism. However, we saw that claims about *actual* disagreement *might* undermine objectivism, if certain empirical conjectures were to turn out true. However, we have seen that the relevant empirical conjectures, While somewhat plausible, are hard to establish conclusively. And thus, epistemic arguments from disagreement provide only limited reason for skepticism about moral objectivity.

This, however, does not mean that considerations about disagreement, as such, fail to undermine objectivism. This chapter has explored a specifically epistemic reading, but other versions of the argument are available. Of particular interest are recent debates over so-called "faultless disagreement". As we will see, the notion of fault employed in these debates can be understood as a non-epistemic interpretation of Wright's notion of a cognitive shortcoming. (Indeed, one of the key proponents of these arguments (Kolbel, 2004a) is explicitly influenced by Wright (Kolbel, 2002).) It is to these arguments which we will now turn, in chapter 2.

CHAPTER 2

ABSTRACT: In this chapter we discuss the second of two versions of the argument from disagreement, namely the argument from so-called faultless disagreement. We begin by sketching an application of the argument to aesthetics. We then take an experimental turn, using data from empirical psychology to defend an ethical application of the argument. We will conclude that the argument from faultless disagreement does provide evidence for a form of moral relativism.

Our aim in this first part of the thesis has been to understand how far arguments from disagreement provide evidence for relativism. We have so far discussed the epistemic argument from disagreement, and found it only somewhat successful. In this chapter, we discuss a different version of the argument – one which appeals to the phenomenon of so-called “faultless disagreement”. We will find that this argument is more successful, and does establish that there is evidence for relativism (and as we will see in a moment, for truth relativism in particular). In subsequent chapters, we will move on from the topic of arguments from disagreement, and discuss the details of the relativist theory.

Our focus here will be in particular on the argument developed by Kolbel (2004a).⁴² Kolbel’s interest is in domains where there is a strong presumption in favour of some form of subjectivity; domains such as aesthetics, personal taste, and comedy. But there is obvious scope to apply his argument to ethics, and this will be our project. Although (as with any philosophical argument) it is controversial whether Kolbel’s argument goes through even in aesthetics (etc.), these are the “easy cases” in the sense that they are the cases where the argument’s fundamental data point is clearest, and in which the argument looks most likely to succeed. Therefore, we will begin by sketching out the basic structure of Kolbel’s argument in terms of one of these domains. Doing this is helpful, because it is helpful to separate out the

⁴² Although this chapter is much influenced by Kolbel’s classic (2004a), a similar argument was independently developed by (Lasersohn, 2005). In recent years, there have been many excellent discussions of this topic, including: (Egan, 2014) (Kolbel, 2005b) (Zeman, 2020) (Eriksson & Tiozzo, 2016) (Palmira, 2015) (Hills, 2013), (Eriksson, 2016), (Pietroiusti, 2020) (Stojanovic, 2019) (MacFarlane, 2014) (Schafer, 2011) (Schafer, 2014) (Björnsson, 2015).

structural features of the argument from the question of whether it can be successfully applied to particular domains, such as ethics, and because it is easiest to see the structural features in the easy case. Once we have a sketch of the argument in front of us, we will turn to the ethical application.

1. Faultless disagreement in aesthetics

Consider the following case of disagreement:

Novel: Adrien and Baringa disagree about whether Hemingway's *The Old Man and the Sea* is a good novel. Adrien thinks it's a good novel, but Baringa thinks it's not a good novel. Suppose both Adrien and Baringa are properly acquainted with the novel, know a lot about literature in general, and are in the right frame of mind to appreciate the novel. But suppose, nevertheless, that they disagree, because the novel just strikes Adrien as good and strikes Baringa as bad.

Bearing in mind, here, the discussion of chapter 1, it seems obvious that Adrien and Baringa's disagreement could be epistemically faultless, at least in an internalistic sense. Kolbel's insight though is that, more than this, it seems that disagreements like this can be (and often are) faultless in a stronger respect. According to Kolbel, there is *no sense* in which either Adrien or Baringa is at fault, or can be said to have made a mistake. So the point is not merely that neither party is irrational, nor that neither party has defective evidence, etc., but rather, that neither party can be faulted relative to *any* norm to which they are subject⁴³, and thus that (for both parties) there is *no sense* in which changing their mind would constitute an improvement.

Now what gives bite to this claim, and this is the second key move in Kolbel's argument, is that it is independently plausible, that *truth* is *normative* for *belief*. That is, while a false belief need not be *irrational* (nor otherwise *epistemically* defective), there is nevertheless a clear sense

⁴³ At least in virtue of believing as they do. Of course, they might be faultable in other ways (e.g. maybe Adrien has irrational beliefs about some other, unrelated topic).

in which any belief which is false is thereby *incorrect*, such that changing one's mind would constitute an improvement. Indeed, some have argued that *all it is* to be a belief, rather than some other mental state, is to be correct when true and incorrect when false (a conjecture can be false, for example, but need not thereby be incorrect). This is controversial⁴⁴, but even those who do not accept this strong view tend to accept the weaker claim that there is *some* normative sense in which any false belief is incorrect, and therefore that any individual who believes something false has thereby made some kind of mistake.⁴⁵

As Kolbel notices, though, once we combine this (relatively) uncontroversial claim with the claim that Adrien and Baringa's disagreement can be *fully* faultless, there is a straightforward route to relativism. For we can now see that if a disagreement is fully faultless, then it follows that neither side's belief is false (for given the normative claim, if one party's belief is false, then this fact should show up in our normative intuitions about the case). And if neither side's belief is false, then (abstracting from cases of indeterminacy) both side's beliefs are true. If we accept that there can be disagreements where both parties' beliefs are *true*, though, then this just amounts, surely, to accepting a form of relativism.

Nor does the proviso about indeterminacy vitiate the force of this argument. For notice that the case in question does not seem to be one which plausibly involves any form of indeterminacy. After all, *The Old Man and The Sea* is a paradigmatic case of a determinately good novel, it won a Pulitzer prize, and its author was awarded a Nobel prize. Indeed, it's plausible that Adrien can *know* that *The Old Man and the Sea* is a good novel, at least inasmuch as aesthetic knowledge is possible in general. This makes it all the more striking that the disagreement can apparently be fully faultless. Even though *The Old Man and the Sea* is a paradigmatic case of a good novel, and even though it seems that Adrien knows that *The Old Man and the Sea* is a good novel, it still seems that Baringa can think that *The Old Man and the Sea* is *not* a good novel without having made a mistake of *any* kind.

⁴⁴ (see Chan, 2013)

⁴⁵ This is, for a number of reasons, independently attractive (c.f. (McHugh, 2012)).

Let's spell out Kolbel's argument with a bit more perspicuity. It appeals to four claims. The first and second are independently motivated, theoretical, commitments.

Truth norm: if one believes something untrue, one has made a mistake.

Univocality in Disagreement: two individuals disagree only if one rejects a content that the other accepts. (more on this in a moment)

The third and fourth are intuitive data points:

Appearance of Full Faultlessness: in cases like Adrien and Baringa's, neither individual seems to have made any mistake.

Appearance of Disagreement: in cases like Adrien and Baringa's, the differences of opinion seem to amount to genuine disagreements.

If Truth Norm and Univocality hold, then if the appearances described by the Faultlessness and Disagreement claims are taken at face value, truth relativism follows. This is because, if we accept that Adrien and Baringa have made no mistakes (Appearance of Full Faultlessness), then it follows (by the Truth Norm) that they believe true contents. But if we also accept that they disagree (Appearance of Disagreement), then it follows (by Univocality) that there is a particular content which one party accepts and the other rejects. Thus, if we accept all these claims at face value, we are led to the conclusion that Adrien and Baringa believe contradictory contents, both of which are true. (Schematically, Adrien believes p , Baringa believes $\neg p$, and both believe something true). Only *truth relativism*⁴⁶ is consistent with this, though. So what we are ultimately led to is the conclusion that truth relativism is true.

Above, I said that if we accept that there can be disagreements where both parties' beliefs are true, then this, surely, just amounts to accepting a form of relativism. We can now recognise the role Univocality in Disagreement plays in precisifying this point. For if Univocality is false,

⁴⁶ See the introduction for presentations of truth and content relativism.

then it is consistent with *content relativism* that there can be disagreements where both parties' beliefs are true, and the form of relativism we accept is content relativism. But if Univocality is true, then accepting the Appearances at face value leads us, if we accept Truth Norm, to accepting truth relativism.

1.1 Contours of the debate

So this is the basic structure of the argument that we are going to attempt to apply to the ethical case. Before we move on, though, it is worth briefly sketching the contours of the debate in the aesthetic case. For the same main moves are relevant in both domains.

It is helpful to break down the available responses by the available alternative semantic theories: content relativism⁴⁷, invariantism⁴⁸, and expressivism^{49, 50}. For content relativists, the natural response is to deny Univocality in Disagreement. Content relativists, recall from the introduction, hold that the contents of Adrien and Baringa's beliefs are straightforwardly compatible. On a toy view, what Adrien believes is that *The Old Man and the Sea* is highly rated by Adrien's sensibility, and what Baringa believes is that *The Old Man and the Sea* is highly rated by Baringa's sensibility. If this is right, then there seems to be no obstacle to accepting Truth Norm and the Appearance of Full Faultlessness, because it is clear that both of these contents can be true. However, if one accepts Univocality in Disagreement, then it follows, on this view, that Adrien and Baringa do not genuinely disagree. Thus, content relativists must either deny Univocality in Disagreement, or deny that Adrien and Baringa genuinely disagree. Some content relativists seem happy, given certain provisos, to accept that Adrien and Baringa may not genuinely disagree.⁵¹ However, a more promising strategy is to offer an account of disagreement which does not require Univocality. Most straightforwardly, content

⁴⁷ e.g. (Finlay, 2014)

⁴⁸ e.g. (Schafer, 2011)

⁴⁹ e.g. (Gibbard, 2003)

⁵⁰ Although there are few aesthetic error theorists, error theory may seem to constitute a fourth option, when we turn to the case of ethics. However, the error theorist's distinctive claims are metaphysical, not semantic. And at the level of semantics, error theorists endorse invariantism. Therefore, since the question of faultless disagreement is a purely semantic one, error theory is not a fourth option here.

⁵¹ (Stojanovic, 2007).

relativists can argue that Adrien and Baringa disagree in virtue of having clashing non-cognitive attitudes, and not in virtue of having incompatible beliefs.

What if one is a semantic invariantist? In that case, one will have no reason to deny the otherwise plausible Univocality in Disagreement, for invariantism straightforwardly entails that Adrien and Baringa's disagreement is univocal. From an invariantist point of view, the natural response to the problem is, instead, to deny that the disagreement is fully faultless. For if we take the Appearance of Full Faultlessness at face value, and accept that neither party has made a mistake of any kind, then Truth Norm will lead us, on invariantist assumptions, inexorably to a contradiction. Instead, invariantists will want to *explain away* the appearance of full faultlessness. They will not want, implausibly, to deny that there is *any* interesting kind of faultlessness present in cases like Adrien and Baringa's. However, they may argue that disagreements like Adrien and Baringa's are merely *epistemically* faultless, albeit in a particularly strong and systematic respect.⁵²

If one is an expressivist, finally, one has two main options. First, if one is a quasi-realist (a la Blackburn (1984), Gibbard (2003)) one should offer essentially the same response as the invariantist. After all, quasi-realism aims to be equivalent to realism with respect to the linguistic data. And furthermore, deflationism about truth is incompatible with relative truth (and thus incompatible with a situation where A believes *p*, B believes $\neg p$, and both believe something true). The details of the epistemological story may differ. But this need not concern us here.⁵³

Not all expressivists are quasi-realists. All contemporary expressivists seek equivalence with realism on *some* points (e.g. to my knowledge all hold that moral judgements are beliefs). But some stop short of claiming full equivalence (Schroeder, 2008). For these expressivists, it is not obvious whether faultless disagreement should be admitted. On the one hand, these expressivists can explain the Appearance of Disagreement with relative ease – by appeal to

⁵² See e.g. (Schafer, 2011).

⁵³ Classical noncognitivism and hybrid expressivism are left out here. But few contemporary metaethicists endorse classical non-cognitivism. And discussion of hybrid theories would take us too far afield.

disagreement in attitude. On the other hand, it is not obvious whether expressivists of this kind can explain the Appearance of Full Faultlessness. If they say, as quasi-realists tend to, that calling a moral judgement mistaken just amounts to expressing a favourable non-cognitive stance towards that judgement, then this will rule out faultless disagreement. For one cannot coherently (in whatever sense of 'coherent' quasi-realists prefer) express a favourable stance towards two conflicting moral judgements. Alternatively, they may say that whether an aesthetic judgement is mistaken depends on the non-cognitive attitudes of the one who makes the judgement. This would allow them to account for faultless disagreement, because it would allow them to grant that neither of two conflicting judgments need be mistaken. However, it would open them up to the accusation that their theory has subjectivist consequences (Jackson & Pettit, 1998) (Suikkanen, 2009). Quasi-realists, for the most part, seem concerned to avoid this problem (Schroeder, 2014). Inasmuch as they are this way inclined, the options available to them seem again to parallel those available to the invariantist.

So these are the contours of the debate. We have seen that the argument from faultless disagreement depends on two intuitive data points and two theoretical commitments. We have also seen that opponents of truth relativism, in order to resist the argument, must either reject one of the theoretical commitments, or explain-away one of the data points. As we have also seen, the key moves are (for invariantists and expressivists) explaining away the Appearance of Full Faultlessness, and (for content relativists) denying Univocality in Disagreement.

Of course, we have so far seen only the barest outlines of these moves. We have not evaluated them. And so, we do not yet have any reason to think that the argument from faultless disagreement succeeds, even within the domain of aesthetics. After all, for all we have seen so far, it may be that one of these moves is sound, and the argument can be resisted – even within the domain of aesthetics.

Nevertheless, we will not pursue the aesthetic argument in any further detail here. While the contours of the debate are broadly the same, this does not necessarily mean that the details of

the various lines of argument will play out in the same ways (for example, the correct theory of aesthetic disagreement may differ from the correct theory of moral disagreement). Let's now turn, therefore, to the question we are really interested in: whether an argument of this form can be applied successfully in the domain of ethics.

2. Faultless Moral Disagreement

At least initially, the prospects for a moral version of the argument may seem poor. The most plausible cases of faultless moral disagreement are those which puzzle moral philosophers: trolley cases, duties to future generations, the doctrine of double effect, etc. However, invariantists might plausibly hold that any appearance of faultlessness which attaches to such questions is due to indeterminacy, or due simply to our inability to know the answers to such questions.⁵⁴ More challenging for invariantism would be the claim that disagreement over moral propositions which seem clearly true, to us, would be faultless. Such a claim would also be more faithful to the cases of paradigmatically faultless disagreement. Thus, imagine that Adrien and Baringa disagree not about whether *The Old Man and the Sea* is a good novel, but rather about whether it is permissible to enslave foreigners, or kill those who disrespect you. It's at least somewhat less plausible that such a disagreement could be faultless in the same way.

However, we can do a bit more to pump relativist intuitions. As Bernard Williams (1981) once emphasised, the plausibility of moral relativism increases at distance. Consider a disagreement between a contemporary western liberal, and a Bronze Age Greek chieftain or mediaeval samurai (Williams' examples). Suppose the chieftain thinks that it is OK to enslave foreigners, and the samurai thinks it is OK to kill those who disrespect him. Of course, we disagree. But it is harder, perhaps, to see that our interlocutor, in this kind of case, must be at fault.

⁵⁴ (Schafer, 2011) appeals to indeterminacy as does (Parfit, 2011). See also n21 of chapter 1.

Matters are complicated by the fact that, as we saw in chapter 1, we cannot rule out that even these disagreements involve some epistemic shortcoming (i.e. that they would not persist in wide reflective equilibrium). Thus, they may not amount to faultless disagreements, in the relevant sense, even if relativism is true.

To clarify the situation, then, suppose we take the parameter of cultural distance to its (hypothetical) extreme:

Hyperspace Bypass: The Vogons are a species of extraterrestrials. A representative vagon, Jeltz, believes that it would be morally permissible to demolish the earth to make way for a hyperspace bypass. Arthur Dent, a representative human, believes that demolishing the earth to make way for a hyperspace bypass would be wrong. Both hold these beliefs in states of wide reflective equilibrium.⁵⁵

For many people, this case has some of the intuitive timbre of Novel. This is because, for many people, it is hard to see that Jeltz is making a mistake, even though it is intuitively clear that he and Arthur disagree. Of course, every reader will have the intuition that demolishing the earth to make way for a hyperspace bypass would be wrong (just as, at least for Hemingway fans, it is clear that the *Old Man and The Sea* is good). But taking a step back from a morally engaged point of view, many find it hard to see that at least one party must be making a purely cognitive mistake. Therefore, cases like this may constitute an intuitive data point for an argument from faultless moral disagreement, in a way which methodologically parallels the role of cases like Novel in the Kolbellian argument for aesthetic truth relativism.

However, reflection on such cases, alone, is not sufficient to construct an argument for moral relativism. For although some have the intuition that cases like *Hyperspace Bypass* instantiate the Appearance of Full Faultlessness, others do not. And these differences in intuitive judgement seem to be fairly tightly connected to prior theoretical commitments.⁵⁶ This is in

⁵⁵ See (Adams, 2007). Notice that the relativist is committed only to the *possibility* of faultless moral disagreements, and not to the thesis that there actually are faultless moral disagreements, hence the science-fictional case.

⁵⁶ C.f. (Enoch, 2009).

marked contrast to the aesthetic domain, where theorists of all stripes accept, at least, that there is some substantive kind of faultlessness worth explaining. Given this lack of consensus, an argument for relativism which appeals to these cases alone may be dialectically ineffective.

3. Going Experimental

There is a way through this impasse, however. Instead of focusing on philosophers' intuitions, I propose that we focus on the ways in which ordinary people actually think and speak. Conveniently, recent work in empirical psychology gives us some substantial insight into how ordinary people think and speak about moral disagreement. And importantly, the results of this empirical research are independent of philosophers' intuitions for or against realism. Therefore, putting an experimental spin on arguments from faultless disagreement makes it possible to extend these arguments to ethics in a dialectically effective manner.

It is important to understand that this move is licensed by the fact that truth relativism is a claim in natural language semantics – a claim about the meanings of the words people ordinarily use. In these experiments, competent speakers of English are presented with an assertion of an English sentence and are asked to agree or disagree with the assertion, with the aim being to elicit their judgement as to whether the assertion is correct. Now, for most assertions, the assent of the majority of competent speakers obviously does not provide decisive reason to accept that the assertion is correct. This is because the correctness of most assertions depends not only on the semantic properties of the associated sentence, but also on the way the world is extra-linguistically, and the judgements of competent speakers are at best fallible guides with respect to the extra-linguistic facts.

For some assertions, however, this is not the case. The correctness of some assertions does not depend on the way the world is, but only on the semantic properties of the associated sentence. In these cases, the judgements of the majority of competent speakers (in the right conditions) cannot be mistaken, because it is these judgements that *determine* the semantic properties of the relevant sentences. Since the assertions which feature in the experiments

discussed below are of the latter type, there can be no objection of the form “most people may just be mistaken”. There can, of course, be legitimate objections of the form “these experiments fail to capture the right judgments, under the right circumstances”. I will attempt to anticipate some objections like this below.

Here is the experimental data. In a pair of recent studies, Sarkissian et al. (2011) and Khoo & Knobe (2016) investigated ordinary speaker’s attitudes towards faultless moral disagreement. The participants in these studies were divided into three groups, each of which was asked to consider a moral disagreement. The first group received a vignette describing a disagreement between two members of their own culture. The second group received a vignette describing a disagreement between a member of the participants’ own culture and a member of an isolated tribe. And the third group received a vignette describing a disagreement between a member of the participants’ own culture and a member of an alien species. All participants were then asked whether at least one party to the disagreement must be mistaken/wrong.

At least *prima facie*, disagreement with this statement indicates endorsement of faultless disagreement. Now admittedly, ‘mistaken’ and ‘wrong’ can be understood in different ways. Therefore, while the latter response indicates endorsement of faultless disagreement simpliciter, it’s not obvious how this response should be interpreted. For now, let’s focus simply on whether ordinary speakers endorse faultless disagreement, simpliciter. In section 4, we’ll discuss possible disambiguations.

Now importantly, both studies found a striking relation between cultural distance and intuitions of faultlessness. While the first group (same culture condition) tended to reject faultless disagreement, the second group (other culture condition) tended to suspend judgement, and the third group (extraterrestrial condition) tended to endorse faultless disagreement.

The first of these studies (Sarkissian et al., 2011) tested for intuitions of faultlessness only. The results of one experiment from this study are displayed in fig.1.

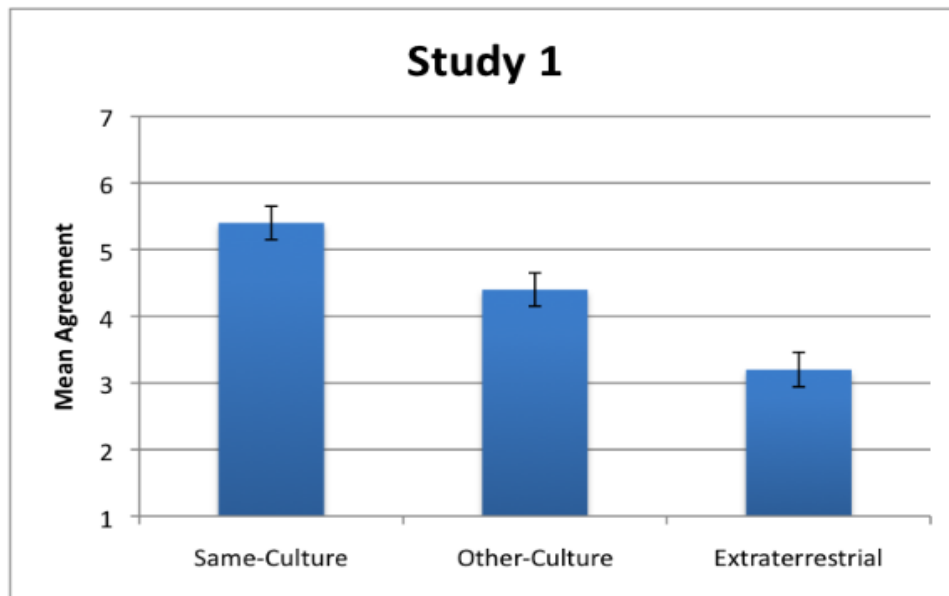
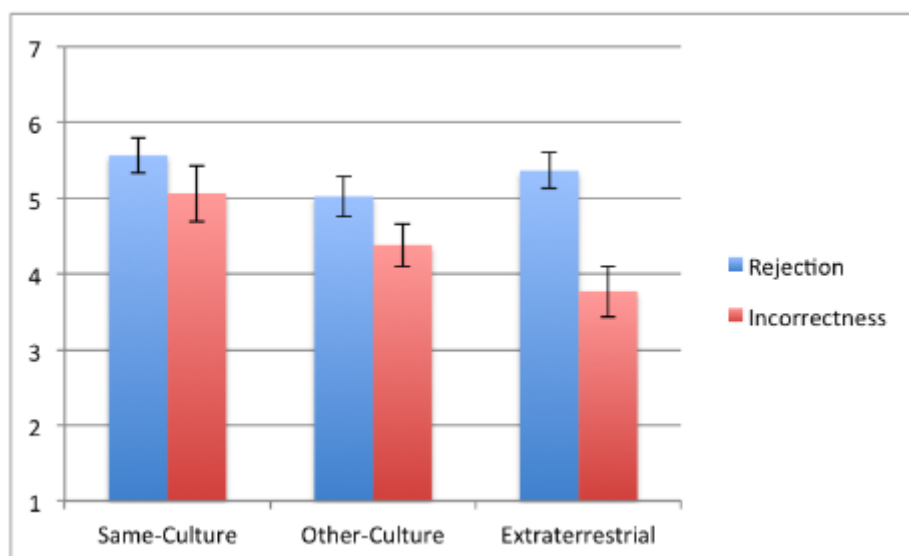


Figure 1. Mean agreement with the statement that at least one must be wrong. Error Bars show SE mean.⁵⁷

The second study tested, in addition, for intuitions about disagreement. This study found a similar pattern with respect to faultlessness, but also found that all three groups regarded the relevant differences of opinion as involving genuine disagreements. The results of one experiment from this study are displayed in figure 2.



⁵⁷ The wording of this question varies slightly across the two studies. (Khoo & Knobe, *ibid*) use the standard “mistaken” wording, whereas (Sarkissian et al., *ibid.*) switch between the “mistaken” wording, and wording in terms of whether at least one party must be “wrong”.

Figure 2. Mean agreement with the statement that it is appropriate for one party to reject the other's statement by saying "no" (rejection), and mean agreement with the statement that at least one must be incorrect (incorrectness). Error bars show SE mean.

These results are, of course, open to interpretation. But *prima facie*, they suggest that ordinary speakers endorse the possibility of faultless moral disagreement, at least in contexts involving extreme cultural distance.

3.1 Worries About the Experiments

One common concern about these results turns on the anecdotal observation that some undergraduate students are explicit moral relativists. Thus, since experiments like these often exclusively recruit undergraduate students, these results may merely reflect explicit theoretical commitments of these students, rather than any broader discursive practice.

This is a legitimate concern with respect to (Sarkissian et al., *ibid*), which exclusively recruited undergraduates. However (Khoo & Knobe, *ibid.*), recruited participants with average ages of 28, 29 & 36 (across three experiments), all through Amazon Mechanical Turk. As we've seen, the two studies found roughly the same results. So there is reason to doubt that these results depend on idiosyncratic undergraduate relativism.

Another worry is that participants may be illicitly influenced by the experimental materials. In particular, participants may assume that information about cultural distance is provided because experimenters take this information to be relevant. If participants assume that *experimenters* think the information is relevant, this may illicitly influence them in favour of faultless disagreement.

If this were the case, however, we would expect to see a similar effect when cultural distance is varied in cases of *non-moral* disagreement. However, (Khoo & Knobe, *ibid.*) tried and failed to find any such effect. In one experiment, half of participants were presented with a non-moral disagreement, and half were presented with a moral disagreement. Importantly, *both*

received the extraterrestrial condition. But in this experiment, participants tended to reject faultless disagreement in the non-moral case, while tending to endorse it in the moral case. Thus, it does not seem that an effect of the relevant kind explains the main findings.

Finally, one might wonder whether ordinary speakers really posit *genuine disagreement* in the culturally distant cases. This is because Khoo & Knobe tested for attitudes to disagreement by asking participants whether it would be *appropriate* for a third party to reject the extraterrestrial's claim, using the word 'no'. Linguistic felicity of rejection using 'no' is normally taken as indicative of genuine disagreement. However, one might worry that participants interpret the question of whether rejection is appropriate in some irrelevant sense (e.g. morally appropriate, polite).

However, Khoo & Knobe also tested for perceptions of disagreement directly, by asking participants whether characters in the relevant vignettes disagree. And this test found roughly the same pattern: consistent endorsement of *disagreement* across variation in cultural distance, with a tendency to endorse *faultlessness* only at greater cultural distance. Thus, even if we throw out the denial data, Khoo & Knobe's results still suggest that ordinary speakers endorse faultless (genuine) disagreement.

This concludes my discussion of methodological objections. I now want to consider a worry about the strength of these results. The worry is that it may seem unclear, from a truth relativist point of view, why there is not stronger endorsement of faultless disagreement in the extraterrestrial condition. Relativism, after all, straightforwardly predicts that the extraterrestrial disagreement is faultless. Why, then, did the experiments find only *weak* endorsement, in the extraterrestrial condition?⁵⁸

However, notice that the results show similarly weak *rejection* of faultless disagreement in the same-culture condition. Turning the objection on its head, we might suggest that this is equally puzzling, from an invariantist point of view. But in fact, neither result is puzzling. This is because no matter how plausible a statement is, there will always be a non-negligible

⁵⁸ Thanks to an anonymous reviewer at *Australasian Journal of Philosophy* for raising this objection.

proportion of respondents who deny it. For example, in a related study, a majority of participants endorsed faultless disagreement with respect to the statement ‘frequent exercise usually helps people to lose weight’.⁵⁹ Thus, results of this type inevitably contain a certain amount of noise. And this plausibly explains, consistent with invariantism, why we see only weak rejection of faultless disagreement in the same-culture condition. Equally though, it explains, consistent with relativism, why we see only weak acceptance in the extraterrestrial condition.

A second problem with this objection is that it fails to consider the broader *pattern* of results. After all, if invariantism is true, cultural distance is straightforwardly orthogonal to the possibility of faultless disagreement. However, these results show quite clearly that respondents take cultural distance to be relevant. And this is a favourable result for the relativist.

Finally, it might be objected that since the extraterrestrial condition involves a far-fetched case, responses for this condition are unreliable. I can think of three motivations for this objection. First, it might be objected that, quite generally, judgements about exotic cases are unreliable. Second, it might be objected that, in particular, laypeople’s intuitions about such cases are unreliable. Third, it might be objected that laypeople’s judgements about exotic cases, even if normally reliable, are unreliable in experimental contexts.

If the first objection is intended, the worry is a reasonable one. However, the worry needs to be motivated by evidence to the effect that judgments about exotic cases are unreliable. For instance, while the case in the extraterrestrial condition is somewhat exotic, it also bears many similarities to ordinary cases of disagreement. To cast doubt on the reliability of our judgements about the former, then, we need to be told why we should be worried about its exotic features, rather than reassured by its familiar features. Furthermore, the worry at issue entails that the argument presented here has many companions in guilt, for many influential philosophical arguments appeal to exotic cases (e.g. Descartes’ demon).⁶⁰ If the second worry

⁵⁹ (Beebe & Sackris, 2016: 915)

⁶⁰ Thanks to an anonymous reviewer at *Thought* for suggesting this response.

is intended, this worry may be compelling, but only if the case is complex and thus difficult for laypeople to understand. For example, ordinary judgments about Putnam's twin-earth case may be unreliable, because philosophical sophistication is needed to understand the case. However, not all exotic cases are like this. Ordinary speakers should be able to understand cases which are merely unlikely, while conceptually straightforward (e.g. Descartes' demon). And in such cases, it is hard to see why ordinary judgements would be unreliable. Thus, since the *experimental* case is quite straightforward, the second worry does not undermine the argument presented here. Finally, if the third worry is intended, the objection is that laypeople's judgements about exotic cases, even if normally reliable, are unreliable in experimental contexts. However, it is hard to see why this would be true *in general*, and specific worries about the design of the relevant experiments have been anticipated above.

2.3 Concluding remarks about the data

Let us take stock. We have seen that the data do suggest that ordinary speakers endorse the possibility of faultless moral disagreement in some cases. However, we have also seen that ordinary speakers' attitudes appear to be sensitive to cultural variation. Thus, the data provide at most qualified support for relativism. Many moral relativists defend an *individualistic* view, where truth is relativised to the standards of individual assessors.⁶¹ But rather than being supported, these individual relativist views are undermined by the data. This is because these views entail that *intra*-cultural disagreement can be faultless. And the data suggest that ordinary speakers deny this.

In this respect, the data we have discussed confirm the results of earlier studies, such as (Goodwin & Darley, 2008)⁶² which do not focus on cultural variation. In these studies, similar vignettes are presented, but the cultural backgrounds of parties to disagreements are left unspecified. It is reasonable to think (as Sarkissian et al. argue) that participants in such experiments default to the assumption that characters in the vignettes are members of the

⁶¹ E.g. (Finlay, 2014).

⁶² See also (Beebe et al., 2015) (Beebe & Sackris, 2016).

same culture. Thus, it is unsurprising, given the results we have surveyed here, that these studies have tended to find that ordinary speakers reject faultless moral disagreement.

On the other hand, this does not undermine the support that *societal* or *group* relativism (that is, the view that truth is relativised to the standards of an assessor's group) does receive from the data. This is because the main data which we have reviewed show exactly the pattern of results that group/cultural relativism predicts. And furthermore, as Sarkissian et al. persuasively argue (p485-6) these data are consistent with the results of (Goodwin & Darley, *ibid.*) and other similar studies. In what follows, societal relativism will be assumed.⁶³

4. Explaining the Data

Now that we are abreast of the experimental data, we need to think about how best to explain it. For just as there are ways for alternative theories to explain the Appearances, in the Novel case, equally, there are ways for alternative theories to explain the experimental data. To briefly recap, and update, the key data points are:

Disagreement Data: participants viewed the relevant differences of opinion as genuine disagreements.

Faultlessness Data: participants took it to be possible that both parties' statements are correct.

The argument for truth relativism also turns on these theoretical commitments:

Truth Norm: there is a sense in which any belief which is not true is incorrect.

Univocality in Disagreement: two people disagree only if one believes a content which the other rejects.

⁶³ See chapter 5 for discussion of the details of the societal elements of the theory.

Thus the experimental Data Points replace the Appearances in the standard, Kolbelian, version of the argument sketched in section 2. Otherwise, the argument is structurally similar. Thus, unsurprisingly, the map of available responses is broadly the same. Invariantists and expressivists can argue that participants in the experiments interpret “correct” and cognates, epistemically. Content relativists can argue that even if participants take the cases to involve genuine disagreements, we cannot take this as evidence for Univocality – because participants may simply be picking up on a disagreement in attitude.

4.1 An Epistemic Interpretation

Begin with the epistemic interpretation. An initial worry about this interpretation is that the experimental results, given the way the questionnaires are worded, are *not* most naturally interpreted in an epistemic sense. Here are the statements which were presented to the participants, first (Khoo & Knobe):

Since Jim and Sam have different judgements about this case, at least one of their judgements must be *incorrect*.

And second, (Sarkissian et al.):

Since your classmate and Sam have different judgements about this case, at least one of them must be *wrong/mistaken*.⁶⁴

Imagine you are someone who thinks, even if only implicitly, that moral disagreements could be epistemically faultless, but *could not* be fully faultless (i.e. faultless in the sense relevant to Truth Norm). In that case, what would be the most natural response to this statement? To my ear, at least, disagreeing with the statements in question would not be a natural way to express this combination of views. Rather, the natural thing for such a person to say, at least to my ear, would be “yes, at least one must be *incorrect/wrong/mistaken*.” While adding “Of course,

⁶⁴ Some of Sarkissian et al.’s experiments use the “wrong” wording, others use the “mistaken” wording.

both might be justified in their beliefs, given their differing cultural backgrounds and upbringing. But still, at least one must be *incorrect/wrong/mistaken*".⁶⁵ This is because all three of these words ('incorrect', 'wrong', 'mistaken') are such that it is in general much more natural to hear them as adverting to a truth norm, than it is to hear them as adverting to an epistemic norm.

Ideally, though, we would like to have an experimental test for this hypothesis. And fortunately, Sarkissian et al. provide just such a test. They split participants into four groups. The first group received a vignette describing a cross-cultural moral disagreement and was asked whether at least one party must be mistaken. The second group received a vignette describing a cross-cultural *non*-moral disagreement, and was asked the same question. The third and fourth groups received the same vignettes, respectively, but were asked whether at least one party must *lack good reasons* to believe as they do.

Why does this help? The key point is that participants are presented with *the same* scenarios, but while one group is given a prompt with a distinctly epistemic flavour, the other group is given the standard "mistake/incorrect" prompt. This is exactly the test we need for the epistemic hypothesis, because what the epistemic hypothesis says is that participants in the main experiments *just are* reading the standard prompt in an epistemic way. Therefore, this hypothesis *predicts* that when we present participants with an explicitly epistemic prompt, we should get *the same* results as in the main condition. However, this is not what Sarkissian et al. found. Instead, they found that participants drew a clear distinction between correctness and epistemic justification (Sarkissian et al., *ibid.*: 499-500). Therefore, these results provide reason for skepticism about the epistemic hypothesis.

4.2 Disagreement Without Univocality

⁶⁵ Admittedly, the "wrong" wording is ambiguous between a reading in terms of the normativity of truth, and moral normativity (ordinary speakers use "wrong" in both ways). However, it seems unlikely that participants understood "wrong" in a moral sense, here. Because if they did, then they would, effectively, be making the (first order) claim that the extra-terrestrials' judgements are not *morally* wrong, and this seems unlikely.

Turn, then, to the content relativist alternative. As we've seen, the difficulty for contextualism is that it implies that moral disagreements involve beliefs with straightforwardly compatible contents. In *Hyperspace Bypass*, for example, a content relativist might hold that the content of Jeltz' belief that taking over the earth is fine is the proposition that the norms which regulate *Vogon* society permit taking over the earth, and that the content of Dent's belief that taking over the earth is wrong is the proposition that the norms which regulate *human* society forbid taking over the earth. The obvious problem is that if *this* is all the two sides believe, then it is hard to see how they disagree, given that these propositions are straightforwardly compatible.⁶⁶

This problem, furthermore, should show up in the experimental data. We would not expect ordinary speakers to attribute genuine disagreement in cases of superficial sentential negation, for which a standard content relativist semantics is correct. For example, we would not expect ordinary speakers to attribute genuine disagreement if the vignette involved Jeltz saying "I am a Vogon" and Dent saying "I am not a Vogon". Why then, do they attribute disagreement in the experimental vignettes, if the semantic story is broadly the same as in the "I am not a Vogon" example?

As we've already seen, there is a standard form of response to this objection, namely to develop an account of disagreement which does not require incompatibility at the level of first-order content. The most straightforward (and in my view most plausible) way to do this is to appeal to disagreement in attitude, the now familiar notion of disagreements constituted by clashing non-cognitive attitudes – which can obtain in the absence of clashing beliefs.⁶⁷ For

⁶⁶ Different versions of contextualism say different things about the contents of moral beliefs. On many views, the story will be more sophisticated than the toy example in the text. For example, a more promising view would be that the content of Jeltz belief is that demolishing the earth has a certain natural property, where the relevant property is determined by his standards. However, the key point is that on all of these views, we can imagine cases where, intuitively, people disagree, but where contextualism entails that they believe contents which are straightforwardly compatible.

⁶⁷ Another option is to appeal to so-called "metalinguistic disagreement" (see e.g. Plunkett & Sundell, 2013). Eliding much detail, the idea is essentially this. Since contextualism entails that the contents of the relevant first-order moral beliefs are determined by different standards, it follows that there is no disagreement at the level of first-order content. However, the suggestion is that when we put forward moral claims, we implicitly advocate for the uptake of our preferred standards by other parties to the

example, inasmuch as Jeltz wants to demolish the earth, and wants Dent to let him do so, but Dent wants Jeltz to desist, this alone is sufficient to ground a sense in which they disagree, according to many, even though there is no mention here of any *proposition* that Jeltz and Dent disagree about.

Once we recognise that content relativists can offer these kinds of accounts, the dialectic between the content relativist and the truth relativist takes on a more expansive form. What the disagreement in attitude account shows is that contextualism is compatible with the view that there is a kind of genuine disagreement present in these cases. Therefore, it shows that contextualism is compatible with the experimental data. What we now want to think about, therefore, is whether contextualism best accounts for *all the evidence* vis a vis disagreement, (including our philosophical intuitions). Fortunately, these intuitions converge to a greater degree than intuitions about faultlessness (where, recall, our original motivation for going experimental was a lack of convergence). Philosophers tend to agree that cases like Hyperspace Bypass involve genuine disagreement, and that this is something that a moral semantics should explain. But while the appeal to disagreement in attitude shows that content relativists can provide *an* explanation, it's doubtful, as I'll now argue, that their account is overall as attractive as the account that the truth relativist can offer.

The reason for this is that the core semantic commitments of contextualism generate some pretty wild implications, pertaining to disagreement in belief/assertion, which the bolting on of disagreement in attitude does nothing to block. This is because, quite generally, the notion of semantic (and psychological) content is implicated in our ordinary thinking about disagreement, so that an account of disagreement which is *independent* of this notion (as the disagreement in attitude account is, given contextualism) is likely to be unstable.

conversation. Thus, there is a higher order proposition to disagree about, namely the proposition that all parties should accept such and such standards. However, while this idea may account for disagreement simpliciter, it is not clear that it can account for faultless disagreement. After all, there is an obvious question about the semantics of this higher order disagreement. If we accept a content relativist semantics, then the present problem simply recurs at the metalinguistic level (claims about what standards to accept will turn out to be compatible, so that there is no genuine disagreement). But if we accept an invariantist or expressivist semantics at this level, this will entail that the higher-order disagreement is not faultless.

This is exactly what we find in the case of the combination of contextualism and disagreement in attitude. For consider the following. Suppose we consider the Vogons' moral beliefs. Inasmuch as we are competent speakers of English, and contextualism is true, we will know implicitly that the contents of the Vogons' beliefs are straightforwardly true. And thus, when we talk to each other about *what they believe*, it should be clear to us that what they believe is straightforwardly true, just as it should be clear to us that what they say when they say "we are aliens" is straightforwardly true (even though it is not something we could correctly say ourselves). Indeed, it should even be clear to us that *everything they believe* about morality is such that we believe it too. For everything they believe about morality is of the form "according to Vagon standards, such and such"⁶⁸.

Now, since the content relativist account of disagreement floats free of these considerations about content, it leaves us free to say that we disagree with the Vogons. But this combination now looks unstable. On the one hand, we should say "everything the Vogons believe about morality is true, and everything that they believe about morality is such that we believe it too", but in the same breath we should say "of course, deep moral disagreements exist between them and us". This is bizarre.⁶⁹

Notice that none of the other views we have considered face this problem, even if they adopt the same account of disagreement in attitude. And fundamentally, this is because none of the other views allow questions about disagreement to come apart so dramatically from questions about semantic content. In the case of expressivism, this is because semantic contents *just are* non-cognitive attitudes. Therefore, when we talk about what the Vogons believe, we just are talking about the noncognitive attitudes with which we disagree (and we can call them false in a deflationary sense). In the case of invariantism and truth relativism, on the other hand, this is because our disagreements in attitude are parasitic on disagreements in belief, about

⁶⁸ Content relativists may see this as an unfair caricature, for few sophisticated content relativists literally hold that moral contents have the form "according to my community's standards, such and such". However, the key point is that content relativists *do* hold that the Vogons' beliefs will have contents which are straightforwardly true, and that this is something that ordinary speakers can implicitly know.

⁶⁹ (Ridge, 2014) makes this point.

shared contents, and therefore, on these views, people cannot rationally believe what the Vogons believe, nor think that what the Vogons believe is true.

This may seem less obvious if truth relativism is true, than if invariantism is true. Granted, truth relativism rules out our believing what the Vogons believe. However, one might wonder: isn't truth relativism compatible with there being a sense in which we should accept that what the Vogons believe is true? After all, truth relativism introduces a notion of truth at a context of assessment, as fundamental to the evaluation of semantic contents. So if truth relativism is true, should not ordinary speakers say that what the Vogons believe is true relative to their contexts of assessment?

But truth relativism does not have this implication. For, to repeat points made in the introduction, it's crucial – if we want to understand the truth relativist framework – that we distinguish the truth predicate that features in the metalanguage of the semantic theory from the ordinary truth predicate. Ordinary speakers of the object language (i.e. English) do not normally speak in terms of the semantic theory. Therefore, we should not expect them to use the technical semantic truth predicate: 'true at n context of assessment', because this is not a well-formed expression of English. Instead, we should expect them to evaluate contents with the ordinary truth predicate. And, recall from the introduction, since the ordinary truth predicate can be applied to propositions which (according to truth relativism) have assessment sensitive contents, it must itself get an assessment sensitive semantics. Thus, its extension is determined relative to a context of assessment, so that a speaker can correctly call a proposition true (in the ordinary sense) only if this proposition is in the extension of the ordinary truth predicate at the speaker's context of assessment. Of course, what the Vogons believe is not true at the humans' context of assessment. And therefore, humans cannot correctly say that what the Vogons believe is true, in the ordinary sense. Instead, they should say what the Vogons believe is false in the ordinary sense.

Notice that no parallel move is available to the content relativist. For according to content relativism, what the Vogons believe is an ordinary proposition, which is true or false in just the same way that claims about matters of prosaic fact are true or false. And this makes it very

hard to see how the content relativist could give a semantics for the ordinary truth predicate which entails that what the Vogons believe is not in the extension of the ordinary truth predicate at the humans' context of use. This, after all, would seem to require that the *non-moral* sentence for which the Vogons' moral utterances are elliptical is *also* not in the extension of the ordinary truth predicate at the humans' context of use (in other words, that humans could *not* correctly say "it is true that invading the planet is permissible according to Vagon morality").

The most detailed reply to these worries that I am aware of is developed in Steven Finlay's (2014). Finlay defends a sophisticated form of contextualism which has two main features. First, it offers a unified content relativist analysis of the semantics of normative language, including its moral uses. Second, it develops a sophisticated pragmatics, to account for the philosophically puzzling aspects of moral uses of this language, without departing from the unified semantics. Finlay argues that in light of his account of moral pragmatics, we should not expect people to utter (or find felicitous) the kinds of bizarre sentences mentioned above. However, he concedes that these sentences should be strictly *coherent*, at least in some contexts, if his view is true.

Finlay is willing to chalk this down to a degree of semantic blindness on the part of ordinary speakers (for ordinary speakers would not find these sentences coherent, in any context). But he thinks this is an acceptable price to pay, given the overall strengths of his account (particularly, its unifying semantics), and given problems that beset other theories (including truth relativism).

To decide whether this is right, we would have to assess Finlay's account in detail. This assessment will have to wait until chapter 4. But what it is fair to say at this stage, and it is not clear that Finlay would disagree with this, is that truth relativism does seem to hold an explanatory edge⁷⁰ here. For to repeat a point made earlier, the data that we ultimately aim to

⁷⁰ At least with respect to the narrow issue of faultless disagreement. Finlay's skepticism about relativism appears to derive from broader theoretical considerations concerning representation and communication. I say some more about such considerations in chapter 6.

explain with our semantic theories are ordinary speakers' intuitions about correct use of sentences. And thus appeals to semantic blindness are *ceteris paribus* nonideal, even if (as Finlay insists) the best theory, in the final analysis, may have to make some such appeals somewhere. Therefore, since truth relativism explains the case without any such appeals, we should conclude that truth relativism outperforms content relativism, at least on this particular issue. Of course, that does not imply that truth relativism will win out in the final analysis, but it does add positive weight, in favour of truth relativism, to the overall calculation. Thus, we can now conclude, a truth-relativistic explanation of the experimental data is at least somewhat superior to *both* realism/expressivism *and* content relativism. (Again, the details of Finlay's theory will be discussed in chapter 4, where I will argue that truth relativism clearly provides a superior explanation.)

4.3 Clarifying the Commitments of Truth Relativism

To be maximally clear, it should be noted that truth and content relativism diverge in their predictions *only* in respect of data about *alethic evaluations* of the *contents* of beliefs and assertions. When it comes to normative evaluations of beliefs and assertions themselves (i.e. qua mental states and speech acts), content and truth relativists *make all and exactly the same predictions*. For example, truth relativism predicts that competent speakers will judge that *what the Vogons believe* is false, even though the Vogons' beliefs are correct (and/or that Arthur can correctly say this). Content relativism, in contrast, predicts that competent speakers will judge not only that the Vogons' beliefs are correct, but also that competent speakers will judge that what the Vogons believe is true.

This means that the experimental data alone are not sufficient to draw an inference favourable to truth relativism, for the data do not discriminate between the relevant alethic and normative judgements. Rather, the argument for truth relativism, and against content relativism, turns at this point in the dialectic on an appeal to intuition. Specifically, it appeals to the claim that statements of the form "everything the Vogons believe about morality is true, and everything that they believe about morality is such that we believe it too... nevertheless, deep moral disagreements exist between them and us" are not, intuitively, correct.

5. Conclusion

In this chapter, we've finally seen how an argument from disagreement can provide evidence for moral relativism. In particular, we've seen that putting an experimental spin on Max Kolbel's argument from faultless disagreement allows us to see that there is reason to accept a truth relativist moral semantics over rival views.

In the final section of this chapter, we briefly discussed the dialectic between truth relativism and content relativism, vis a vis the problem of disagreement. Our discussion focused on what we might loosely call the negative side of the problem. That is, we focused on *objections* to the content relativist account, while arguing that truth relativism avoids these objections. On the positive side, we assumed that both truth relativists and content relativists can explain disagreement by appeal to disagreement in attitude. However, we saw that truth relativism is less vulnerable to objections that arise from the resulting combination of semantic descriptivism with disagreement in attitude.

What we have not considered in detail, however, is whether truth relativists are able not only to endorse moral disagreement in attitude, but also to provide an account of moral disagreement in belief. It might seem obvious that truth relativists can do so, for they hold that contradictory moral beliefs have contents which are each others' logical negations. In fact, though, this is less clear than it initially seems. Although truth relativism entails that there is a unique content which the Vogons reject and humans accept, recent research has shown that this is *not* sufficient for genuine disagreement in belief. It thus remains an open question whether truth relativism can provide a positive account of disagreement in belief.

To be clear, the openness of this question does not undermine the argument of this chapter. For even if truth relativism does not provide a positive account of disagreement in belief, we have already seen enough to know that it provides a more attractive account of (faultless) disagreement *simpliciter* than contextualism does (for it provides a positive account in terms

of disagreement in attitude, and this account does not clash with its core semantic commitments). However, it is an interesting question (and one that has so far received only limited attention) whether truth relativism can go further, and provide a positive account of disagreement in belief as well. It is this question which will concern us in chapter 3.

CHAPTER 3

ABSTRACT: This chapter examines the question of whether truth relativists can account for disagreement in belief. Relative to the previous chapter, it initially focuses more closely on the domain of aesthetics, before turning to ethics in its later sections. MacFarlane's recent truth relativist proposal is discussed, as is Dreier's objection to this proposal. The chapter argues that MacFarlane's main account fails. However, it develops and defends an alternative suggestion of MacFarlane's. The chapter's thesis is that the resulting account provides a positive account of disagreement which essentially involves the truth relativist's distinctive semantic notions.

In the previous chapter, we discussed the argument from faultless disagreement. A key claim in the argument was that truth relativism provides a better account of disagreement than does content relativism. And, this claim turned partially on recognising that truth relativists, just like content relativists, can appeal to the phenomenon of disagreement in attitude (while avoiding objections that befall content relativists attempts to appeal to this phenomenon).

However, at the end of the previous chapter, we noted that it would be surprising if truth relativists could *only* appeal to disagreement in attitude, and could offer no account of disagreement *in belief/assertion*. After all, truth relativists hold that moral judgements are, in a non-minimal sense, beliefs. And they hold that when two people make contradictory moral judgements, these judgements are incompatible beliefs about a particular proposition (likewise, *mutatis mutandis*, for moral assertions). One might reasonably expect, therefore, that truth relativists would hold that when two people make contradictory moral judgements, they will disagree in belief, and not just in non-cognitive attitude (again, *mutatis mutandis* for contradictory assertions). The goal of this chapter is to argue that the truth relativist *can* defend such a claim.

Unlike the previous chapter, this chapter discusses disagreements about *aesthetics* in considerable detail. Indeed, it spends *more* time doing this than it does discussing moral disagreement. This is because, as noted in the previous chapter, it is helpful to understand how truth relativism works in the “easy cases”, such as aesthetics, before one turns to hard cases such as ethics. However, in the previous chapter, our topic was one where the application of truth relativism to aesthetics had already been worked out in detail. And, we were therefore able to help ourselves to a framework which we could then focus on applying to ethics. In *this* chapter, however, our topic (namely, disagreement in belief/assertion for truth relativists) has been less extensively discussed. So, there is a fair amount of new ground to be broken, even in the “easy case”. Not only is it intrinsically interesting to break this new ground, it is also an important preliminary step to an effective discussion of the ethical case.

Here is the plan of how the rest of the chapter will go. Sections 1 – 4 discuss disagreement in aesthetics. Section 1 sets up the discussion by explaining why it is non-obvious that truth relativism is compatible with disagreement in belief/assertion, and why there is a challenge for the truth relativist to meet if he or she wants an account of disagreement in belief/assertion. Section 2 discusses MacFarlane’s (2014) attempt to meet this challenge. MacFarlane’s account, I argue, fails. But the failures are instructive. In section 3, I build on lessons learned from the failure of MacFarlane’s account to develop an alternative account (although this account is not fully original to this thesis, but rather develops a suggestion of MacFarlane’s). Section 3 also rebuts objections to this account. Section 4 applies the account to ethics.

1. Background

Begin with an example. The film *Only God Forgives* is one of those rare works of art which sharply divides critics. Some critics think it’s determinately very good, others think it’s determinately very bad. Few seem to be unconvinced either way. The Guardian’s Peter Bradshaw, for example, calls the film “gripping” and praises its “lethal formal brilliance”. Bradshaw rates the film “five stars”. Time Out’s Dave Calhoun, on the other hand, criticises the film’s “emptiness and sheer silliness”. Calhoun rates it “one star”. Bradshaw and Calhoun, we may say, disagree about whether *Only God Forgives* is a good film:

Critics: Bradshaw and Calhoun disagree over whether *Only God Forgives* is a good film. Bradshaw thinks it is a good film. Calhoun thinks it is not a good film. Both are expert film critics, and are properly acquainted with the film itself.

To briefly rehearse points made in the previous chapter, many people have taken this kind of case to provide evidence for a truth relativist semantics for aesthetic evaluations.⁷¹ There are basically two reasons for this. On the one hand, the disagreement seems faultless in the sense that both parties' beliefs are correct. On the other hand, it seems that Bradshaw and Calhoun genuinely disagree. Traditional theories have trouble explaining this. Invariantists have trouble explaining the appearance that both parties are correct (as do sophisticated expressivists). Content relativists have trouble explaining the sense that Bradshaw and Calhoun disagree.

If we zoom in on the second point, and again to repeat somewhat, the difficulty for the content relativist is that content relativism entails that Bradshaw and Calhoun believe compatible contents: Bradshaw believes a content which is determined relative to his context of utterance, whereas Calhoun believes a (different) content which is determined relative to *his* context of utterance. On a toy view: Bradshaw believes that Bradshaw likes the film and Calhoun believes that Calhoun likes the film.⁷² This makes it hard to see how Bradshaw and Calhoun can disagree in *belief*, if content relativism is true (although they may disagree in attitude).

As we saw in the previous chapter (again, a brief recap) truth relativism avoids this difficulty because it entails that there is a unique content that Bradshaw accepts and Calhoun rejects, namely the content that *Only God Forgives* (OGF) is a good film. According to the truth relativist, this content gets evaluated for truth as used and assessed at a pair of contexts (essentially, as used at a world and assessed at an aesthetic standard).⁷³ Thus, Bradshaw's belief is correct just in case its content is true as used at his world and assessed at his standard,

⁷¹ See note 1 chapter 2.

⁷² Actual proponents of content relativism will take a more sophisticated view of belief content. For example, they could hold that the content of Bradshaw's belief is the claim that watching the film promotes some goal of Bradshaw's. The key point, though, is that even on this view, Bradshaw and Calhoun's claims will be straightforwardly compatible.

⁷³ See the introduction for presentation of the truth relativist framework.

likewise with Calhoun's belief. This allows Bradshaw and Calhoun to both be correct, even though there is a unique content that one accepts and the other rejects.

Now, it might therefore seem that truth relativism straightforwardly secures an explanation of the intuition that Calhoun and Bradshaw disagree in belief/assertion. For it entails that Bradshaw accepts a content that Calhoun denies, and vice versa. However, matters are not so simple. Another case will help us to see why.

Cross-worlds: Alba thinks that Moscow is the capital of Russia. Anna thinks that St Petersburg is the capital of Russia. Alba inhabits the actual world. Anna inhabits a counterfactual world where St Petersburg is the capital of Russia.⁷⁴

Intuitively, Anna and Alba do not genuinely disagree. However, given standard semantic assumptions, Anna and Alba hold contradictory views about a unique content. The cross-worlds case therefore shows that shared content is not, by itself, sufficient for genuine disagreement. Therefore, it undermines our naïve thought that truth relativism successfully accounts for disagreement *merely* in virtue of entailing univocality.

The point is not simply that Cross-worlds is a counterexample to the view that univocality is sufficient for disagreement. There is a deeper worry, which is this. Possible states of the world play a role in standard truth-conditional semantics which is not dissimilar to the role played by aesthetic standards in the truth relativist semantics. Standards in the truth relativist semantics, like worlds in the standard semantics, are the things against which contents get assigned truth values. And so, to put the worry pithily: if truth relativism is true, why are tastes not just like worlds? Granted, Anna and Alba are not talking *about* different things (Anna is not talking *about* w1 and Alba is not talking *about* W2).⁷⁵ Nevertheless, their respective

⁷⁴ This kind of case is discussed in MacFarlane (2007).

⁷⁵ People sometimes want to hear more about the following point. Why can we not simply say that Alba believes that *at the actual world* Moscow is the capital of Russia? Then, we could explain Anna and Alba's non-disagreement in terms of their believing compatible contents. However, this would imply, implausibly, that ordinary empirical claims (such as that Moscow is the capital of Russia) express necessary truths. This is because the proposition [in the actual world, Moscow is the capital of Russia] is true in all possible worlds.

beliefs are *made true* by different things. And this is presumably part of the reason why it does not seem to us that they disagree. But then, why does the analogical line of reasoning not run for Bradshaw and Calhoun, if truth relativism is true? Granted, their beliefs are not *about* different things. Nevertheless, their beliefs are *made true* by different things. So why, if we are truth relativists, should we conclude that they disagree?

So this is the problem that the first three sections of this chapter aim to solve. Why think that Bradshaw and Calhoun disagree, even though Anna and Alba do not?

Before we get into the details of the problem, three brief points of clarification. First, it's important to get clear on how we understand the notion of an aesthetic standard. Following MacFarlane (2014; p143) I understand an aesthetic standard as something psychological, but non-intellectual. On this view, a standard is not a set of abstract principles about what makes for a good movie, but rather a set of dispositions to like or dislike certain movies (which we can think of as a kind of psychological function from movies to liking-states). In other words, we can think of a standard as a sensibility in the sense of (Blackburn, 1984: 192).

Second, I want to insist on a degree of idealisation surrounding the notion of an aesthetic standard. In particular, I want to say that the standard that gets assigned at a context of assessment is not the assessor's actual sensibility, but rather the sensibility they *would* have under ideal conditions. (I understand being in ideal conditions as, at least, being properly informed about the non-aesthetic aspects of the relevant work, being free from parochial bias or narrow-mindedness, and being in an otherwise psychologically normal state (i.e. not being drunk or overly tired, etc).)

This commitment to idealisation is independently motivated. To see this, imagine someone who thinks OTG is bad just because they have a personal hatred for the director⁷⁶. Even if we are attracted to truth relativism, we would not want to say that this person's belief is correct. Of course, there is room for debate about how much to pack into the idealisation. But examples like this show that we are going to want to idealise to at least some degree. It is not

⁷⁶ Hume appeals to an example like this, to make a similar point (Of the Standard of Taste, 227).

so important, for now, how strong the idealisation is (although we will come back to this point).

The final point of clarification builds on the first two. The point is that it's crucial that we understand the Critics case in such a way that Bradshaw and Calhoun's aesthetic standards are taken to diverge, even after idealisation. This is because if Bradshaw and Calhoun's standards would *converge* after idealisation, then (given the points just made) their beliefs will get evaluated relative to *the same context of assessment*. But if their beliefs get evaluated relative to the same context of assessment, then there will be no special difficulty in explaining why they disagree. For in that case, they will disagree in exactly the same sense as people ordinarily disagree about objective matters: the sense in which at least one of their beliefs must be incorrect. The interesting case for our purposes, then, is rather one where Bradshaw and Calhoun's standards diverge even after idealisation. It is only in that kind of case where there is any interesting problem for the truth relativist. Now since we will come back to this point, let me introduce a label for this kind of disagreement. Let us say that an aesthetic disagreement is *fundamental* just in case the parties' idealised aesthetic standards would output incompatible evaluations of the object of the aesthetic disagreement. Thus, a disagreement which counterfactually depends on a relevant shortcoming is non-fundamental in this sense.

2. MacFarlane's view

MacFarlane (2014) discusses our problem at some length. MacFarlane's solution is unsuccessful, or so I will argue. However, it is instructive to see why this is so. Understanding the difficulties that face MacFarlane's view will help us understand what is attractive about the alternative solution which we will go on to develop. Again, to be clear, the alternative is a development of a suggestion that MacFarlane himself offers. So the basic idea is not original to this thesis. The contribution that this discussion makes is rather to show that this suggestion of MacFarlane's can be developed into a workable account.

MacFarlane begins by arguing that cases like Critics instantiate disagreement in attitude.⁷⁷ Of course, this does not by itself get us any closer to thinking that they instantiate disagreement *in belief*. And furthermore, as we noted in the previous chapter, it does not help us understand how truth relativists can endorse a form of disagreement which is stronger than that which content relativists and expressivists can endorse. However, MacFarlane thinks that the fact that disagreement in attitude is instantiated in the case gives us the materials to develop an account of a stronger form of disagreement, which is, as it were, proprietary to truth relativism.⁷⁸

MacFarlane thinks that in cases of disagreement in attitude, we normally try (in non-cognitive terms) to change each other's sensibilities. That is to say, we normally try and get our interlocutor to abandon their non-cognitive attitude, and come to share ours, thereby resolving the disagreement in attitude. For example, we can imagine Bradshaw trying to get Calhoun to share his liking for *Only God Forgives*, thereby changing his sensibility, and resolving the disagreement in attitude.

MacFarlane's theory offers a special way of capturing the sense in which this phenomenon (of trying to change each other's sensibilities) adds some additional, proprietary, robustness to an account of the disagreement present in the case. Central to MacFarlane's idea is the speech act of retraction. This is the speech act one performs when one says things like 'I take that back', or 'I retract that'.⁷⁹ Retraction can target many kinds of speech acts, such as

⁷⁷ MacFarlane has in mind disagreements which concern matters of taste, such as over whether a particular food is tasty. But nothing important, as far as I can tell, hangs on the differences between aesthetic and taste discourse, at least not for our purposes.

⁷⁸ It is somewhat unclear whether MacFarlane would view this as a form of disagreement in belief/assertion, as opposed to a stronger form of disagreement in attitude (and see footnote 2). Nothing hangs on this.

⁷⁹ For a nice example of retraction in the wild, see here: <https://youtu.be/kPPyBLC5ddM?t=170>. It is important for MacFarlane's view that retraction is understood as something over and above mere assertion of the negation of the content of the assertion targeted. But this understanding is plausible. For one can permissibly retract an earlier assertion even if one cannot permissibly assert the negation of the earlier assertion. For example, suppose one becomes less confident in *p*, so that one does not want others to rely on one's earlier assertion. One would then be permitted to retract an earlier assertion that *p*. However, one may still be sufficiently confident in *p* that one cannot permissibly assert that not-*p*. Again, different views of the fundamental nature of speech acts will offer different explanations of

questions, offers, and commands. But the most interesting case, for us, is where it targets an assertion. Plausibly, this use is governed by the norm that one is obliged to retract earlier assertions whose contents are false as used at one's earlier context and assessed at one's current context. But what has this got to do with disagreement?

On MacFarlane's view, the content of an earlier aesthetic assertion may be false as used at the earlier context (of assertion) and assessed at the later context (of retraction). Thus, if I used to hate broccoli, but now I love it, I am obliged to retract my earlier assertion of 'broccoli is delicious' because the content of this assertion is false as used at the earlier context and assessed at the later context.⁸⁰ Now, though, recall the point about disagreement in attitude – in particular, about how we normally try to change each other's (non-cognitive) minds. Suppose Bradshaw successfully gets Calhoun to change his tastes, so that he comes to like OTG. Calhoun will now be in a position where he is obliged to *retract* his earlier assertion, for his aesthetic standards have changed, so that the content of his earlier assertion is now false as used at the earlier context and assessed from his current context. According to MacFarlane, there is a sense in which, if one is trying to get one's interlocutor into a position where they are obliged to retract their previous assertion, it makes sense to think of oneself as *attempting to refute* one's interlocutor. And thus there is a sense in which it makes sense for Bradshaw to see himself as attempting to refute Calhoun's earlier assertion. Attempting to refute, though, is characteristic of genuine disagreement. Thus, it seems, MacFarlane can elucidate a sense in

why this makes sense. But we can see that it plausibly does make sense, independently of accepting any particular deep explanation of the practice.

⁸⁰ Here is a brief recap of the assessment sensitivity machinery. Start with standard (Kaplanian) content relativism. On this view, sentence meaning is doubly relativised to a context of use and circumstance of evaluation. On this view, circumstances of evaluation *cannot* contain taste-parameters, and are initialised by the context of use. Now consider conservative (Kolbelian) relativism. On this view, the same double-indexing machinery is in place, but the circumstance of evaluation *can* contain taste parameters. Crucially, the circumstance of evaluation is still initialised by the context of use. Finally on ambitious (MacFarlanian) relativism the circumstance of evaluation is determined by a *context of assessment*, which can contain a taste parameter, and which (crucially) need not be initialised by the context of use. Therefore, it can make sense to evaluate (for truth) what is said in one context *from* a perspective (and taste) determined by some *other* context. This cannot be done in the standard content relativist and conservative truth relativist frameworks. See MacFarlane (2014: chs.3-4), and the introduction to the thesis.

which Calhoun and Bradshaw genuinely disagree: they are attempting to refute each other's assertions.

Notice that this account is not available to the content relativist. For since Calhoun has different tastes at t_1 and t_2 , he asserts different contents at t_1 and t_2 (and these contents are true or false absolutely). Thus, there is no sense in which the content of his earlier assertion can be false at t_2 , for the content of his earlier assertion is something which is true absolutely (only relative to a world). But since his earlier assertion has a content which is true absolutely, it is hard to see how it could be obligatory for Calhoun to retract this assertion. And without an obligation to retract, the story about attempting to refute collapses.

2.1 Initial problems for MacFarlane's view

Now that we are clear on how MacFarlane attempts to solve the problem, we can turn to examine the reasons why it fails. There are three difficulties.

First, it is not obvious that there is any genuine form of disagreement in attitude in Critics. Much depends on how we construe the non-cognitive attitudes that Bradshaw and Calhoun have. For not all pairs of opposing non-cognitive attitudes amount to genuine disagreements, and the line between disagreement and mere difference is somewhat fuzzy.

For those not familiar, the basic idea of the disagreement in attitude strategy is that there are at least some cases of disagreement which seem to involve clashing non-cognitive attitudes, rather than clashing beliefs (Stevenson, 1937) (Björnsson & McPherson, 2014) (Ridge, 2014). The classic case is one where two people are deliberating about what to do together, and they develop conflicting preferences which cannot both be satisfied. For example, suppose Bradshaw and Calhoun are discussing which restaurant to dine at after a press screening. Suppose they have agreed that they'll eat together, at the same restaurant. Suppose Bradshaw wants to eat Italian food and Calhoun wants to eat Japanese food (suppose there are no restaurants which serve both). Many find it plausible that these cases instantiate a genuine

disagreement (about which restaurant to eat at) even though there is no obvious disagreement in belief.⁸¹

Now, we *could* construe the Critics case in a structurally similar way. That is, we could imagine that Bradshaw and Calhoun have agreed to see a film together, and are discussing the question of whether OTG is a good film in the context of deliberating about which film to see. If we think about the case this way, it will be plausible that Bradshaw and Calhoun have a genuine disagreement (in attitude).

However, it's clear that we could equally well construe the case in a way which makes it less clear that there is a genuine disagreement (in attitude). Suppose Bradshaw and Calhoun have no plans to do anything together. Suppose instead, they are discussing OTG in the context of a TV film review show, where their only practical goal is to make the case for their view as to whether OTG is a good film. What non-cognitive attitudes is it plausible to attribute in this case? Bradshaw and Calhoun, plausibly, have opposite likings or preferences. One likes the film, the other dislikes it, one is disposed to enjoy the film, the other is disposed not to enjoy it. Furthermore, Bradshaw and Calhoun plausibly have some opposite practical dispositions. One is disposed to watch the film (under certain conditions) and another is disposed not to watch it. But the problem is that *these* oppositions of attitudes seem too thin to generate a sense of genuine disagreement. If I just like a certain thing, and you just dislike that thing, it's not clear that we would (yet) be in a position to say that we disagree about anything. Nevertheless, intuitively, there is a genuine disagreement (even) in the review-show version of the case. So there is at least a limit to the extent to which we can see disagreement in attitude in the cases we want to explain.

How serious a problem is this for MacFarlane's account? One might think that it is not such a big problem. For it is not clear that the distinctive part of MacFarlane's account *depends* on the existence of a disagreement in attitude. Strictly speaking, MacFarlane's strategy only requires that Bradshaw and Calhoun attempt to change each other's non-cognitive attitudes. And it is conceivable that they would do this even if the attitudes did not already constitute a

⁸¹ The dining case is a classic example (Stevenson, 1937) (Ridge, 2014).

disagreement. Indeed, Stevenson's classic discussion of disagreement in attitude suggests that the line between disagreement and mere difference is crossed *only* when Bradshaw and Calhoun are disposed to attempt to change each other's minds.⁸² So if they do attempt to do this, maybe this is OK for MacFarlane.

This brings us to the second problem with MacFarlane's account however, namely that it is not clear that people normally do try to change each other's minds, at least in the relevant kind of case. To see this, we need to recall the distinction between fundamental and non-fundamental disagreements. Non-fundamental disagreements are disagreements where there is sufficient overlap between the parties' sensibilities (after idealisation) for the disagreement to be understood as turning on some non-aesthetic mistake (for example, lack of familiarity with the film, lack of knowledge of the genre, having viewed the film under the wrong conditions, etc.). Fundamental disagreements, on the other hand, are disagreements where there is insufficient overlap between the parties' idealised sensibilities for the disagreement to be understood in this way, and where the disagreement is therefore likely to persist in ideal conditions.⁸³

Why is this relevant? It is relevant because we respond differently in cases of fundamental and non-fundamental disagreement about matters of taste. In non-fundamental cases, we may well attempt to change each other's minds. For example, we may draw each other's attention to certain aspects of the film which we take to be aesthetically pleasing, or we may attempt to inform each other about the conventions of the genre, or about relevant historical facts, or about relevant biographical facts about the director, and so on. We may reasonably expect that doing this will change our interlocutors' mind. But what about fundamental cases? In fundamental cases, it is hard to see how these tactics can work. For in fundamental cases, by

⁸² (Ridge, 2014).

⁸³ Of course, some domains will have more scope for non-fundamental disagreement than others. In film criticism, there is quite a lot of scope. But in the appreciation of natural beauty, say, there is probably less. The argument in the text does not turn, however, on the assumption that there is a lot of scope for fundamental disagreement in the relevant domain. This is because the key point is that *neither* case is happy for MacFarlane's account. In *fundamental* cases we do not normally try to change each other's sensibilities, but in *non-fundamental* cases there is no need for any non-standard account of disagreement.

hypothesis, the disagreement would persist even if our interlocutor had all the same non-aesthetic knowledge as we have – for their sensibility is just constructed in such a way that they find *different* things aesthetically pleasing and displeasing. Thus in fundamental cases, it is plausible to think that we would *not* normally continue to attempt to change each other's minds. Indeed, it is plausible to think that this is part of what is expressed by the aphorism "there's no disputing about taste". Early proponents of truth-relativistic treatments of these kinds of cases emphasize this point. For these theorists, the fact that there seems something odd about continuing to dispute in cases like this is part of the core motivation for a truth-relativistic treatment. Intuitively, the stance which it is natural to take in cases of fundamental disagreement about taste is one of "agreeing to disagree", rather than one of continued attempt to change the others' mind.⁸⁴

Even if we set these worries aside, however, there is a third problem, posed by Dreier (2009), which is also more interesting because it is more general. This problem is more general in the sense that, although it targets MacFarlane's account, it is a problem which *any* truth relativist account will have to solve.

2.2 The main problem for MacFarlane's view: Dreier's objection to speech act-based accounts

As we have seen, MacFarlane's strategy revolves around an appeal to a kind of speech act, namely retraction (and its role in attempts to refute). In earlier work, MacFarlane offers a Brandom-inspired account in terms of a different speech act, namely challenge (and response to challenge). Additionally, Schafer (2014), a fellow traveller, offers an account in terms of the success conditions of the speech act of assertion. What all three of these accounts have in common is that they identify speech acts which are characteristic of genuine disagreement (for example, if it is appropriate for me to *challenge* an assertion you have made – then this suggests that we have a genuine disagreement), and then attempt to give truth-relativistic accounts of these speech acts. If truth relativists can explain why these speech acts are appropriate, the thought goes, then they can explain why there is genuine disagreement,

⁸⁴ Kolbel (2004a), Lasersohn (2005) and Egan (2014) all make this point.

consistent with truth relativism. This is the basic strategy that all of these accounts have in common.

The worry I want to focus on, which is attributable to Dreier (2009), targets all accounts of this form. And the worry is this. It is not enough, Dreier worries, to simply stipulate a truth-relativistic account of the relevant kinds of speech acts. More than this, we have to understand how the introduction of truth-relativistic contents lets us hang on to the *point* of the practice – how the use of these speech acts continues to make sense once we understand them as having assessment sensitive contents. Dreier doubts that MacFarlane's account allows us to understand this.

Without rehearsing the details of the dialectic between Dreier and earlier truth relativist views that he had in mind, let me offer what I take to be a friendly update of Dreier's objection. In a Dreierian vein we can ask: "does the practice MacFarlane describes really capture what we ordinarily mean by 'attempting to refute'?" If it does not, then the link to genuine disagreement is lost. And it seems that it does not.

This is because, although MacFarlane is right that getting one's interlocutor into a position where they are obliged to retract their previous assertion *normally* counts as a kind of refutation, not just *any* way of doing this counts as such. It matters *how* the change in view gets effected.

For example, suppose you hate marmite. Thus, on MacFarlane's view, it will be correct for you to assert that marmite is disgusting. Now suppose I try to get you to love marmite, thereby making it obligatory for you to retract your earlier assertion. How could I go about this? One way would be to try to get you properly informed with respect to marmite. Perhaps you have never been served marmite *the right way*: maybe marmite is best on toast, and you've never had it on toast, or maybe you've only ever had off-brand marmite, and you need to try the real thing, or maybe marmite needs to be eaten hot, or with a dash of Worcester sauce, or whatever. This, I suspect, is what most people would think of when they think of how a debate over whether marmite is delicious could make sense.

But notice, this can't be what MacFarlane has in mind. Because if this kind of debate would resolve the issue, then the disagreement would be a non-fundamental one (for it would then be the case that both parties are disposed to like marmite under the right conditions, but one party is just not in those conditions). If we're thinking of a fundamental disagreement, then, we can't see the attempt to change each other's non-cognitive attitudes as proceeding in this way. But then, how can we see it as proceeding? Importantly, what we *can't* see is that there could be any way of providing *evidence* that one party's non-cognitive attitudes are incorrect, or *reasons* for any party to change their mind. The best that could be done, it seems, is to proceed by cajoling, or flattery, or ridicule, or threat of force, or some other purely psychological means ("you'd better like marmite if you want to be *one of us*"). But debating tactics that proceed by these kinds of methods, surely, cannot be seen as attempts to refute, at least not if we are clear headed. If I say to you "start liking marmite or I'll take out my gun and shoot you", and if this works, if fear leads you to genuinely start liking marmite, so that you become obliged to retract your earlier assertion that marmite is delicious, then by MacFarlane's lights I have effected a successful *refutation* (of your attitudes). This, though, borders on absurdity.

Examples like this multiply fairly easily. If a mad scientist zaps me with a sensibility-transforming ray, so that I come to hate marmite, and thus am required to retract my earlier assertion that marmite is delicious, has the scientist successfully *refuted* me? What if someone gives me a sensibility pill, or performs sensibility-altering brain surgery on me, etc? The key point is that not just any means of getting someone to be obliged to retract counts as an attempt to refute. Only means of the *right kind* count. But unfortunately for the MacFarlanian account, it seems that the only means that can make sense, in the relevant kind of (fundamental) case, are means of the wrong kind.⁸⁵

3. *The rejection account*

⁸⁵ Schafer's (2014) view also seems to me to be vulnerable to this objection, but I lack the space to discuss it in detail.

So this is why MacFarlane's account fails. Although getting one's opponent into a position where they are obliged to retract *normally* counts as an attempt to refute, MacFarlane fails to explain how we can continue to make sense of this once we introduce truth-relativistic contents into the equation.

Fortunately, MacFarlane provides us with a second option. Instead of appealing to the speech act of retraction, and its role in constructing a notion of attempting to refute, we can appeal to the speech act of *rejection*. MacFarlane chooses not to develop this account, for he thinks it is less clear that there is a distinct speech act of rejection than that there is a distinct speech act of retraction. But MacFarlane does not rule the rejection account out. And here, I want to develop this account. We will return to the worry about the existence of a distinct speech act below.

What is rejection? It is the speech act we perform when we say "I reject that" or "I deny that" (think of the politician who says "I categorically deny the allegations"). Linguistic denial (using the word "no") is also associated with rejection. We can think of rejection as an interpersonal form of retraction, inasmuch as it involves distancing oneself from someone else's assertion, rather than one of one's own previous assertions. Obviously, rejection is characteristic of genuine disagreement. If it makes sense for A to reject one of B's assertions, then this is strong evidence that A and B have a genuine disagreement. And natural norms for rejection, given truth relativism, entail that Bradshaw and Calhoun can reject each other's assertions. For, given truth relativism, it is natural to think that one may reject an assertion iff its content is false as used at the asserter's context and assessed at one's own context.

So the rejection account is essentially this claim: that, given truth relativism, it makes sense for Bradshaw and Calhoun to reject each other's assertions. And therefore, given truth relativism, Bradshaw and Calhoun genuinely disagree.

3.1 Advantages of the rejection account

Now this account has a number of advantages over the refutation account. First, it does not depend on the existence of a disagreement in attitude, for refutation is legitimised simply by the semantics of the relevant assertions. Second, and for the same reason, it does not depend on parties to a disagreement being disposed to change each other's attitudes. In this way, it sidesteps the first two worries mentioned above.

But more importantly, the rejection account avoids the main Dreierian worry, too. To see this, imagine trying to run the Dreierian objection against the rejection account. In this vein, we might ask: could it make sense for successful rejection to stand in perfect harmony with the equal success of the assertion it targets?⁸⁶ After all, truth relativism entails that *both* Bradshaw and Calhoun's assertions are correct, and that they can know this. But surely, it might be objected, the success of rejection has got to be tied to the *failure* of the assertion it targets. If our view of "rejection" entails that I can, perfectly felicitously, reject an assertion which I know to be correct, then should we not draw the conclusion that we are not longer talking about rejection, but rather something else entirely, something merely stipulative? Is our ordinary notion of rejection really compatible with this practice? Does this practice even make sense?

In *this* case, though, there is compelling evidence that this practice *does* make sense, because there is compelling evidence that this *is* the practice that we *already* have. For it is part of the intuitive data about cases of so-called faultless disagreement (cases like Bradshaw and Calhoun's) that we are happy to reject assertions about matters of taste, aesthetics, comedy, etiquette, and so on, *without* any assumption that the success of our acts of rejection require a corresponding defect in the assertions targeted. Indeed, this is part of what is special and distinctive about this kind of disagreement. It is one of the faultlines that marks off these kinds of disagreements (of which we say that there is "no disputing about taste") from the more robust disagreements which characterise other areas of thought and talk.

And notice another faultline. We have already touched on the point that we do not normally take ourselves to be attempting to refute each other when we engage in these kinds of

⁸⁶ Relevant here is Dreier's "throbby" case, which advances a similar argument against a truth relativist account of the speech acts of challenge and response (2009: 102).

disagreements (at least not in fundamental cases). And notice equally, that we do not normally engage in the practice of *challenge and response* in such cases (at least not in fundamental ones).⁸⁷ That is, while a central part of our ordinary practice of disagreement about, say, straightforward empirical matters is a disposition to *challenge* assertions with which we disagree, and *respond* to challenges to our assertions, this practice is conspicuously absent in cases like Bradshaw and Calhoun's (at least in fundamental cases). Indeed, it is plausibly part of the *data* that there would be something odd or misguided about challenging each other's assertions in such cases. After all, this, plausibly, is part of what we mean when we say that "there is no disputing about taste".

So what I want to say is this. Dreier is right that in ordinary cases of disagreement, it is hard to see how a truth relativistic account of the norms governing the speech acts associated with disagreement could make sense. For ordinary cases of disagreement involve the speech acts of challenge and response, as well as attempts to refute. And truth relativism has a hard time making sense of these things. However, the key point here is that cases like Bradshaw and Calhoun's are *not* ordinary cases of disagreement. They are an instance of a special, less robust, but nevertheless genuine kind of "faultless" disagreement, where the full range of speech acts which we would normally associate with disagreement are not used, and would be of dubious coherence, but where a more restricted range of disagreement-associated speech acts (at least, the speech act of rejection) are used, and *do seem to make sense*, even on a truth-relativistic account. This is why the rejection account avoids the Dreierian worry.

3.2 *Why the rejection account is unavailable to content relativists*

Notice that this is an account that *cannot* be appropriated by the content relativist. For (as we saw in chapter 2) if content relativism is true, the contents of Bradshaw and Calhoun's assertions are true absolutely, and they should be able to recognise this. It is thus hard to see how they could sensibly reject each other's assertions.

⁸⁷ See Dreier (ibid.) and (MacFarlane, 2007) for discussion.

In response to this claim, content relativists might point out that rejections can target propositions expressed at the level of *pragmatics*. For example, the rejection in the following dialogue is perfectly felicitous:

A: She's poor but honest.

B: I reject that, it's no surprise that a poor person is honest.

Here, B should not be interpreted as rejecting *what A says*. B may accept that, strictly speaking, the relevant individual is both poor and honest. Rather, B rejects something expressed at the level of the pragmatics of A's utterance, namely the implicature that there is any tension between the claim that it is surprising that a poor person is honest.

Content relativists could argue that something similar is going on in Calhoun and Bradshaw's case. Specifically, they could argue that what Bradshaw can reject is not the content of Calhoun's utterance, but rather something expressed at the level of pragmatics. However, this move would run into a problem of the following form. Notice that the following dialogue is perfectly felicitous:

A: She's poor but honest.

B: I reject that. Strictly speaking, what you say is true. She is both poor and honest. But it is no surprise that a poor person is honest.

But in Bradshaw and Calhoun's case, the parallel is infelicitous:

Bradshaw: OTG is a great film.

Calhoun: Strictly speaking, what you say is true. But I reject that, OTG is terrible.

The lesson to draw from this is that although rejecting an utterance *can* cohere with accepting the semantic content of the utterance. It does not seem that this is what is going on in Calhoun and Bradshaw's case. Rather, it seems that the kinds of rejection which Bradshaw and Calhoun will want to make are incompatible with accepting the semantic content of the targeted

assertion (it would be very odd for Bradshaw, at any point, to say that what Calhoun says is strictly speaking true). And therefore, it is hard to see how the content relativist can account for these rejections.

As an aside, notice that these considerations provide some initial insight on the phenomenon of faultless disagreement. For, again, part of the data that initially motivated interest in this phenomenon was the intuition that there is a kind of disagreement which seems to be less robust than disagreement about objective matters, but nevertheless seems to constitute disagreement worth the name. And, we can now say a bit more about what features of the practice our intuitions are picking up on when we have this intuition. For we are able to recognise that part of what makes these kinds of disagreements less robust is that they *do not* call for attempts to refute or challenge. But we are also able to recognise that part of what makes them seem to constitute disagreements, worth the name, is that they *do* call for rejection.

Interestingly, this account also appears to be unavailable to the conservative, as opposed to the ambitious, truth relativist. As we saw in the introduction to this thesis, conservative truth relativism is incompatible with intuitive uses of the speech act of *retraction*, because these uses require that the relevant (in this case, aesthetic) standard is determined by the context of retraction, while the relevant state of the world is determined by the context of the original assertion. Similar considerations apply to intuitive uses of rejection. For recall the vegetarianism example from the introduction. In that case, nothing turns on whether the vegetarian is rejecting another person's earlier pro-meat-eating assertion, rather than retracting their own. Thus, the rejection account provides reason to prefer an ambitious truth relativist view, rather than a conservative one. This view will henceforth be assumed.

3.2 Objections to the rejection account

The next step is to attempt to apply this account to the domain of ethical thought and talk. But before we do that, we need to discuss some possible objections to the aesthetic version of the account. First, the objection that rejection is not a distinct speech act. Second, the objection that the argument of the preceding section begs the question by appealing to a relativist-friendly

construal of faultless disagreement. Third, the objection that it is still unclear that the practice of truth-relativistic rejection makes sense. Fourth, the objection that the rejection account fails to explain disagreement in the sense of a state that two people can be in (as opposed to an activity that two people can engage in).

First objection: On the one hand, *retraction* has a normative profile that is distinct from mere assertion of the negation of the content of the retracted assertion. For it can be permissible to retract an assertion even if one is not prepared to assert the negation of the content of the assertion retracted (if, for example, one still believes the content, but is no longer sufficiently confident that one is happy for others to rely on one's testimony). However, the orthodox (Fregean) view of *rejection* is that to reject an assertion *just is* to assert the negation of the content of the assertion.⁸⁸ Thus, on this view, rejection has no distinct normative profile. In this respect, according to the orthodox view, rejection is importantly unlike retraction. And if this is right, then rejection does nothing to help us understand disagreement. For we have already seen that asserting the negation of another's assertion is insufficient for disagreement (as in the Cross-worlds case, Anna asserts the negation of what Alba asserts, but Anna and Alba do not genuinely disagree). If this is right, then the fact that Bradshaw can reject Calhoun's assertion does nothing to show that Bradshaw and Calhoun genuinely disagree.

Response: It is true that the orthodox view of rejection is unhelpful to the rejection account. However, discussions of the orthodox view have not envisaged cases like Cross-worlds. And these cases help us see that there *is* something more to rejection than mere assertion of negation. For consider the following.

Imagine Alba is in a philosophy seminar, and has just been convinced of modal realism. Now she is told "there is a counterpart of yours, Anna, in another possible world, where the Bolsheviks never took power, and the capital of Russia was never moved from St Petersburg to Moscow. In fact in this world, the capital of Russia is still St Petersburg. Your counterpart, Anna, has just asserted that St Petersburg is the capital of Russia, how do you respond?"

⁸⁸ C.f. (Ripley, 2011) (Incurvati & Schlöder, 2017).

Notice that the following responses all seem fine:

Alba: Ok well, Moscow is the capital of Russia.

Alba: Ok but, Moscow is the capital of Russia.

Alba: Nevertheless, Moscow is the capital of Russia.

And yet the following seems at least somewhat odd:

Alba: # No, Moscow is the capital of Russia.

Alba: # I reject that, Moscow is the capital of Russia.

Notice that in Critics, rejection seems fine:

Bradshaw: Only God Forgives is a masterpiece.

Calhoun: No, it's terrible.

Calhoun: I reject that, it's terrible.

Thus, it seems that rejection *does* differ from mere assertion of negation. Indeed, it seems that it is precisely rejection which marks the difference between cases of assertion negation which do not amount to genuine disagreement, and cases of assertion negation which do amount to genuine disagreement.

Second objection: The response to Dreier's objection appeals to claims (about cases of paradigmatically faultless disagreement) which are supposed to be independently plausible. In particular, it is claimed that it is independently plausible that Bradshaw can coherently reject Calhoun's assertion, even though he knows that Calhoun's assertion is correct. But this is how *relativists* like to think of these cases. Not everyone agrees. For example, some theorists deny that both assertions are correct, and appeal to a weaker notion of faultlessness to explain the intuition that neither has made a mistake.⁸⁹ The truth relativist cannot *assume* that these

⁸⁹ See e.g. (Schafer, 2011).

alternative explanations fail, but must rather *show* that they do, if they want to appeal to the claim that it is independently plausible that rejections can coherently target correct assertions.

Response: It is true that there are alternative ways of explaining the intuitive data. And if one of these alternative explanations is best, then we cannot appeal to cases like Critics in support of the rejection account. However, in the previous chapter, I argued that these alternative explanations are unattractive, at least in the moral case. And there are, if anything, stronger reasons to think that they are unattractive in the aesthetic case. For in the aesthetic case, the relevant kinds of disagreement are significantly more common, and this makes it harder to explain them by appeal to mere epistemic faultlessness, at least if one wants to hold on to a reasonable amount of aesthetic knowledge. Thus, even if there is some shifting of the bump in the rug here, the bump is getting shifted to an area where the truth relativist, I have argued, has the upper hand.

Third objection: Even if we accept this description of the data point, one wants to know more about why the practice makes sense: why we should take the appearances at face value. The rejection account invites us to notice that disagreements in cases like Critics seem to have a certain special structure: where rejection is warranted, but challenge is not. But even if we grant that this *seems* to be the case, one wants to know more about what *grounds* the difference we seem to be tracking. One wants to be reassured that the truth relativist can tell us a plausible story about the features of the underlying socio-linguistic practice which make sense of the data (and that they can do so in a way that doesn't collapse into some other, non-relativistic semantic view). If no such story is forthcoming (and if non-relativistic explanations of the data fail) then perhaps we should revise our practice, rather than accept truth relativism.

Response: To answer this question, we need to do two things. First, we need to say some more about the essential nature of the relevant speech acts; in terms of their discursive function. Second, once we have an account of the nature of the speech acts in mind, we need to say some more about why it would make sense for us to use these speech acts to engage in a discourse with the relevant structure. All of this needs to be done in a way consistent with a truth relativist semantics.

Suppose, first, we follow MacFarlane in adopting a Brandomian account of the fundamental nature of speech acts.⁹⁰ On this view, what speech acts fundamentally do is create and alter normative relations between speakers. When a speaker makes an assertion, in particular, this does two things. First, it extends a licence to other speakers to take the assertion as testimony. Second, it takes on a commitment to defend the assertion against challenges. Retraction, on this view, simply cancels the normative effects of the assertion. Rejection cannot exactly do this, for one cannot “unmake” an assertion made by another. However, it can do something analogous in the interpersonal case: it can call on others not to take up the licence offered by the targeted assertion (in effect, calling on others to act as if the assertion had been retracted).

This is a non-relativistic account, which requires some modification if it is to apply to truth-relativistic contents. For as we have seen, it is infelicitous to challenge in cases of fundamental disagreement. And therefore, we should not say that assertions of truth-relativistic contents take on a commitment to defend the assertion from challenge *against all comers*. However, the relevant modifications are fairly straightforward. We can say that an assertion of an assessor sensitive content first extends a licence to other speakers who share the asserter’s standards (or who have relevantly similar standards) to take the assertion as testimony, and second takes on a commitment to defend the assertion against challenges from those who share the asserter’s standards (or have relevantly similar standards). We can then say that retraction simply cancels these normative effects. And we can say that rejection calls on those who share the rejecter’s standards (or have relevantly similar standards) to ignore the licence extended by the targeted assertion.

All that this shows, however, is that the Brandomian story can be carried over to speech acts with truth-relativistic contents. What we are still in need of is an explanation of why a practice of using these speech acts in ways governed by the relevant norms would make sense. But such an explanation is not so difficult to construct. Take, as an example, the case of music. The set of norms we have envisioned in fact fits quite naturally with our best understanding of the aetiology of music: which is that music is an evolutionary adaptation which functions to mark

⁹⁰ (MacFarlane, 2014: ch5) (Brandom, 1994).

and strengthen in-group cultural identity⁹¹ (intuitively, think of the importance of taste in music in marking out certain subcultures, or the social role of national anthems and religious music). It is easy to see how the normative contouring of the practice which we have outlined fits with this idea. For first, this function would be served by the possibility of *rejecting* the aesthetic-musical judgements of those who do not share one's tastes. The possibility of this kind of rejection would serve the function of signalling to one's own group that one is *not* aligned with the tastes of the out-group. However, there would be no obvious functional utility to *inter*-group challenges or responses. For resolving intergroup tensions is not central to the function of music, on this view. There would, however, be utility to *intra*-group challenge and response, for this practice would serve to coordinate the aesthetic judgements of members of the group. Thus, the kind of normative practice we have envisioned (rejection *across* taste standards, but challenge and response only *within* taste standards) seems to fit quite naturally with this understanding of the adaptive function of music.

Further review of the empirical literature on other aesthetic domains is necessary. And we cannot complete this review here. However, there is clear evidence that our best understanding of at least one important domain fits well with a Brandomian account. And it is not unreasonable to think that similar considerations would apply to other aesthetic domains.

Fourth objection: Even if the rejection account is preferable to MacFarlane's account in some of the ways discussed here, the two accounts share a fundamental weakness. Like MacFarlane's account, the rejection account seeks to explain the intuition that Bradshaw and Calhoun disagree in terms of a speech act that Bradshaw and Calhoun may perform when they engage in conversation. This may explain the sense that there is an *activity* of disagreement which Bradshaw and Calhoun can engage in. However, it leaves unexplained the sense that Bradshaw and Calhoun would stand in a *state* of disagreement, even if they had never met.⁹²

⁹¹ This research is nicely summarised in (Hallam et al., 2009: ch1).

⁹² See (Cappelen & Hawthorne, 2009) for discussion of the distinction between state and activity disagreement.

Response: Although the rejection account *starts* with the activity disagreement, this does not mean that it can provide no account of state disagreement. For we can say, roughly, that two people are in a state of disagreement in virtue of having beliefs which would make activity disagreement permissible (in light of the norms we have been discussing) *if* these two people were to find themselves in conversation. In particular, we can say that A and B are in a state of disagreement as to whether p iff it would be permissible for B to reject an assertion of p by A in a hypothetical conversation between A and B.⁹³

4. Moral disagreement

Now that we have a blueprint of a truth-relativistic account of disagreement in belief/assertion, it is time to turn our attention to disagreement in ethics. Can the rejection account also explain ethical disagreement in belief/assertion? Or is it limited to application in domains like aesthetics?

I take it that the onus here is on the opponent of moral truth relativism to present reasons why the account *cannot* be applied in ethics, rather than on the truth relativist to come up with reasons why the account *can* be applied in ethics. This is because we have already seen that the account is applicable to at least some domains, so if it is *inapplicable* in ethics this can only be because there are particular features of ethics (features we do not find in aesthetics) which do not fit well with the account. I can think of two such worries.

4.1 First worry about moral disagreement

First worry: There is more emphasis on challenge and response in ethics than there is in aesthetics. We do *not* normally say “there is no disputing about *morality*”. Rather, we generally assume that moral differences *can* and *should* be rationally resolved, and we think of challenge and response as a perfectly sensible way of going about this. But truth relativism, as we have seen, entails that it does not even make sense to issue challenges or responses in cases of

⁹³ (Plunkett & Sundell, 2013) appeal to a response of this kind. Although see n26, chapter 2.

fundamental moral disagreement. Truth relativism is therefore a poor explanation of this aspect of moral discourse.

Response: Certainly, we do not normally take the “live and let live” kind of attitude towards moral disagreements which seems perfectly apt in cases of aesthetic disagreement. We normally do try and resolve moral disagreements via challenge and response, even when these disagreements take place at what seems like substantial cultural distance. However, the key here is to remember that truth relativism predicts that challenge and response ceases to make sense *only* in cases of *fundamental* disagreement. That is, in cases of disagreement which would persist after idealisation. And for reasons we discussed at length in chapter 1, it is not at all clear that there actually are any such disagreements. For it is not at all clear that substantial moral disagreement would persist after idealisation. Given this lack of clarity, and given the *pragmatic* reasons which exist for people to *assume* that moral disagreements can be resolved through idealisation (in other words, to assume that there is some substantial, even if nonobvious, overlap in moral standards between them and their interlocutor), it should not be surprising, even given truth relativism, that we almost always attempt to resolve disagreements via challenge and response, and that this almost always seems to make perfect sense. (See chapter 5 for further discussion).

Even if an interlocutor were to accept this point, though, it might still seem that it is insufficient to deflect the present objection. After all, even if fundamental moral disagreements are extremely rare (indeed, even if they are purely hypothetical) it might still seem intuitively plausible that it would make sense to engage in the issuance of challenges and responses in such cases.

In fact, though, it is not obvious that challenge and response would make sense in such cases. For in fact, it is quite plausible to think that there *is* something defective about carrying on with moral debate once it becomes clear that one’s interlocutor’s standards are *quite so* far apart from one’s own. Now, we have to be careful to note that this is not a matter of mere futility. As Dreier points out (2009), the worry about challenge and response for the truth relativist is not merely that it may be futile. After all, it may be *futile* to argue with flat earthers,

or with creationists, in the sense that there is very little chance of changing their minds. But as Dreier points out, these kinds of arguments make perfect conversational sense even if they are not a good use of one's time. What we have to show here, if we want to defend the truth relativist position, is that there is reason to think that moral debates at extreme cultural distance are *conversationally defective*, not just that they are futile.

However, even with this proviso, we can notice that there *is* reason to think that such debates would be defective. Consider Elizabeth Anscombe's attitude to the Utilitarian who thinks that it could be morally right, in some far-fetched contexts, to execute an innocent person:

But if someone really thinks, in advance, that it is open to question whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration – *I do not want to argue with him; he shows a corrupt mind*. (Anscombe, 1958) [emphasis added]

While this is perhaps a strong stance to take towards a utilitarian, most of us would surely be inclined to take this attitude towards at least *some* interlocutors (Nazis, perhaps, or the medieval Samurai, or the Vogons discussed in the previous chapter). And this attitude *does* seem to be one to the effect that there is something *deeply defective* (not just futile) in carrying on debating with people whose standards are *so* far removed from our own (who are "beyond the pale" as it were).

It is interesting to compare this point with differences in our attitudes towards the value of *open-mindedness* across moral and non-moral debates. In particular, open-mindedness seems somehow more appropriate in *non-moral* cases than it does in moral cases. On the one hand, if we are talking to a flat-earthier, it seems that open-mindedness is a virtue. There is at least something admirable in being open to the possibility that they are right, even if our prior credence in that possibility is vanishingly small, and there is at least something admirable in debating with such a person in good faith. On the other hand, suppose we are having a conversation with someone who endorses slavery, or is a full-blown Nazi. In that case, there no longer seems to be anything admirable about taking an open minded stance, or debating

with them in good faith. Rather, it seems that one should take something like Anscombe's stance in such a case, one should not want to argue with them; they show a corrupt mind.

We must also bear in mind, here, the results of the experiments discussed in the previous chapter. What those experiments showed is that people are willing to say that in cases of extreme cultural distance (such as involving alien species) that both sides' beliefs may be correct. But notice that Dreier's argument *against* the compatibility of challenge and response with truth relativism (against MacFarlane's view) appealed to precisely this state of affairs. In other words, our argument was that in a case where two contradictory assertions are both correct (and both parties to the discussion recognise this), it cannot make sense for the parties to challenge each other's assertions. But if this was right, then it follows that there is (indirect) experimental evidence for the claim that challenge and response does not make sense in such cases. For there is direct experimental evidence that both assertions are correct in such cases, and we have seen that this entails that challenge and response would be defective.

A final point on this is that none of this means that we should take the kind of *laissez-faire* attitude towards moral disagreement that we take towards aesthetic disagreement: not even in the fundamental cases. For even in these cases, there will remain an important form of disagreement in attitude. We will strongly desire that our interlocutors change their behaviour (and even, arguably, that they change their minds), and our interlocutors will strongly desire the reverse. Furthermore, if we ever come into conversational contact with such people, we will have a heated practical dispute, which may be carried on with threats or pleading, or other persuasion tactics, even if not with explicit, reasoned, moral debate. The point then is that we should not expect a "live and let live" attitude to prevail in cases of fundamental moral disagreement, even if truth relativism is true. All we should expect is an absence of moral *debate*. And this prediction is, in fact, quite plausible. (See chapter 5, section 1.2, for related discussion (on the quasi-realistic aspects of truth relativism).)

3.1 Second worry about moral disagreement

Second worry: It's less clear, in the moral case, that there is a satisfying story to be told about how the rejection account connects up with the evolutionary point of the discourse. In the aesthetic case, we saw that there is evidence that the evolutionary point of aesthetics primarily involves intra-group dynamics. This supports the rejection account, for the rejection account seems well suited to serve such a function. But it is less plausible that this is the case in ethics. After all, part of the point of ethics, presumably, is not only to regulate in-group co-operation, but also to resolve co-operation problems *between* groups. Why then, should we think that it makes sense for our discursive practice around *moral* disagreement to be governed by truth relativistic norms?

Response: Here, again, it is crucial to understand that the truth relativist is proposing that challenge and response cease to make sense only in cases of *fundamental* disagreement – such as the kinds of cases which prompted intuitions of faultless disagreement on the part of participants in the experiments discussed in chapter 2. The truth relativist is *not* proposing that challenge and response ceases to make sense in the kind of inter-group disagreements that would have actually featured in morality's evolutionary history, or which we actually encounter today. Rather, what the truth relativist proposes is that most, if not all, actual moral disagreements function exactly as if invariantism were true (i.e. that challenge and response make perfect sense in such cases) but that ordinary speakers will tend to go relativist (i.e. fall back on rejection alone) in cases (such as those involving extraterrestrials or mediaeval samurai, or whatever) where it is clear that the disagreements are fundamental in the relevant sense.

Furthermore, when one reviews the literature on the evolutionary function of morality, one finds that this proposal is not an unnatural one. As Stanford (2018) explains, the most promising explanation of the primary function of moral psychology is that it aims to promote mutually fitness enhancing cooperation, while also (importantly) protecting altruistically motivated individuals against exploitation. Thus, the primary function of moral *discourse* is to seek agreement on moral principles between potential partners in co-operation. The benefit of this agreement is that all parties are aware that moral commitments are experienced as overriding and universally binding. And this provides a guard against exploitation between

those who agree about what is morally required. The key point of relevance for us is that this function is specifically oriented towards agreement with those *who could realistically serve as partners in social co-operation*. It does not, therefore, give us reason to expect a practice of challenge and response in cases of clearly *fundamental* disagreement. For in these cases, our interlocutors are not plausibly thought of as potential co-operators. Of course, it will still make sense for us to *reject* the assertions of our interlocutors, for this will serve to signal our suitability for co-operation to those with whom we share standards. Thus, the account that the truth relativist is proposing in fact fits quite naturally with what we know about the evolutionary point of morality. Indeed, it is not clear that there is any less natural a connection in the moral case than in the aesthetic case. Bernard Williams' "relativism of distance" again seems to nicely capture the intuitions we would expect to find if this explanation of the adaptive point of morality were correct. In cases of "real confrontation", in Williams' idiolect, relativism will seem less plausible. But in cases of "merely notional confrontation", where cooperation is not a realistic possibility, relativism seems more plausible. Our best understanding of the adaptive function of morality gives us no reason to expect a practice of challenge and response in cases of merely notional confrontation. (See Chapter 5, section 2, for further discussion.)

5. Conclusion

This chapter has shown that truth relativists can construct a plausible account of disagreement in belief/assertion. They need not be satisfied only with disagreement in attitude. This conclusion is interesting in itself, but also serves to further the truth relativist's advantage over content relativism on the issue of disagreement. In terms of the bigger picture of the thesis, the point we have reached is this. In the first three chapters, we have discussed moral disagreement. The first chapter found that epistemic considerations, around disagreement, provide only limited reason to reject moral realism. In the second chapter, however, we found that an argument from "faultless disagreement" provides clear reason to accept moral truth relativism in particular. In this chapter, we have shown that an account of disagreement in belief/assertion is available to moral truth relativists.

The form of moral relativism we have focused on in chapters 2 and 3 is an innovation imported from recent debates in the semantics of “subjective” predicates, such as those of taste and aesthetics. One of the key points in the discussion has been that this new-style relativist semantics does a better job, than content relativism, of explaining disagreement. In this respect, new-style relativism (truth relativism) helps solve a long-standing worry about views flying the banner of “moral relativism”, namely that they fail to explain the possibility of moral disagreement. Content relativists have, however, not been silent on this issue. We cannot, therefore, lazily assume that the problems identified for content relativism are fatal. In chapter 4, we will discuss the problem for content relativism in more detail. Only then will be able to conclude that the truth relativist can mount a more effective defence.

CHAPTER 4

ABSTRACT: The arguments advanced so far (in chapters 2 and 3) have appealed to a certain objection to contextualism, namely that it is incompatible with data involving alethic evaluations and belief reports. While this objection has been treated, so far, in a cursory manner, this chapter discusses it in detail, arguing that contextualist responses fail. The most promising strategies, due to Finlay and Suikkanen, are discussed and rejected. Skepticism about the prospects for future development of contextualist strategies is also defended.

1. Introduction

The preceding two chapters have made a case for truth relativism in ethics. A key plank in the argument for truth relativism has been the claim that the truth relativist's account of disagreement is superior to the content relativist's account. In chapter 2 (section 4.2) an argument was given which appealed, in particular, to the infelicity of utterances like:

"Everything my interlocutor believes about morality is true, and everything that she believes about morality is such that I believe it as well. However, we have many deep moral disagreements."

As we saw, the problem for content relativism was that content relativists who endorse a *disagreement in attitude* account of faultless disagreement, seem to be committed to holding that such utterances can be correct. Chapter 3 appealed to a related consideration. There, we saw that content relativists cannot endorse a *rejection account* of disagreement, because they would then be committed to the view that utterances like:

“strictly speaking, what my interlocutor just said is true, but I reject that”

could be felicitous. But since both utterances are deeply infelicitous, these implications pose a problem for the content relativist’s account of faultless disagreement.

Let us recap some of the dialectic. The discussion has centred around cases like:

Hyperspace Bypass: The Vogons are a species of extraterrestrials. A representative Vagon, Jeltz, believes that it would be morally permissible to demolish the earth to make way for a hyperspace bypass. Arthur Dent, a representative human, believes that demolishing the earth to make way for a hyperspace bypass would be wrong. Both hold their beliefs in states of wide reflective equilibrium.⁹⁴

Experimental data (Sarkissian et al., 2011) (Khoo & Knobe, 2016) suggest that ordinary speakers regard such cases as involving both a strong form of faultlessness, and genuine disagreement. While both content relativism and truth relativism succeed in accounting for the faultlessness data, the challenge for contextualism is that it is *prima facie* inconsistent with the disagreement data.

(As a reminder, the reason for the focus on the exotic, science-fictional case is that it is only in such cases that the possibility of faultless disagreement derives clear support from experimental data).

The challenge for contextualism arises because it entails that the *contents* of Arthur and Jeltz’ beliefs are determined by their differing standards. Thus, on an example view, what Jeltz believes is that destroying the earth has some natural property, $\langle$, and what Arthur believes is that destroying the earth does not have some *other* natural property $\otimes$. Thus, the propositions they believe are straightforwardly compatible. Relativism, in contrast, entails that one and the same content, the proposition that demolishing the earth to make way for a hyperspace bypass is permissible, is accepted by Arthur and rejected by Jeltz.

⁹⁴ (Adams, 2007)

In response, as we have seen, contextualists can appeal to the phenomenon of disagreement in attitude. For, as we have seen, it is in general plausible that there can be disagreements which involve clashing non-cognitive attitudes, even in the absence of incompatible beliefs. An example that will be useful for our purposes, in this chapter, is this. Suppose Billy's mother is worried about the dangers of American football, and wants him not to play. Suppose Billy agrees (in belief) about the dangers, but wants to play anyway (Stevenson, 1937) (Ridge, 2014). It is natural to think that Billy disagrees with his mother, even though they seem to agree in all their beliefs. Rather, the disagreement seems to stem from their clashing non-cognitive attitudes: Billy wants to play, his mother wants him not to play. Helpfully for the content relativist, it is independently plausible that superficially inconsistent moral judgements involve clashing non-cognitive attitudes, of the right kind to generate disagreements in attitude.⁹⁵ Thus, content relativists can plausibly claim to secure a sense in which Jeltz and Arthur disagree, despite the compatibility in the contents of their beliefs.

However, as we saw in chapter 2, the content relativist is not yet out of the woods. Disagreement in attitude was originally put to work in order to show that *non-cognitivists* could account for moral disagreement. On these early non-cognitivist views, the problem) was not that Arthur and Jeltz have beliefs with compatible contents. Rather, the problem was that Arthur and Jeltz do not have moral beliefs *at all*, and thus *cannot* disagree in belief. Thus, appeals to disagreement in attitude were originally intended as a stand-alone account of disagreement, which was intended to harmonise with the preferred account of semantic content.

⁹⁵ Although it's non-trivial to specify exactly which attitudes are involved, this is plausibly everyone's problem. This is because one does not have to think that moral judgements *just are* non-cognitive attitudes in order to think that moral judgements are at least systematically accompanied by non-cognitive attitudes (even if this is merely a deep feature of contingent human, or Vagon, psychology). It is thus hard to see why anyone would deny that moral disagreement in attitude exists. There is perhaps a lingering problem for the proposal to account for all disagreement in this way, namely that it seems that so-called "amoralists" (hypothetical individuals with moral beliefs but no non-cognitive attitudes) can have moral disagreements. But I will not explore this point here.

If one is instead a content relativist, there is still a sense in which disagreement in attitude can provide a stand-alone account of disagreement, since disagreement in attitude is in general *sufficient* for disagreement simpliciter. However, one might reasonably worry that this account contains an internal tension. After all, according to content relativists, Arthur and Jeltz *do* have moral beliefs with descriptive contents, and importantly they seem quite capable of *agreeing in respect of these beliefs*. Thus, there seems to be a tension between the cognitive and non-cognitivist aspects of the content relativist's proposal. In particular, content relativism seems to be committed to the view that Arthur and Jeltz both agree and disagree, by making the moral judgements that they do.

To put the objection another way, it seems (as mentioned above) that contextualism entails that Arthur could correctly say:

(1) Everything Jeltz believes about morality is true. Indeed, everything Jeltz believes about morality is such that I believe it too. Nevertheless, we have a moral disagreement.⁹⁶

This is because, according to content relativism, *what Jeltz believes* is simply the proposition that destroying the earth has a certain ordinary natural property (determined by Vagon standards), and Arthur Dent could know that this proposition is straightforwardly true.

(1), however, seems clearly incorrect. Instead, it seems that Arthur can correctly say:

(2) "what Vagon Jeltz believes about morality is false. And I reject his moral beliefs. Thus, we disagree".

But since content relativism entails that the contents of Jeltz' beliefs are straightforwardly true, and that Arthur could know this, *and yet that they disagree*, it seems to follow from contextualism that Arthur would be making a mistake by uttering (2).

⁹⁶ (Ridge, 2014).

Indeed, contextualism seems to entail, paradoxically, that Arthur could correctly say:

(3) I agree with everything Jeltz believes about morality, but I Jeltz and I have many deep moral disagreements.

This is not to say that simultaneous agreement in belief and disagreement in attitude are in general incoherent. Recall Billy and his mother's disagreement. Plausibly, Billy could correctly say "everything mom believes about the dangers of football is true, but we disagree – I want to play and she won't let me." Rather, the problem is specific to moral discourse. It does *not* seem that we can both agree and disagree in virtue of making a particular pair of superficially opposing moral judgements, as the infelicity of (1) and (3) shows.

It might be wondered what kind of knowledge needs to be attributed to ordinary speakers to make these objections work. Certainly, it would be implausible to attribute to Arthur explicit knowledge of the inner workings of the semantics of moral utterances. For example, it would be implausible to attribute to him the explicit knowledge that Jeltz' belief can be paraphrased as 'demolishing the earth to make way for a hyperspace bypass has natural property ®'. Nevertheless, it is reasonable to expect ordinary speakers to be able to track the variations in content brought about by semantic context-sensitivity, and to be able to deploy this implicit knowledge in assessing utterances like (2). For example, consider:

(4) the NBA commentators said that the point guard was short. What they said was true. But he's not *short*.

Admittedly, it would be felicitous to make the relativisation explicit here ("but he's not short in general/short for an adult male human"). Nevertheless, not doing so (as in (4)) is perfectly felicitous, and arguably more natural. Uttered in a context where ordinary standards for height are salient, (4) would be correct, and competent speakers would be able to recognise this. Thus, there is *prima facie* reason to expect competent speakers to be able to see that (1) is correct, given content relativism. Thus, the fact that competent speakers do not view (1) as correct, provides apparent evidence against content relativism.

Thus far, all we have done is flesh out considerations mentioned briefly in chapters 2 and 3. But in previous chapters, we discussed these points only in passing. Content relativists have responded to this objection at some length, and, in chapter 2, we examined content relativist responses only very briefly. The argument would thus be incomplete without a more detailed treatment of these responses. The aim of this chapter is to provide this treatment.

To be clear, and again to recapitulate, the challenge for content relativists is not to explain disagreement *per se*, for, as we have seen, it is consistent with content relativism that Hyperspace Bypass instantiates a disagreement in attitude. Nor is the challenge, more than this, to explain why Arthur Dent and Vogon Jeltz have a disagreement in belief. This demand would arguably be unfair. For it is not clear that our intuitions are fine grained enough to provide evidence for the instantiation of some particular form of disagreement, rather than the instantiation of at least one genuine form of disagreement. Instead, the challenge for content relativists is to predict the right results about belief ascriptions and alethic evaluations, thus explaining the infelicity of (1), as well as (cognitive) agreement/disagreement ascriptions, thus explaining the infelicity of (3).

Indeed, we should notice that although this challenge is closely connected with the phenomenon of disagreement, it can arise in contexts where disagreement is not salient. For example, suppose the Vogons do not destroy the earth. Suppose that many years later, earth historians review the incident, they say: “interestingly, the Vogons believed that demolishing the earth was morally permissible. We do not believe this ourselves, of course. But we want to understand how Vagon morality developed in this way.” Admittedly, these historians would seem to be *committed* to disagreeing with Jeltz, but they need have no interest in expressing this disagreement. Nevertheless, the example illustrates a data point that content relativists would seem unable to explain, namely that the historians can disavow the content of Vagon Jeltz’ belief.

Indeed, there is a challenge for contextualists to explain even *the commitment to disagreement* (Ridge, 2014), for it is not obvious that the historians would express non-cognitive attitudes

via this disavowal, and it is thus not obvious (on the present content relativist proposal) why they would be committed to agreeing or disagreeing in virtue of disavowing the belief. The problematic data involving alethic evaluations and belief ascriptions are thus present in a broader range of contexts than a narrow focus on agreement and disagreement might suggest.

There is another way, too, in which a narrow focus on disagreement might obscure the breadth of the problem. For a similar problem arises in cases of agreement, and cases where one party suspends judgement. First, in the case of agreement, suppose that Vogon Jeltz believes that hitch-hiking is morally permissible. Suppose Arthur Dent believes this too. But suppose that Vogon Jeltz is making a mistake in the sense that Vogon Morality does not, in fact, permit hitch-hiking. In that kind of case, we get a structurally similar problem: contextualism would seem to imply that Arthur Dent could correctly say “Jeltz just said that hitch-hiking is permissible, I believe that hitch-hiking is permissible, but what Jeltz just said is false, and I reject it”. This is because the content of Jeltz’ belief is the proposition that hitch-hiking has a certain natural property, and it is false that hitch-hiking has this property, and Arthur Dent could know this. Second in the case of suspension of judgement, suppose Jeltz believes that torture is permissible in at least some cases. Suppose Arthur is unsure whether torture is ever permissible, and thus suspends judgement about whether torture is permissible in at least some cases. It seems correct for Arthur to say “Vogon Jeltz just said ‘torture is sometimes permissible’ but I am not sure whether that is true”. But content relativism seems to instead imply that Arthur could correctly say “Vogon Jeltz just said ‘torture is sometimes permissible’ I am not sure whether torture is ever permissible, but I am sure that what he said is true”. This is, again, because the content of Jeltz’ belief is determinately true, and Arthur could know this.⁹⁷

A final variation on the problem is this. As Schroeder (2009) points out, inferences of the following form seem obviously valid:

⁹⁷ As we noted in chapter 3, rejection dialogues constitute an additional data point. Contextualism seems to entail that utterances of the form “Jeltz just said ‘demolishing the earth is fine’”. Strictly speaking, what he said is true. But I reject that, it would be wrong to demolish the earth”.

Arthur: Jeltz just said that demolishing the earth to make way for a hyperspace bypass would be morally permissible.

Arthur: But demolishing the earth to make way for a hyperspace bypass would be wrong.

Arthur: Therefore, what Jeltz just said is false.

But content relativism seems to imply that this inference is *invalid*. After all, according to content relativism, the content of Jeltz' belief is perfectly compatible with that of Arthur's. And thus, there should be no valid move from the truth of what Arthur believes to the falsity of what Jeltz believes. But intuitively, as this example shows, such moves *are* valid. So there is another piece of data, here, that content relativism seems unable to explain.

In what follows, I will for simplicity's sake set these additional data points aside and concentrate on utterances like (1), (2) and (3). But it is important to notice that the data that need explaining are more widespread and varied than this focus may otherwise suggest. Perhaps the most important *prima facie* advantage of truth relativism, over content relativism, is that, unlike contextualism (as I will here argue), truth relativism straightforwardly explains all these data, by endorsing the intuitively plausible claim that what is said and believed is invariant across speakers and thinkers.

Finally, it should be clarified that this problem arises only in cases involving potentially faultless disagreement; that is, cases where the speaker's or thinker's moral codes would diverge after idealisation. This is because if the speaker's or thinker's codes would *converge* after idealisation, then contextualists could hold that their moral utterances and beliefs make reference to the same sets of natural properties. And contextualists could therefore hold that such cases involve straightforward disagreement in belief. It is fair to object that if this *only* applies to highly esoteric cases such as Hypersapce Bypass, then the putative theoretical advantage associated with truth relativism may be a relatively narrow one. However, without any further response, this would render content relativism somewhat a hostage to empirical

fortune. For as we have seen in chapter 1, it is *unclear*, on empirical grounds, whether there actually are any faultless moral disagreements. And this response thus leaves open that there *may* be actual cases (perhaps between different societies of ancient history, more on this in chapter 5, section 2) in which the content relativist is saddled with the implausible commitments just detailed.

With these points in hand, let us turn to the main aim of the chapter, assessing putative content relativist solutions to this problem. We will eventually see that these solutions are not satisfactory, and will therefore finally be able conclude the argument for relativism initiated in chapter 2.

Although there is a large literature on content relativist accounts of disagreement, much of this literature does not focus on the problem just outlined. Rather, most discussions focus on securing disagreement *per se*, either via disagreement in attitude (Dreier, 2009) or via an alternative strategy which appeals to so-called meta-linguistic negotiation (Silk, 2017) (Plunkett & Sundell, 2013).⁹⁸ I will not discuss these strategies, as I am happy to grant that content relativism has the resources to secure genuine disagreement in cases like Hyperspace Bypass. Indeed, I believe this should be fairly uncontroversial, as it merely requires that moral assertions express clashing non-cognitive attitudes, which is independently plausible. Other content relativists go in the opposite direction, and simply deny that there is disagreement in cases like Hyperspace Bypass (or at least deny that this is a datum that our theories should aim to secure (Dowell, 2016)). I will not discuss this argument, either, since I regard it as a last resort, to be avoided if the content relativist can do better.

Instead, I will focus on two content relativist proposals which not only aim to secure disagreement *per se* but, more than this, try to defend against the objection outlined above, namely Finlay's (2014)⁹⁹ and Suikkanen's (2019).

2. Finlay's response

⁹⁸ (Finlay, 2017) provides a nice summary of this research.

⁹⁹ An earlier version of Finlay's response appears in (Finlay & Björnsson, 2010).

Finlay concedes that utterances like (1) are correct in the sense that (1) is strictly speaking true at Arthur's context of utterance. However, he argues that utterances like (1), while correct in this sense, are incorrect in another, and more important, pragmatic sense. This tension between truth-normative and pragmatic requirements is resolved, on Finlay's view, by the postulation of a certain kind of error theory for this narrow fragment of moral language.

Let's spell out Finlay's proposal in more detail. We will then see why it fails to solve the problem for content relativism.

First, the key pragmatic point for Finlay is that our primary interest in moral discourse is communicating information about each other's behavioural dispositions, determined by the non-cognitive components of each other's moral mental states. Finlay also endorses a fundamental principle governing pragmatic requirements in general, namely that people always speak in the way that they believe best serves their conversational ends. Thus, in this context, the combination of these two claims predicts that Arthur will speak in the way that best serves the goal of communicating information about Jeltz' behavioural dispositions, as conditioned by his non-cognitive mental states.

Thus, on Finlay's view, although our moral assertions do express descriptive propositions, these are of secondary conversational importance. Rather, moral assertions are of conversational interest primarily in virtue of the non-cognitive attitudes that they also express (at the level of pragmatics). Equally, although there *can* be moral disagreements in belief (as occur whenever parties to a disagreement have standards which would converge in reflective equilibrium), these disagreements are in general of secondary interest, relative to the associated disagreements in attitude.

In cases of ordinary (non-potentially-faultless) disagreement, no difficulty arises from this. For in cases of ordinary disagreement, our disagreements in belief and disagreements in attitude will line up neatly with one another. For example, suppose Josh and Jess live in the same society. If Josh the meat eater says that eating meat is fine, he asserts a content which

Jess the vegetarian believes to be false, namely the content that eating meat has some natural property such that, we all agree, actions which have this property should not to be disapproved of. Jess is thus able to say 'that's false', and thereby commit herself to a belief which is systematically associated with a non-cognitive attitude of disapproval of eating meat. Thus, uttering a correct alethic evaluation, in this context, also allows Jess to speak as is pragmatically best.

However, in cases of potentially faultless disagreement, such as Hyperspace Bypass, these claims (in concert with Finlay's contextualist semantics) lead to a tension between pragmatic and truth-normative requirements. This is because, on the one hand, it is correct given the semantics to assert (1). But, on the other hand, the first two sentences of (1) "everything Jeltz believes about morality is true. Indeed, everything Jeltz believes about morality is such that I believe it too" might easily be misunderstood as expressing pro-attitudes towards Jeltz' behavioural dispositions. (This is because in non-potentially-faultless contexts, which are much more common, asserting these sentences would express a pro-attitude towards Jeltz' behavioural dispositions, as we have just seen). Uttering (1), thus, while technically correct, runs a great risk of directly vitiating Arthur's conversational ends, by conveying the exact opposite of the non-cognitive attitudes which it is his primary conversational aim to convey. Conversely, there is a similar tension concerning (2). In contrast to (1) Finlay's semantics implies that (2) is technically *incorrect*. However, also in contrast to (1), uttering (2) *would* serve to further Arthur's conversational ends, because (given the story just told) Arthur could anticipate that the first two sentences of (2) are likely to be interpreted as expressing precisely the non-cognitive attitudes that Arthur intends to express.

Having identified this tension between semantic and pragmatic considerations, Finlay proposes a neat solution. As he points out, there is a proposition which is closely related to the one which is actually picked out by 'what Jeltz believes', and which is such that, if it were to be picked out by this locution, the pragmatic tension just described would disappear; namely, the proposition Jeltz *would have believed* if he had shared Arthur's standards. If we understand Arthur, in uttering (2), as evaluating *that* proposition, then the semantic

incorrectness of (2) disappears (as does, *mutatis mutandis*, the semantic correctness of (1)), and the pragmatic and truth-normative requirements fall into alignment.

This proposal involves attributing a mistake to ordinary speakers, like Arthur. For Finlay does not hold that, in cases like *Hyperspace Bypass*, Arthur's evaluation *actually* targets the proposition Jeltz would have expressed in Arthur's context.¹⁰⁰ Rather, Finlay holds that speakers *intend* to alethically evaluate the proposition that their interlocutor's utterance would have expressed, but that, in doing so, they fall into a subtle misunderstanding of the meanings of the words.

Finlay provides detailed explanations as to why this error might occur. However, we should understand, at outset, that this is a move which, in general, there is some presumption against. For in general, it would be surprising if entire communities of competent speakers were systematically confused about the meanings of their own words (in this case, the meaning of 'what is said/believed' and relevant demonstratives, e.g. 'that' (as in 'that's false', 'I don't believe that')). Although such gaps in semantic competence are perhaps possible, positing such gaps in just the right places to preserve one's semantic theory may seem *ad hoc*, without further explanation.

Finlay, however, provides plenty of further explanation. In addition to the pragmatic story just told (which already provides some explanation, in virtue of showing that pragmatic considerations favour the error) Finlay offers two additional considerations which are intended to explain, first, why ordinary speakers would fall into the relevant confusion, and, second, why ordinary speakers may, on some level, consciously view the relevant pattern of speech as a useful fiction.

Finlay first points out that the relevant contexts are especially cognitively demanding, as they require speakers to track two subtle distinctions, which speakers do not ordinarily have to track. First, Arthur has to account for both his own and Jeltz' context of use, and correctly

¹⁰⁰ Although Finlay defended this view in his earlier (2010), (with Gunnar Björnsson), it is retracted in his (2014).

register the semantic significance of both at different points in the same sentence (Jeltz' context for the first two sentences, and Arthur's context for the third sentence). And second, in order to recognise the semantic correctness of (1) Arthur has to be able to draw the distinction between disagreement in belief and attitude. This is because if Arthur fails to recognise this distinction, he may easily make an inferential move from the existence of a disagreement simpliciter to the existence of a disagreement in *belief*, and thus to a commitment to the falsity of what Jeltz' believes. Since it is plausible that ordinary speakers do not recognise this distinction, it is reasonable to expect that ordinary speakers would systematically draw this faulty inference.

The second plank of Finlay's explanation is an appeal to another key aspect in his overall theory, which he calls "rhetorical objectivity", which it will be helpful to briefly explain. The primary role that the notion of rhetorical objectivity plays in Finlay's theory is to explain the fact that even though moral utterances encode relativised contents (on Finlay's view) moral assertions are almost never explicitly relativised. This appears problematic, since utterances of paradigmatically context-sensitive expressions, such as 'tall' can normally be made explicit (as in "when I said 'he's tall', all I meant was that he's tall for a basketball player"). Finlay's appeal to "rhetorical objectivity" is the idea that this absence of relativisation in the moral case is explained by the fact that suppressing the relativisation expresses a pragmatic presupposition that the relevant contextual parameter is shared by all parties to the conversation. By leaving the relativisation of moral utterances implicit, speakers thus, in a sense, prescribe that all parties to the conversation share their ends. For Finlay, this helps explain a key feature of moral discourse (although one which is not our central interest here) namely that statements of moral requirements appear categorical. On Finlay's view, statements of moral requirements appear categorical, because they carry a kind of pragmatic prescription which is independent of the audience's actual ends.

Returning to the present point, Finlay draws on this independent feature of his theory to buttress his defence of the localised error theory that is our present interest. In particular, he argues that if rhetorical objectivity is systematic enough, speakers may simply assume that the objectivity is semantic, rather than pragmatic, thus (mistakenly) endorsing an invariantist

semantics. This would explain why they regard utterances like (1) as incorrect, for such utterances would be incorrect given an invariantist semantics. Furthermore, even insofar as speakers do recognise the merely rhetorical status of the practice of rhetorical objectivity (as Arthur should be able to, since the presupposition of shared standards is obviously false, in *Hyperspace Bypass*), it may be that rhetorical objectivity is so closely associated with distinctly moral discourse that any assertion which gives the lie to the practice may signal a lack of moral motivation, something against which there is, plausibly, significant social pressure.

Finlay thus offers a subtly mixed explanation via the appeal to rhetorical objectivity. On the one hand, the relativisations of moral discourse (in virtue of which utterances like (1) come out technically correct) may for the most part run “under the hood” unnoticed by most ordinary speakers, most of the time. On the other hand, inasmuch as speakers are, on some level, aware of this relativisation, it may be a kind of “elephant in the room” in the sense of a phenomenon which many people understand, but few are willing (on pain of social censure) to acknowledge. This suppression may not, moreover, be entirely irrational. Many who reject moral objectivity have argued that moral objectivity would be highly useful, and should be preserved as a fiction by those who share their anti-objectivist views. According to Finlay, this benign motivation may already be shared, in fact, by many ordinary speakers.

2.2 Response to Finlay

Finlay thus provides a nuanced and sophisticated explanation of the apparent infelicity of (1), compatible with his theory. Many of the moving parts in his argument appeal to plausible observations about the role of moral discourse in general, and features of cases like *Hyperspace Bypass*, in particular. However, Finlay’s account is not ultimately persuasive. Here’s why.

First, Finlay’s account is inconvenienced by the fact that the data discussed in chapter 2 suggest that ordinary speakers *are* happy to endorse the claim that Jeltz’ belief may be *correct*. The reason this is an inconvenient data point for Finlay is that his arguments would in general

seem to put equal pressure on such endorsements. For example, failure to draw the disagreement in attitude/disagreement in belief distinction would be equally likely to lead speakers to make a structurally similar inference, from the existence of disagreement simpliciter to the conclusion that Jeltz' belief must be incorrect. And, endorsement of the claim that Jeltz' belief may be correct would equally well undermine the useful fiction of rhetorical objectivity. Finlay's explanations are thus something of a blunt object in this respect. Even if they succeed in accounting for the apparent incorrectness of (1), they do so only at the cost of an account of the sense that the disagreement is faultless in the first place (including the experimental data from chapter 2), as this requires that ordinary speakers are able to recognise that Jeltz' belief is correct.

Even if we set aside this worry, however, it is in fact unclear that Finlay's account *does* as much as succeed in accounting for the apparent incorrectness of (1). Although Finlay provides many explanations for why ordinary speakers would find (1) infelicitous in ordinary contexts, some of these explanations are simply unconvincing, and others should be cancellable via tinkering with the characterisation either of the speaker or of the context. In other words (as (Ridge, 2014) puts the point) there should, in these cases, be *some way* to isolate contexts where, in lieu of unexplained semantic error, competent speakers should, even given Finlay's defensive strategies, be able to recognise the correctness of (1), if contextualism is true. In none of these cases, though, does (1) come to seem correct. Indeed, it is hard to imagine a context in which it would.

Begin with Finlay's view of the main conversational point of moral discourse, that it primarily aims at communication of non-cognitive attitudes, with the aim of understanding and regulating the behavioural dispositions of various parties. As a general claim about the primary function of moral discourse, this is, again, plausible. But inasmuch as this is purely a pragmatic matter (and thus independent of the semantic contents of moral claims, individuation of moral judgements, or the intrinsic nature of moral facts) we should be able to imagine conversational contexts in which this end is not primary. Consider, for example, the retrospective conversation between two academic historians, discussed in the introduction. To repeat, let us imagine that, in the end, the Vogons decide not to demolish the

earth. We can imagine that centuries after the events of Hyperspace Bypass, earth historians look back. They say “The Vogons believed that demolishing the earth to make way for a hyperspace bypass was morally permissible. Of course, we do not ourselves believe what they believed. But as historians, we do not aim to moralise, we aim rather to understand the Vogons’ point of view.” Here, it is not at all clear that the speaker’s main *conversational* end is to communicate information about his or her non-cognitive attitudes. And yet, the disavowal of *what the Vogons believed* looks entirely appropriate. In the absence of conversational pressure in favour of using the belief report to express the speaker’s non-cognitive attitudes, this sense of appropriateness would seem, on Finlay’s view, to amount to brute semantic error.

Turn, next, to Finlay’s claims about the cognitively challenging aspects of contexts like Hyperspace Bypass. First, although it is true that these contexts are challenging inasmuch as they involve tracking two contexts at once, it is not clear that this in general prevents competent speakers from understanding the semantic properties of context sensitive expressions. For example, if I am having a telephone conversation, in December, with my friend in Rio (while I am in Edinburgh) this would not normally lead me to misuse phrases like ‘here’ and ‘it’s raining’. It would be surprising, for example, if I used ‘that’s true’ to respond to the proposition my friend *would* have asserted if they had used ‘it’s raining’ in my context. Rather, if my friend (in Rio) says “it’s raining” (in December), one would expect me to be able to (correctly) say “that’s false”. Second, while, it is more plausible that difficulty in drawing the disagreement in belief/disagreement in attitude distinction would render ordinary speakers vulnerable to the relevant kind of error, this leaves it unexplained why trained philosophers continue to find (1) infelicitous on reflection (as evidenced by the longstanding status of (1) as the cause of worries about content relativism).

If we turn, finally, to the strategy of appealing to rhetorical objectivity, we need to consider the two faces of this strategy separately. For, recall, Finlay suggests *both* that competent speakers may, in virtue of rhetorical objectivity, fail to even understand that their words are semantically context sensitive, *and* that competent speakers may regard rhetorical objectivity as a useful fiction. But in the case of the former strategy, this merely seems to move the bump

in the rug, since the strategy seems to explain the narrow error theory by appeal to a broader one. And this broader error theory stands, itself, in need of explanation.

In the case of the second strategy, the suggestion is that people recognise the context-sensitivity of their language, and thus recognise that (1) is correct. However, they refrain from uttering (1) out of either benign, moralistic, motives, or fear of social sanction. This claim, however, does not seem to provide a satisfactory explanation of the resistance we feel in uttering (1). To see this, imagine a sentence which is technically true, but which is either socially censored or morally wrong to say (e.g. “we didn’t invite you to the party, because everyone hates you”). Very plausibly, our resistance to (1) is not of a piece with our resistance to sentences like this. Rather, (1) strongly seems to be not just socially censored or morally wrong to say, but rather *incoherent*. It is hard to see how the “elephant in the room” strategy can explain this.

Here is a proposal for an empirical test of the fictionalist face of Finlay’s appeal to rhetorical objectivity. Suppose we conduct an experiment with two conditions. In the first condition, we present participants, as a group, with statements like (1) and (2), and ask them to publicly assess these statements. In the second condition, we present participants with the same statements, but ask participants to assess them in private, and offer large rewards for correct responses. Finlay’s account would seem to predict significant divergence in responses across the two conditions, for the pressures in favour of maintaining the fiction are absent in the second condition. We might not expect complete endorsement of (1) in the second condition, given Finlay’s view, but we should expect to see some difference. This would be a surprising empirical result, though. What seems more likely is that judgements in the two conditions would not differ significantly.¹⁰¹

Overall, then, it does seem that each of Finlay’s strategies falls short in some way. The appeals to the cognitive demandingness of tracking multiple contexts, and to the notion of rhetorical objectivity simply fail to satisfactorily explain the apparent incorrectness of (1). And, while the other considerations (e.g. ordinary speakers’ ignorance of the disagreement in

¹⁰¹ Thanks to Guy Fletcher for this suggestion.

attitude/belief distinction) do make some progress on the apparent incorrectness of (1), they seem to admit of exceptions, where Finlay's view would seem to predict that (1) seems correct to competent speakers. Since (1) does not seem correct in these exceptional contexts, there remains reason for skepticism about Finlay's account.

Furthermore, even if these explanations were more successful in their own terms, there is something at least somewhat unappealing about the extent of variation in the mechanisms that Finlay appeals to. Although Finlay's account is highly creative and ingenious, it seems somewhat convenient to find that so many disparate phenomena work in unison to plug the gap in Finlay's theory. Such a disparate story might be true, but without wanting to digress too far into debates over criteria for theory choice, it does seem that the relativist's explanation (by appeal to the single claim that content is invariant across thinkers and speakers) is at least somewhat more attractive, on theoretical grounds, than Finlay's account, other things being equal.

3. Suikkanen's response

Perhaps, though, alternative contextualist-friendly strategies may fare better. Suikkanen (2019) defends a form of contextualism which is closer to the version of relativism defended here inasmuch as it holds that moral truth is relative to the standards of a social group, rather than those of an individual. Suikkanen builds on this feature of his contextualism, combined with recent proposals in the semantics of epistemic modals, to develop a response which differs interestingly from Finlay's. As we will see, however, Suikkanen's response, while interesting and innovative, is not persuasive.

Suikkanen's idea is that instead of putting forward a single proposition, moral assertions typically put forward a "cloud" of propositions, determined by the various possible ways of carving up the contextually relevant society or culture. Thus, for example, suppose Boris is at a conservative London members' club and asserts 'it was right for the U.K. to leave the European Union'. On this view, his utterance can be understood as putting into play the proposition that relative to the club's culture, Brexit was right; the proposition that relative

metropolitan London culture, Brexit was right; the proposition that relative to U.K. culture, Brexit was right; and so on. This set of propositions forms the “cloud” that is the content of Jones’ assertion.

How does this solve the present problem? To see how, consider the following way of framing the difficulty with a traditional single-content contextualist model. Since, for the contextualist, contents are true or false absolutely, the single-content model entails that an assertion can be correct only if the content it puts into play is true absolutely. But this, problematically, implies that if two superficially contradictory assertions are both correct, both must have contents that are true absolutely, and the present problem ensues (for it will be correct for each speaker to say “what my interlocutor just said is true”).

Now here’s how the cloudy model solves this problem. The cloudy model allows that many of the propositions that a speaker correctly puts into play may in fact be false. This is because on the cloudy model, it need only be the case that *at least one* of the propositions in the cloud is true, for the assertion to be correct. Thus, two parties can correctly make superficially contradictory assertions, even if both assertions put into play many contents which are straightforwardly false. And thus, there is at least potential for this model to explain how interlocutors can evaluate *what the other has said* as false, if the right kind of story can be told about *which* content in the cloud is picked out by phrases such as ‘what my interlocutor just said’. Take our Brexit example. On this model, as long as Brexit was right relative to *the club’s* culture, this is sufficient for the assertion to be correct. But this leaves open that Brexit may have been wrong relative to, say, metropolitan London culture, and it thus leaves open that some other club member (who may bear greater allegiance to the latter, let’s call them Tony) can correctly say “what Boris just said is false”. For it may be that Tony can use ‘what Boris just said’ to pick out the content that Brexit was right relative to *metropolitan London culture*. After all, this is a content that Boris did indeed put into play via his assertion.¹⁰²

¹⁰² An initial complication is that this view appears to entail that for a given sentence, *S*, a given individual in a given context may be able to assert both *S* and $\neg S$. For example, Boris would seem able to assert both “Brexit was right” and “Brexit was wrong”, because Brexit is right relative to his club’s culture and wrong relative to metropolitan London culture, and both cultures are relevant at his context. However Suikkanen (p.c.) notes that this problem can be solved if the contextualist holds that the speaker’s pro-attitudes determine a necessary condition for correct assertion (that is, Jones must

And Suikkanen does provide a nice (pragmatic) story about how the referent of locutions such as “what Boris just said” gets determined. Like Finlay’s, Suikkanen’s pragmatic story starts with a plausible claim about the *conversational purpose* of moral exchanges, namely that the primary function of moral discourse is to influence behaviour in service of certain socially beneficial ends (such as cooperation and conflict resolution), and in particular, in the Brexit case, that Tony’s conversational end, in uttering the alethic evaluation, is to influence Boris in ways which serve these wider goals.

Now, how does this help us understand which of the contents, put forward by Boris, Tony should react to?¹⁰³ Suikkanen offers three plausible criteria: first, the content should be one determined by a group such that Boris shares the core concerns of this group’s code; second, the content should be one determined by a group such that Boris’ personal identity is bound up in membership of this group; and third, the content should be determined by a group which is sufficiently tight-knit to be able to exert social pressure on Boris, and to be able to effectively take sides for or against Boris in case of a dispute. Why are these criteria plausible? They are plausible because reacting to contents determined by groups which did *not* meet them would be unlikely to influence Boris’ behaviour (compare: “all your close friends think that what you did is wrong”, vs “everyone born at 4am on a Monday thinks that what you did is wrong”).

While these pragmatic principles are natural, given the cloudy model of assertion, the cloudy model itself may seem somewhat ad hoc. But the cloudy model is not without independent motivation. Suikkanen imports the proposal from the literature on epistemic modals, where content relativism faces similar problems. (Suikkanen credits the proposal, in particular, to von Finkel and Gillies (2011).) And versions of the cloudy approach have also been defended in the area of run of the mill vagueness (MacFarlane, 2020). Part of von Finkel and Gillies’

approve of Brexit in order to assert “Brexit was right”, and thus cannot assert “Brexit was wrong”). Assume the contextualist endorses this move.

¹⁰³ The reason that this is a pragmatic and not a semantic question is that the semantics of “what Jones just said” plausibly underdetermines a particular referent, given the assumption that the content of Jones’ assertion is a cloud of propositions.

motivation is a desire to avoid relativism for epistemic modals, on theoretical grounds. We will set this aside. But they also offer the following story. Suppose you are at a conference and a stranger asks you where you are from. Plausibly, the context leaves open a number of interpretations of this question. The interpretations “where is your home institution”, “where is your visiting appointment”, “where did you get your PhD”¹⁰⁴, “where did you grow up”, and so on, may all be equally legitimate. It is not, therefore, obvious how to understand the content of this question, at least if we refuse to let go of the assumption that every utterance has a particular, determinate, content. And yet, the question is clearly meaningful, and clearly achieves some form of communicative success. The cloudy model purports to offer a solution to this independent puzzle. And thus, to the extent that it solves the puzzle well, it is independently motivated. If we do accept that it is well-motivated in the case of run-of-the-mill contextual underspecification, then the appeal of plundering the model for the moral case will be obvious.

So this is one way in which the cloudy model is independently motivated. But there is another way, namely that it offers to solve another puzzle, although this time one that is more local. The puzzle is that if we endorse the view that moral truth is relative to a social group’s standards, then whether we are relativists or contextualists, there is a worry that this will undermine the determinacy of vast swathes of moral discourse. For societies and cultures are not in general sharply demarcated, and it is thus far from obvious, in any particular case, which of the many relevant ways of carving up the social milieu at a context is relevant for determining content (on contextualism) or truth value (on relativism). The cloudy view offers a neat solution to this puzzle, by simply suggesting that *all* the relevant groups determine a content which is put into play, and that an assertion is correct just in case *at least one* of the contents in the cloud can itself be correctly asserted. (I will say more about how truth relativism solves this puzzle in chapter 5, section 2).

3.1 Response to Suikkanen

¹⁰⁴ The example is due to von Finkel & Gillies.

While the cloudy model may be independently motivated, this is not to say that it is independently uncontroversial. In fact, it faces a number of independent objections (MacFarlane, 2020). However, I want to simply grant the viability of the framework. For even if we do so, Suikkanen's response still fails to explain the relevant data as effectively as relativism.

There are two reasons for this. First, Suikkanen's proposal fails to predict the right results even in some everyday, terrestrial, contexts. But second, in radically cross-cultural cases (like *Hyperspace Bypass*) which are our most central interest (because they are most plausibly capable of being faultless) Suikkanen's response fails to even get off the ground.

First, notice that it is not even clear that Suikkanen's response predicts the right results in the Brexit case. For which content should Tony react to? Which content best satisfies the three conditions? Plausibly, it is the content determined by Boris' club's culture, rather than the content determined by metropolitan London culture. And, we have assumed, this content is objectively true (Brexit was right relative to the club's standards). But then, Suikkanen's view implies that the content that Tony should react to is one which is objectively true, despite the fact that Tony and Boris disagree in attitude. It thus licenses Tony in uttering, precisely what we wanted to avoid, sentences of the form:

(5) "Boris just said that Brexit was right. What Boris said is true. And indeed, I believe it too. But Boris and I disagree, morally, over Brexit".

Admittedly, this example is perhaps set up in a way which makes it unfriendly for Suikkanen's view. We could equally well construct a case in which his proposal is more likely to get the right results (for example, by saying that Boris identifies more closely with some anti-Brexit group). But the point is that the case is equally coherent (and indeed likely) in its current form. Thus, it should in general be relatively easy to generate cases which are unhappy for Suikkanen's proposal, even if it is also possible to generate cases wherein his proposal gets the right results.

Furthermore, in fairness, Suikkanen appears to intend a *satisficing* interpretation of the pragmatic requirements, rather than a maximising one. Thus, he could respond that Tony is not necessarily *required* to assert (5), but rather is merely permitted to assert (5). And since London metropolitan culture also satisfies (i) – (iii) to some extent, he would also be permitted to assert (6) “Jones said that Brexit was right. What Jones said is false, and I reject it. Thus we disagree.”

But there are two problems with this move: first, as long as Jess is *permitted* to assert (5), that seems bad enough. For (5) does not seem permissible, it seems incorrect. Second, it now seems, on the satisficing view, that Jess is committed to a contradiction, since the satisficing view entails that she can correctly assert *both* (5) and (6) (for *both* Metropolitan London culture, *and* the club’s culture, will satisfy (i) – (iii) to some degree). Thus the satisficing view is, arguably, even less plausible than the maximising view.¹⁰⁵

Moving on to the second problem for Suikkanen’s view, notice that in the case we initially wanted to explain, Hyperspace Bypass, it is unclear that there is *any* content which satisfies (i) – (iii), and which Arthur can correctly evaluate as false. After all, we can stipulate that all sub-cultures within Vogon society are in favour of demolishing the earth to make way for a hyperspace bypass. Thus, *every* content which satisfies (i) – (iii) will be straightforwardly true. As a result, Arthur will not be able to correctly assert (2), and may be committed to asserting (1). Thus, the cloudy approach helps *not at all* in explaining the data in cases like Hyperspace Bypass.

Suikkanen’s suggestion thus, while interesting and innovative, fails to explain the relevant data as well as relativism, which, in comparison, neatly explains *all* such cases, via its

¹⁰⁵ As in n.9 Suikkanen could here appeal to Jess’ attitudes to determine the content she can react to. That is, he could hold that the content she could react to must be one which satisfies (i) – (iii), and which is such that it matches up with her non-cognitive attitudes in the sense that she approves of the relevant action (event, character trait, etc.) if the content is true and disapproves if it is false. However, Suikkanen’s view would then look in danger of collapsing into a form of quasi-realism, for use of the ordinary truth predicate would then appear to be very closely tied to expression of non-cognitive attitudes.

commitment to univocality. We can thus conclude that Suikkanen's response does not save content relativism from its difficulties with alethic evaluations and belief ascriptions.

4. Conclusion

Finlay and Suikkanen's proposals are the most promising efforts, of which I am aware, to respond to the objection we have been pressing. I thus conclude, that since Finlay and Suikkanen's responses fail, content relativists have so far proved unable to explain the relevant data as well as relativists can. Of course, this does not rule out the possibility that some ingenious content relativist solution, which we cannot currently envisage, may present itself. However, there are reasons to be skeptical of this eventuality. To see why, note that since the problem is generated by the content relativist's commitment to the existence of relativised descriptive contents, the content relativist has only three possible options available to solve the problem. First, they might claim that assertions and beliefs *lack* relativised descriptive contents for belief reports and alethic evaluations to concern ("*subtract* descriptive contents"). Second, they might claim that alethic evaluations and belief reports pick out or are intended to pick out some relativised content *other* than the semantic value of the interlocutor's belief or assertion (*change* descriptive contents). Third, they might claim that there are *additional relativised* contents associated with the relevant beliefs and assertions, in the hope that some of these additional contents will have the right alethic properties (*add* descriptive contents).

The reason for skepticism about any future content relativist responses is that each of these options seems to face serious difficulties. The first option, of course, simply abandons content relativism, and would thus undermine content relativism if successful. The second option seems to inevitably involve either some ad hoc stipulation or some form of semantic error theory. For alethic evaluations and belief reports do not normally pick out contents other than the semantic values of the interlocutor's belief or assertion. And, inasmuch as they do, this should be cancellable in at least some contexts (such as the historians case). It is thus hard to see how the second option can be effected without at least disadvantaging content relativism relative to the attractively simple and non-ad-hoc explanation offered by relativism. Finally,

the third option simply seems unlikely to get the right predictions in *all* cases, even if it may in some. This is because adding additional contents, if done in a non ad-hoc way, seems antecedently equally likely to generate contents which have the “wrong” alethic properties as it is to generate contents which have the “right” ones. As we saw in the Brexit case, Suikkanen’s story about which additional contents get put into play and reacted to, although generating the right results in some cases, leads in other cases to a situation where the original problem resurfaces, because the contents determined, in that case, have the “wrong” alethic properties – that is, alethic properties which revive the original problems. It thus seems unlikely that any principled addition of extra contents into the mix will avoid the implication that (1) is correct in at least some cases.

I thus conclude not only that current content relativist offerings fail to solve the problem posed in this chapter, but also that there is reason for pessimism about the prospects of future content relativist solutions. Of course, this does not conclusively rule out the possibility that a successful content relativist response is out there somewhere, in logical space. However, it seems reasonable, given the available evidence to conclude that a relativist account of the data is and will remain preferable.

The upshot is that we can finally make good on the argument, initially advanced in chapter 2, that relativism gains support, by IBE, from experimental data concerning faultless disagreement. A key plank in this argument was that content relativist accounts of the Appearance of Genuine Disagreement, in such cases, are inferior to relativist accounts. And we have now seen why this is true. Although content relativists can explain why such cases seem to instantiate disagreement, via appeal to disagreement in attitude, this appeal clashes with their core semantic commitments. And the resulting implications undermine the content relativist’s account.

So far in this thesis, I have played offence, trying to show that there is positive reason to accept (a truth-relativist version of) metaethical relativism. But many of those who reject metaethical relativism do so less because they are skeptical of the arguments in favour of it, and more because they believe it faces fatal objections. One important objection is that relativism is

unable to account for disagreement. In the course of the positive discussion, I have shown, preemptively, that this objection fails. However, other objections remain. Therefore, I now want to play some defence: to say a bit about why remaining objections are less problematic than is often assumed.

CHAPTER 5

ABSTRACT: This chapter defends moral relativism against two substantive metaethical objections: first, that it is objectionably subjectivistic and second that the societal form of relativism (assumed in chapter 2) leads to an unacceptable level of indeterminacy. Discussion of these objections provides opportunity to flesh out the semantic framework in metaethically relevant ways. The chapter concludes by comparing relativism with other popular metaethical theories.

This thesis has so far focused on making a positive case for truth relativism. However, many philosophers hold that relativism (of any form) faces fatal objections. Although these objections are not extensively discussed in the contemporary literature¹⁰⁶, they seem to explain much of moral relativism's lack of popularity among contemporary metaethicists. Thus, even if the case made so far is accepted, many readers may feel that the evidence in favour of truth relativism, deriving from the arguments advanced so far, is outweighed by evidence against it, deriving from the relevant objections.

To close out the thesis, then, I want to play some defence, and explain why truth relativism has the resources to meet some of these common objections. First, in section 1, I will discuss the worry that truth relativism is *objectionably subjectivistic* in its implications. In particular, some worry that truth relativism is an "anything goes" view of morality, where any moral code is just as correct as any other and where, and as a result, we are unable to criticise those who endorse moral codes with which we disagree. I will explore a number of ways of developing this objection, and detail the resources available to truth relativists in meeting it. I

¹⁰⁶ Although see the useful exchange between (Finlay, 2008), (Joyce, 2011), (Finlay, 2011). Also of recent interest are (Prinz, 2007) (Street, 2016) (Finlay, 2014) Francen (2016) Francen (2007) (Dyke, 2020) (Korman & Locke, 2020) (Woods, 2020) (Einheuser, 2006).

will show that this objection has far from the decisive weight that it is often afforded. Indeed, properly understood, the truth relativist's position, on these issues, is a plausible one.

Section 2 will begin by recapping the experimental evidence which motivates endorsing a societal form of relativism, rather than an individualistic one. It will then detail two additional rationales for doing so. Many have argued that societal relativism faces distinct challenges of its own, challenges which are not faced by individualistic relativism. For example, it is sometimes pointed out that societies are not sharply delineated¹⁰⁷. A given individual may be a member of many societies, and it may therefore be a vague matter which social or cultural standard determines a given individual's context of assessment. Consequently, some worry, the determinate correctness and/or incorrectness of our moral beliefs appears threatened. In section 2, I will show that relativists have the resource to meet this objection, too.

Section 1 will assume an individualistic form of relativism, even though the ultimate aim is to defend societal relativism. This is because things are made simpler if the worry about objectionable subjectivism is discussed prior to introducing the details of the societal theory. Furthermore, from a dialectical point of view, the worry from objectionable subjectivism should be if anything *more* acute for the individualist version of the view, than it is for the societal version. So this way of setting things up should not do the truth relativist any illicit dialectical favours.

This thesis has so far retained a fair amount of distance from substantive metaethical questions often associated with moral relativism: questions about what degree of moral objectivity can be admitted under a relativist framework, about how far moral judgements are grounded in non-cognitive attitudes, about how far evolutionary psychology should inform metaethical theory, and about the extent to which moral truth should be understood as grounded in sociological facts. Responding to the objections described above will provide natural opportunities to flesh out the theory defended here along these parameters, and this will also set us up to conclude by comparing some of the theory's advantages and disadvantages relative to other substantive metaethical theories. Since, furthermore, the focus of the chapter

¹⁰⁷ (Suikkanen, 2019) discusses this objection.

is broadly metaethical, rather than narrowly semantic, I will use the generic term ‘relativism’ to refer to the view I defend – a truth relativist semantic framework will nevertheless be assumed.

1. Objectionable subjectivism

According to an influential objection, relativism is an “anything goes” view of morality. Relativism entails that moral propositions are true relative to moral standards. But moral standards, it might be objected, can themselves admit of moral evaluation. Thinkers can have standards which are correct, plausible, wrong, abhorrent, or bizarre. But according to relativism *all it takes* for a belief to be correct is for it to be true relative to the standards relevant at a believer’s context of assessment – even if these standards are wrong, abhorrent or bizarre. That, it might be objected, does not seem right.¹⁰⁸

After all, it does not seem that morality is an “anything goes” type of discourse. Rather, morality seems to be something that we can coherently reason about. It seems that this reasoning can result in right or wrong answers, and it seems that although this reasoning may be successful, it may also be unsuccessful, and incorrect moral beliefs may result. If relativism entails that “anything goes” when it comes to moral belief, it will fail to explain this salient feature of moral discourse.

One way to develop this worry would be to claim that relativism makes moral judges *infallible*. Since it obviously is coherent to wonder if one’s standards are correct, this would be a serious flaw. However, even naïve subjectivist views do not quite have this upshot. For consider the naïve view that a thinker’s standards are determined by their own actual present states of approval and disapproval. For example, on this view, an individual’s belief that phi-ing is wrong is correct if and only if the relevant individual disapproves of phi-ing. Notice, though, that even this crude and simplistic form of subjectivism does not entail that moral thinkers are *infallible*. For thinkers are in general imperfect judges of their own psychologies

¹⁰⁸ (Huemer, 2005) (Shafer-Landau, 2003).

(Williamson, 2000), and thus thinkers are plausibly imperfect judges of their own affective states in particular.¹⁰⁹ Thus, even on this view, it is possible for thinkers to be mistaken about whether their own moral judgements are correct.

Instead, the worry should be that even if relativism is compatible with *some* degree of fallibility, it is an objectionably subjectivistic view to the extent that the relativist's notion of fallibility is *insufficiently robust*. And indeed, if *all the relativist could say*, by way of giving content to this notion, was that one can be mistaken about what one's own standards entail, this would clearly be an objectionably subjectivistic position. Even if an individual is correct about what their own standards entail, it is still surely possible for this individual to be morally mistaken. To connect the worry up with some intuitive data, it might be pointed out that relativism appears to have some implausible implications vis a vis the correctness of certain repugnant moral beliefs. For example, it might seem to imply that many of Hitler's moral beliefs were correct, since many of the things Hitler believed were true relative to Hitler's standards.

But although this kind of objection is influential, the threat that it poses to sophisticated forms of relativism is often overestimated/ As I will show, sophisticated forms of relativism are compatible with a robust form of, what I would call, pseudo-objectivity. And, sophisticated forms of relativism have plausible implications about the correctness of particular moral beliefs.

In particular, (1.1) there are two natural idealising moves that relativists can make in this context. First, they can appeal to the notion of reflective equilibrium in their explication of the standards determined at a context of assessment. Second, they can appeal to the claim that there are substantive conceptual moral truths. These moves allow the relativist to deny that Hitler's moral beliefs, *inter alia*, were correct.

¹⁰⁹ Note that the point would still hold on the even more naïve view that an individual's standards are determined by their own moral beliefs. For since thinkers are fallible with respect to their mental states in general, thinkers are presumably also fallible with respect to their moral beliefs.

Still, (1.2) anti-relativists may seek to press the objection against relativism by appealing to hypothetically idealised versions of actual repugnant moral thinkers, such as an idealised Hypothetical Hitler. However, relativists can raise methodological objections against this move. Appealing to the quasi-realist aspects of this theory, they can point out that in assessing a Hypothetical Hitler, it is difficult to disentangle our first-order moral intuitions (in respect of which, relativism allows us to say everything that objectivism does) from our intuitions about the non-moral norms governing Hypothetical Hitler's beliefs (wherein realism and relativism diverge). When we, furthermore, consider cases which avoid this methodological difficulty, the relativist can argue that the force of this objection is much diminished.

Finally, (1.3) relativists can call our attention once again to the experimental data discussed in section 2. Although, they should agree, it is not plausible that Hitler's moral beliefs were correct, we should not be misled by this into thinking that *no* moral beliefs which we disagree with could be correct. Rather, as we have seen, there is experimental evidence that some moral beliefs which we strongly disagree with are correct. The upshot is that not all subjectivism is objectionable. Rather, our aim should be to find the right balance. Only relativism can accomplish this. For only relativism is compatible with at least some degree of subjectivism.

These points, I will argue, provide a compelling defence against this version of the objection. However, (1.4) the critique might be renewed by appeal not to straightforward claims about the correctness of repugnant beliefs which we disagree with, but rather by appeal to *counterfactual* claims about the correctness of beliefs we ourselves would have, if our values were to suitably change. For example, it might be claimed that if our values were to change in such a way that we would endorse Nazism (in reflective equilibrium), it would then be the case that we could correctly believe exactly the same things that Hitler did. In response to this objection, I appeal to modified versions of the points discussed in (1.2) and (1.3). The quasi-realist aspects of the theory take much of the sting out of this objection. And on methodological grounds, the objection is dubious. I conclude that relativism, charitably understood, is not objectionably subjectivistic. I then turn (2) to consider societal relativism.

1.1 Idealising moves

As we saw in chapter 3, it is in general plausible that contexts of assessment should be determined in ways which idealise, to at least some extent, from a speaker or thinker's actual standards. For example, it is very plausible that an individual who dislikes a certain piece of art purely because they are jealous of the artist's success should not, thereby, be taken to believe correctly that the work of art is a bad one. Furthermore, not only is this independently plausible, it should be consistent with the relativist's core claim that moral disagreements can be faultless. For, obviously, the claim that there *can* be faultless disagreements does not imply that all disagreements are faultless. Relativism should, therefore, be in general consistent with at least some independently plausible claims about what constitutes a cognitive fault in any given domain. In terms of the machinery of the theory, relativists can account for such claims by incorporating relevant idealisations into the account of how contexts of assessment get associated with speakers and thinkers. In the case of the example just mentioned, they could say that a speaker/thinker's context of assessment is determined by the aesthetic standards they *would* have if they disregarded their jealousy of the artist's success (presumably, other conditions should be included as well).

What kind of idealisation is plausible in the moral case? In the literature on moral epistemology, realists and anti-realists alike agree that our best model of successful moral reasoning is the method of wide reflective equilibrium.¹¹⁰ This method involves broadly Bayesian revision of moral judgements (including moral intuitions), at various levels of generality, in order to achieve greater internal coherence (coherence between moral beliefs themselves) and external coherence (between moral and non-moral beliefs). Although it is not suggested that this is a process which can feasibly be completed, the degree to which it is successful can be characterised in terms of a thinker's distance from an ideal limiting state of reflective equilibrium, where their moral beliefs are fully internally coherent, and their moral beliefs fully cohere with a complete and accurate set of background non-moral beliefs.

¹¹⁰ (Rawls, 1951) (Cath, 2016) (Daniels, 1979) (DePaul, 1986).

Not all theorists accept that reflective equilibrium is a complete account of epistemic success in ethics. But almost all accept that it is a large part of the account¹¹¹. Thus in general, it should be a desideratum of metaethical theories that they provide some account of *why* this is the case – why it is that the more closely one approaches a state of reflective equilibrium the more epistemically laudable one's thinking becomes. Relativists can offer a straightforward explanation. They can claim that a thinker's context of assessment determines, rather than their actual standards, the standards that they would have on reaching a state of wide reflective equilibrium. Why, then, are a thinker's beliefs more justified as they approach states of reflective equilibrium? Because, on this view, it is trivial that the more closely one's beliefs resemble the beliefs that one would have in wide reflective equilibrium, the more likely one's beliefs are to be correct.

This move goes a long way to showing that relativism has no implausible implications vis a vis the correctness of actual moral beliefs. This is because in the standard kinds of cases which are thought to undermine relativism (such as that Hitler's moral beliefs) it is often implausible that the relevant beliefs would persist in wide reflective equilibrium. For example, Hitler's belief that it was right to exterminate Jews was based on a deep and broad network of false non-moral beliefs (such as that Jews were causally responsible for German defeat in World War I). In more general terms, although various kinds of prejudice seem to come naturally to humans, it is also true that the moral components of these prejudices are normally interwoven with false non-moral beliefs which would not survive in reflective equilibrium. While it is non-trivial that the purely moral elements of racist and bigoted belief systems would not survive in reflective equilibrium, once detached from their non-moral supports, it is nevertheless quite plausible that this is so. One reason for this is the fact that when such beliefs face off, in moral reasoning, against universalising principles¹¹² such as that people should be treated equally unless they are different in morally relevant ways, and when false non-moral beliefs are discarded, the universalising principles seem often to come out on top, for it seems that many people attach very high prior probabilities to such universalising principles.

¹¹¹ (Kelly & McGrath, 2010).

¹¹² Which are robustly cross-cultural – (Stanford, 2018).

It is therefore quite plausible, in the case of Nazi and more generally xenophobic or racist beliefs, that few if any actual individuals would maintain such beliefs in wide reflective equilibrium.¹¹³ Certainly, it is plausible that most early 20th century Germans would not remain committed to such beliefs in wide reflective equilibrium. Admittedly, this is not something we can know with certainty. And therefore, it might be insisted that the relativist is committed to the claim that we do not know with certainty that Hitler's moral beliefs were incorrect. This, it might be insisted, is implausible enough to undermine relativism.

But notice that there is dialectical pressure also on the *critic* of relativism to hold that Hitler's moral beliefs would not survive in reflective equilibrium. For recall the arguments of chapter 1. Very briefly, one conclusion that we reached, there, was that if (enough) actual moral disagreements would persist in wide reflective equilibrium, then this would undermine objectivism. For this would mean, if objectivism is true, that we would have compelling reason to suspend judgement about the propositions that the disagreements concern. Thus, if there were enough such propositions, a form of broad moral skepticism would turn out to follow from objectivism. Relativism, as we saw, avoids this problem, because such disagreements need not undermine knowledge if relativism is true.

As we saw in chapter 1, it is unclear how much actual disagreement would persist in wide reflective equilibrium. And therefore, it is unclear how much support relativism gains from the truth of this conditional claim. However, there is a trade-off here. If a critic wants to insist that Hitler's beliefs are correct, according to relativism, (because they want to insist that Hitler's beliefs would persist in reflective equilibrium) and this critic is an objectivist, then the arguments of chapter 1 would then entail that the critic must accept we do not know that the holocaust was wrong, that Nazism is evil, and so on, and thus, arguably, that we cannot ourselves assert or believe such things (because, the arguments of chapter 1 show that this is an upshot of the claim that Hitler's moral beliefs would persist in wide reflective equilibrium). This, though, is arguably *even more* unattractive than the claim that Hitler's beliefs were correct. For relativism would *at least* leave our engaged first-order judgements intact, even if

¹¹³ As we will see below, societal relativists can provide additional reasons to think that substantive moral disagreement would not persist in wide reflective equilibrium.

it were to turn out that it saddles us with some unattractive commitments at a metaethical level. Objectivism, instead, would rob us of both. For even if Nazism is wrong, and Hitler's beliefs are incorrect, it would turn out that these are not things we can know, believe or assert.

I have said that this is a problem for an objectivist critic, but the same dialectical issues may apply to the debate between relativists and quasi-realist expressivists. For the arguments of chapter 1 would seem equally sound given deflationary accounts of the relevant notions (such as knowledge, justification and truth). The basic dilemma is that it is hard to avoid skepticism, in the face of disagreement in wide reflective equilibrium, without endorsing some form of relative truth or correctness. Thus, inasmuch as expressivists want to avoid relative truth or correctness (as many do)¹¹⁴, it is unclear how they would avoid the same dilemma. Admittedly, expressivism is a vibrant research program, and there are many recent expressivist proposals on the table. I will say some more about these options below. Still, we can see that rather than providing any devastating objection to relativism, these objections place a burden on the *critic* of relativism to explain how the objection to relativism can be formulated in a way which does not saddle them with even more unattractive commitments than those which are putatively faced by relativism.

The appeal to reflective equilibrium also gives rise to an interesting question about how the notion of reflective equilibrium connects up with that of a moral standard. In chapter 3, we saw that at least in the case of aesthetics, standards are taken to be *sensibilities*: complex psychological dispositions have certain non-cognitive states in response to certain works of art. A similar view looks promising in the ethical case. It is attractive to relativise moral truth to standards understood as non-cognitive states, because relativising moral truth to standards understood as *cognitive* states would threaten circularity. After all, what set of *beliefs* could plausibly play the role of moral standards, without themselves being moral, or at least genuinely normative?¹¹⁵ But if standards are understood as sets of moral beliefs, they can hardly play a useful role in explicating the truth conditions of the contents of moral beliefs.

¹¹⁴ e.g. (Blackburn, 1998) (Schroeder, 2008).

¹¹⁵ One way to go here would be to construe an agent's standards in terms of their instrumental practical reasons (Smith, 1994) (Dyke, 2020) (Copp, 1995). I prefer not to go down this route, because it seems to me that morality and practical reason are distinct normative domains, which may be

If we go in this direction, however, it may seem less obvious how a standard can be said to be held in reflective equilibrium. After all, reflective equilibrium is a property of sets of beliefs, not a property of sets of non-cognitive attitudes. But here is what the relativist can say. On a plausible relativist view, moral judgements, while individuated as cognitive states, are systematically associated with non-cognitive states. And a thinker's non-cognitive moral psychology runs in a state of dynamic alignment with their cognitive moral psychology (that is, when a thinker's moral beliefs change, her associated set of non-cognitive states also change). Thus, it is perfectly coherent to speak of the sensibility that an individual would hold in wide reflective equilibrium. This is the complex set of non-cognitive states (understood as dispositions to have certain attitudes towards certain actions, individuals, institutions, etc.) which would be systematically associated (in virtue of contingent psychological laws) with the moral beliefs that the individual would hold in wide reflective equilibrium.

What non-cognitive states are these? Although there is a large debate about how characteristically moral non-cognitive states should be understood, which it is not possible to do justice to here, I favour an account of the kind defended by (Björnsson & McPherson, 2014). On this view, moral judgement is understood as a multiply realisable functional state, one key component of which is to dispose individuals to deploy social hostility against those who behave in proscribed ways, and against those who make sufficiently divergent moral judgements. This account connects up in interesting ways with the societal version of relativism I want to defend, and with empirical research on the evolutionary origins of moral discourse. (More on which later). For now, an individual's sensibility can be glossed as the set of Björnsson & McPherson-style functional states they would have in wide reflective equilibrium, which involves, *inter alia*, a set of dispositions to deploy social hostility against those who act in ways proscribed by the moral beliefs they would hold in wide reflective equilibrium.¹¹⁶

interrelated, but which cannot be reduced to each other. But this is obviously not an issue I can discuss in detail here.

¹¹⁶ Björnsson & McPherson explicitly anticipate a form of relativism which appeals to their version of non-cognitivism in determining the truth conditions of moral claims, this proposal is intended in that spirit.

Now, a critic of relativism might respond to this first idealising move by appealing to merely hypothetical cases, rather than actual ones. As we saw in chapter 1, merely possible cases of disagreement which would persist in wide reflective equilibrium need not undermine the epistemic status of our actual beliefs. Thus, a critic might claim, relativism is committed to the view that a hypothetical Hitler, who believes exactly as Hitler did, but whose beliefs are stipulated to be such that they are persistent in wide reflective equilibrium, would have correct moral beliefs. This, they might claim, is an objectionably subjectivistic consequence of relativism. This claim, furthermore, need not commit the critic to any form of skepticism, since it turns merely on the *mere possibility* of moral disagreement which would persist in wide reflective equilibrium (and as we saw in chapter 1, this possibility need not commit the critic to any form of skepticism).

But before I move on to confront this objection directly, I want to highlight another idealising move that the relativist can make. Although this move does not obviously help solve the problem of how to evaluate a hypothetical Hitler. It is an important addition to the theory, and now is as good a time to mention it as any.

The point is that, as Philippa Foot (1958) once noticed, it is independently plausible that there are cases in which we would intuitively view putatively moral beliefs as incorrect, even if they involved no error in reasoning, and even if not because we are committed to the existence of objective moral facts. For example, we would think it bizarre if we were to encounter thinkers who appeared to believe that it is seriously morally wrong to look at hedgehogs in the light of the moon, or run around trees right handed. These beliefs could be made intelligible if it were to emerge that the thinkers in question connected looking at hedgehogs in the light of the moon (via some strange background beliefs) with some recognizably moral consideration. For example, if these thinkers believed (for whatever reason) that looking at hedgehogs in the light of the moon would anger the gods and thereby cause natural disasters, this could make sense of their putatively moral belief that doing so is wrong. But in the absence of any such background belief, it would seem that a substantively moral belief of this kind could not be correct. Plausibly, even if morality is not objective, it cannot be the case that looking at

hedgehogs in the light of the moon is, inherently, seriously morally wrong. (Just as, even if etiquette is not objective, it cannot be the case that someone is rude if and only if they behave according to prevailing social conventions).

Foot's diagnosis of such cases is that there are at least minimal constraints on the application of moral concepts. It is often noted that a puzzling aspect of moral language is that it is less tightly constrained in its conditions of use than ordinary descriptive language. If someone did not agree that water is the wet, potable stuff that fills rivers and lakes, falls from the sky etc., we would assume that we were mistranslating this person's use of the word 'water'. In contrast, moral language seems to admit of much greater variation, without loss of univocality.¹¹⁷ We do not automatically assume that Hitler was not using the concept of moral permissibility, just because he applied it to racism and genocide. Foot's insight, however, is that there is a limit to this phenomenon. Although moral concepts seem to permit greater referential variation than ordinary descriptive contents, such as water, there is also referential variation that they do not permit. As Foot comments "If I say 'I am sitting on a pile of hay' and bring as evidence the fact that the object I am sitting on has four wooden legs and a hard wooden back, I shall hardly be described as thinking, even mistakenly, that I am sitting on a pile of hay; all I am doing is to use the *words* 'pile of hay'". Similarly, on Foot's view, someone who says that looking at hedgehogs in the light of the moon is seriously morally wrong may be using the words 'morally wrong', but they have not succeeded in genuinely judging that looking at hedgehogs in the light of the moon is seriously morally wrong.¹¹⁸

So this is an additional layer of (pseudo) objectivity that the relativist can admit. No matter what standards an individual would endorse in wide reflective equilibrium, an assertion of the sentence 'looking at hedgehogs in the light of the moon is, other things being equal, seriously morally wrong' could not be correct, because it is a conceptual truth that this sentence is false. This should be perfectly compatible with relativism, since it does not appeal to the existence of context-insensitive truth conditions for moral claims, and since the strategy seems, in general, perfectly consistent with an intuitive anti-realism (such as about etiquette).

¹¹⁷ C.f. (Horgan & Timmons, 1991) (Dowell, 2016).

¹¹⁸ C.f. (Cuneo & Shafer-Landau, 2014) (Evers & Streumer, 2016).

1.2 Hypothetical cases

Although this is an important point, it may not (to return to the main thread) make contact with the objection that relativism problematically entails that a *hypothetical* Hitler's moral beliefs were correct. Hitler's beliefs (and hypothetical versions thereof) are more plausibly thought of as substantively wrong, rather than bizarre or confused (in the way that the belief that looking at hedgehogs in the light of the moon is seriously morally wrong is bizarre or confused).

But in response, it is plausible that the appeal specifically to a hypothetical Hitler is methodologically suspect. And, when we reformulate the objection in a way which avoids methodological difficulties, we will see that it is considerably less compelling.

To see why the appeal to a hypothetical Hitler is methodologically suspect, notice first that the version of relativism defended here is to a large degree quasi-realistic. This is because the assessment sensitivity framework allows the relativist to make all and exactly the same *first-order* moral judgements as the objectivist does. Thus, relativism is consistent with condemning Hitler in the strongest possible first-order moral terms: thinking that his actions were wrong, evil, outrageous; and that there were weighty moral reasons to prevent him from acting as he did (as well as having all the associated non-cognitive attitudes). It is only when we step back, and take a metaethical vantage point, that the form of relativism defended diverges from objectivism. For it does so only in its predictions about which beliefs and assertions are permitted by the relevant non-moral norms (see chapter 2, section 1). As in paradigmatic cases of faultless moral disagreement, judging that a belief is free from fault (in the relevant non-moral sense) is perfectly consistent with rejecting these beliefs (chapter 3), and with any and all first-order judgements we might want to make about them (just as in paradigmatic cases of faultless disagreement, such as over whether some dish is tasty: judging that the dish is tasty is perfectly compatible with judging that someone else's judgement that it is not tasty is correct). This includes judgements about whether tolerance towards such judgements is

morally required. It is quite consistent with relativism that tolerance is not required, for tolerance may not be required by our standards.

Thus, notice that it is important, if we want to get clear on the costs and benefits of relativism, that we consider cases wherein we are confident in our ability to disentangle our substantively moral intuitions from our intuitions about the non-moral norms governing belief and assertion as such. And plausibly, cases involving Hitler, or other familiar moral villains of history, fail to meet this desideratum. This is because our moral judgements about the actual Hitler are so salient and important in our actual first-order moral discourse, that there is a danger that our disposition to avoid saying anything that might implicate non-acceptance of such judgements may go beyond the disposition that is strictly rational in purely cognitive terms. To borrow a point of Finlay's (see chapter 5) the social costs involved in indicating that one thinks that Hitler's actions were permissible are so high, that this may easily lead us to overcompensate, by adopting dispositions to deny not only that Hitler's actions were permissible, but also to deny any claim that might feasibly be mistaken for the claim that Hitler's actions were permissible (such as that Hypothetical Hitler's beliefs were in a certain sense correct) independently of the truth of that claim. Furthermore, it may be that an effective way of adopting such dispositions is to *internalise* them, so that we may not only be disposed to deny such claims, but may also genuinely find them counterintuitive.

This point does not, of course, provide any positive reason to think that it is true, as relativism predicts, that Hypothetical Hitler's beliefs were correct. But what it does do is give us reason for caution about the use of cases involving Hypothetical Hitler (and his ilk: Hypothetical Pol Pot, Hypothetical Saddam Hussein, etc.). Furthermore, if, when we consider structurally similar examples which are not methodologically problematic in this way, our anti-relativist intuitions tend to soften, this may provide some evidence that the kind of conflation just mentioned is indeed in play.

In this spirit, a more charitable and methodologically desirable test case for assessing the implications of relativism is one which we have already discussed (chapters 2, 4):

Hyperspace Bypass: The Vogons are a species of extraterrestrials. A representative vagon, Jeltz, believes that it would be morally permissible to demolish the earth to make way for a hyperspace bypass. Arthur Dent, a representative human, believes that demolishing the earth to make way for a hyperspace bypass would be wrong. Both hold these beliefs in states of wide reflective equilibrium.

The Vogons, here, are structurally similar to Hypothetical Hitler, in the sense that we can stipulate that their psychologies are such that relativism entails that their moral beliefs are, in the relevant sense, correct. Indeed, demolishing the earth to make way for a hyperspace bypass should be, if anything, *worse* from a substantive moral perspective than, say, Hitler's authorisation of the holocaust (since it involves killing roughly 1000x as many people). It would thus be puzzling, from an anti-relativist perspective, if we found it any less counterintuitive that the Vogons' beliefs are correct (that they are free from purely cognitive, that is non-moral, fault).

But many people, I suspect, will already find it significantly less counterintuitive that the Vogons' beliefs are correct. And let me do a bit more to pump this intuition. As Street (2016) argues (and this is a second methodological point), it is important when we are considering hypothetical agents whose psychologies differ from ours in radical ways that we do our imaginative due diligence – that is, that we make an effort to fully and vividly imagine the psychological makeup of the relevant agents. For otherwise, we may implicitly import assumptions about the psychologies of familiar human agents, assumptions which will corrupt the case. In that Street-influenced spirit, here is some more by way of colour to the Hyperspace Bypass case.

The case is drawn from Douglas Adams' comedy radio play and novel *The Hitch Hikers' Guide to the Galaxy*. In *The Hitch Hikers' Guide*, a key feature of Vagon social psychology is a commitment to the importance of (comedically Kafkaesque) bureaucratic process. According to the Guide itself:

Vogons are one of the most unpleasant races in the galaxy – not evil, but bad-tempered, bureaucratic, officious, and callous. They wouldn't even lift a finger to

save their own grandmothers from the ravenous Bugblatter Beast of Traal without orders, signed in triplicate, sent in, sent back, lost, found, queried, subjected to public inquiry, lost again, and finally buried in soft peat for three months and recycled as firelighters.

If we try, in all seriousness, to make this intelligible, we might offer the following story. Suppose the Vogons evolved in a context much like the human one, but with a few exceptions. Suppose one key exception was that the Vogons lived on a planet studded with special radioactive volcanos. Suppose that every evening, on this planet, at least one of these volcanoes would erupt. The resulting heat, mudslides, clouds of ash, and so on, were no threat to the safety of proto-Vogons, since they lived far enough away from the Volcanoes not to be affected. However, these special volcanoes also gave off a special kind of radiation, which quickly infused the atmosphere of the entire planet, and which immediately wiped the memory of any organism it came into contact with. As a result, for the majority of Vagon evolutionary history, Vogons were never able to remember what had happened at any time prior to the present day. Suppose, however, that in all other ways, the Vogons were extremely well suited to life on this planet, and had no evolutionary rivals. They thus became the dominant species on the planet. They did this by developing a strong disposition to record their discoveries and decisions in some objective form (perhaps starting out as cave paintings or lines in the sand), and by developing an equally strong disposition to take actions only if they had been extensively confirmed by such systems of recordings. Over time, we can imagine, the Vogons found ways to protect themselves from the effect of the volcanoes. But via the twists and turns of cultural development, we can imagine this deep psychological disposition ultimately manifesting itself, in modern Vogons, as a strong disposition to perform actions only if these actions have been ratified by systems of record-keeping and procedural approval, which would strike us as absurdly bureaucratic.¹¹⁹

Turning to the case of Hyperspace Bypass, we can imagine that the Vogons' insistence that demolishing the earth is permissible is grounded in the fact that the proper paperwork (to object to the demolition) was not filed in time at the intergalactic planning office. Since, for the Vogons, the ultimate principle of practical reasoning is a bureaucratic one, we might

¹¹⁹ This story is similar to one outlined in (Street, 2016).

imagine that this is a conversation-stopping consideration as to whether an action is morally permissible.

This feature of Vagon psychology is vestigial. But this does not distinguish it from many important principles of human moral psychology. For example, many ordinary humans think (with very high confidence) that incest is morally wrong, even if both parties willingly and informedly consent¹²⁰, and even if effective birth control is used. Similarly, many human beings think that it is permissible to aid one's kin, even if doing so comes at a huge opportunity cost in the form of aiding more needy non-kin. For example, many people in the U.K. spend hundreds of thousands of pounds putting their children through fee-paying education, when this money could plausibly save many lives if donated to effective charities. Similarly, people typically think it is not only morally permissible, but morally *required* to use their networks to assist their children in securing internships, jobs, university places, etc. even though competition for these places is zero-sum, and the consequences for individuals of securing them (or failing to do so) are great. These judgements, too, are plausibly vestigial. There is presumably some evolutionary story to be told about why many humans make such judgements. But we might hesitate to say (even if we disagree) that they are incorrect, at least if we assume that they would persist in wide reflective equilibrium.

1.3 The experimental data again

Finally, recall that as much as I have attempted to argue that the implications of relativism are intuitive, it is not obvious that the burden is on the relativist to convince any and all critics to share these intuitions. For recall (from chapter 2) that there is experimental evidence that ordinary speakers tend to judge that there can be faultless moral disagreements in cases much like Hyperspace Bypass. Thus, at least insofar as we are interested in giving a theory which is accurate as a descriptive semantics for ordinary moral discourse, there is independent reason to accept that this theory should have at least some "subjectivistic" consequences. Exactly where to draw the line is an interesting question. But the assumption should not be that *any* form of subjectivistic implication is a theoretical drawback. Rather, if anything, theories which

¹²⁰ (Haidt, 2012).

rule out *any* subjectivist consequences (such as realism, error theory, and some forms of expressivism) are the ones which thereby lose plausibility.

1.4 Counterfactual Conditionals

In response to all this, the assault on relativism might be renewed by appealing to a slight modification of the objection. Instead of focusing on claims about whether Vogon Jeltz' beliefs are correct, the critic could focus on claims about whether our own beliefs would be correct, if we were to come to have a psychological profile which would allow us to endorse Vogon values in reflective equilibrium. In particular, they could claim that relativists are committed to:

Counterfactual: if our fundamental values were to change in suitable ways, so that we would have relevant pro-attitudes (after idealisation) towards demolishing the earth to make way for a hyperspace bypass, it would be correct for us to believe that destroying the earth to make way for a hyperspace bypass is morally permissible.

I want to make three points in response to this version of the objection, each of which adapts one of the points made above, to this context. First, I want to appeal to the experimental data. Second, I want to appeal to the importance of properly and vividly imagining the relevant psychological profile. Third, I want to appeal to the quasi-realist elements of the relativist theory.

First, although it is less obviously so, Counterfactual should also be relatively straightforwardly an upshot of the experimental data. If it is possible for the Vogons' moral beliefs to be correct, while occupying the same world as ours, then it is hard to see how there is not also some scenario in which we could correctly believe the same things. For if correctness is not tied to bare species membership (as seems unlikely), then the possibility of correct Vogon beliefs must be underwritten by some aspect of Vogon psychology or social psychology. And since psychological states are plausibly multiply realisable, it seems that whatever feature of Vogon psychology is responsible for the correctness of their beliefs must also, in principle, be capable of rendering beliefs that we ourselves could have, in the same

propositions, correct. Thus, it is hard to see how one could coherently accept that Hyperspace Bypass instantiates a faultless disagreement, without also accepting Counterfactual.

Second, we should again notice that there is a first-order version of Counterfactual which relativists can deny. For notice that relativism does not entail:

*Counterfactual**: if our fundamental values were to change in suitable ways, so that we would have relevant pro-attitudes (after idealisation) towards demolishing the earth to make way for a hyperspace bypass, demolishing the earth to make way for a hyperspace bypass would be morally permissible.

Here, it is worth repeating the point (made above) that there is a danger, when we assess examples of this kind, that our substantive *moral* judgements may be difficult to disentangle from our judgements about the non-moral norms governing belief and assertion (and the point made above is here particularly acute, since Counterfactual is a claim about the normative status of *our own* beliefs). There is thus a possibility that at least part of any hesitation we may feel towards accepting Counterfactual may really derive from an (internalised) desire to avoid accepting Counterfactual*. And inasmuch as this is so, the relativist can happily remind us that rejection of Counterfactual* is entirely compatible with their view.

This is because Counterfactual* is a first-order claim. It claims that, under conditions where our attitudes have changed in suitable ways, demolishing the earth to make way for a hyperspace bypass would be permissible. It is thus of a piece with claims like 'infidelity is permissible if no one ever finds out' or 'it is permissible to strike preemptively if one knows one is about to be attacked'. These claims are obviously first-order claims, and can thus be correctly believed only if they are true relative to the believer's standards (Korman & Locke, 2020). No relativist with whom I am acquainted, however, has standards relative to which Counterfactual* comes out true. So every relativist with whom I am acquainted can correctly reject Counterfactual*.

If doubts remain as to whether Counterfactual* is a first-order claim, try thinking of it in this way. Compare the present objection with the following toy argument against ordinary contextualism about indexicals, like 'here' (Einheuser, 2006). We can imagine an argument to the effect that, since contextualism entails that the referent of 'here' varies with the context of use, contextualism entails that following is true:

SUN: If I had said this in Rio, then it would now be sunny here.

Suppose I utter this sentence on December 1st in Edinburgh. Intuitively, this sentence would then be false (for even if I had been in Rio, it would still have been raining in Edinburgh). So if contextualism predicts that this sentence is true, that would be bad news for contextualism.

But it is pretty uncontroversial that contextualism gives the right semantics for 'here'. So where does this toy argument go wrong? The key point is that the context of use is not shifted by uttering the antecedent. The context of use for the whole sentence is the context in which I actually utter it (in Edinburgh), hence the whole sentence is false.

If we are relativists, we can make a similar point. Conjecturing that we are in some other context of assessment doesn't shift the context of assessment we are actually in, in the same way that conjecturing that we are in another context of use doesn't shift the context of use of the utterance of the conjecture. Or at least, there is no reason to assume that things are any different in the assessment sensitive framework. Thus the burden of proof is shifted onto the critic of relativism. Absent some reason to think that things are different in the assessment sensitive framework, it is reasonable to assume that the context of assessment is not shifted by the antecedent of a counterfactual conditional. And given this assumption, Counterfactual* will come out false on relativism.

Michael Huemer (2005) anticipates this response, and replies that converting the offending sentences to the indicative mood renews the threat to (at least contextualist forms of) relativism. Consider, (Huemer's example):

MINNEAPOLIS: If I *have* flown to Minneapolis (such that I am now in Minneapolis), then Minneapolis is here.

To Heumer, MINNEAPOLIS seems true. If this is right, then this would seem problematic for relativists, because we could easily formulate an indicative version of Counterfactual*.

But is MINNEAPOLIS true? Although it might initially seem true, there is reason to doubt that it is. Compare:

SUN*: If I have said this in Rio, then it is now sunny here.

Again, we should assume that I am uttering SUN* on the 1st of December in Edinburgh. Intuitively, SUN* is false. All it says is that if I have (at some past time) uttered the sentence 'SUN' in Rio, then it is now sunny here. But this is *obviously* false. The fact that I have uttered SUN* in Rio at some time in the past (when it may have been sunny there) obviously does not imply that it is sunny here, now, in Edinburgh.

Perhaps the reason why MINNEAPOLIS initially seems true is the parenthetical clause '(such that I am now in Minneapolis)'. This may mislead us into thinking that the speaker is actually in Minneapolis. But this would be to set up the case incorrectly. For presently, we are thinking about people who have non-Vogon standards uttering a sentence which conjectures that they have Vogon standards. So similarly, we should imagine someone who is not actually in Minneapolis conjecturing that they have flown to Minneapolis. But again, it is *obviously* false that if I have flown to Minneapolis then Minneapolis is here. (The reader can imagine that I, the author, have indeed flown to Minneapolis (perhaps, many years ago), obviously – this does not mean that Minneapolis is here, where I am writing this chapter. Edinburgh is where I am writing this chapter.

Perhaps a better test then (given the past-tense reading of 'have') is this:

SUN': If I am saying this in Rio, then it is now sunny here.

This is a closer comparison to the indicative moral conditionals that Huemer thinks renew the threat against relativism. But again, if we imagine my saying this in Edinburgh on December 1st, it seems clearly false (if somewhat infelicitous). It therefore seems safe to conclude that these conditionals do not renew the threat to relativism, unless critics can show that things are different in the assessment sensitive framework.

So this is why relativists are not committed to Counterfactual*. I have conceded, however, that they are committed to Counterfactual. And I want to make one last point about why this commitment is not problematic.

Above, we discussed Street's insistence that, when considering objections arising from counterfactual scenarios, we do our imaginative due diligence – we fully and vividly imagine the psychological profiles of the relevant thinkers. But notice that what this involves, if we are evaluating Counterfactual, differs in important ways from what it involved above. For in this case, what we are asked to imagine is not some strange extraterrestrial with a radically different psychology (and radically different values). Instead, we are asked to imagine *ourselves* having a radically different psychology and, in particular, radically different values. However, it is not clear that there is *any* story we can tell which will allow us to fully and vividly imagine this. Our values are so deeply embedded in our conceptions of ourselves, that it is tempting to think that this conjecture is not even fully coherent (if my values were to change this radically – would I still be me?) But the relativist need not commit themselves to this (and it may not be strictly speaking true, according to whatever is the correct theory of personal identity). All they need hold is that, plausibly, it is extremely difficult, if not impossible, to fully and vividly imagine ourselves having radically different values. Thus, it is probable that in evaluating Counterfactual, we cannot help but implicitly assume that the relevant thinkers (that is, ourselves) have something like our actual values. In this respect, relativism (given plausible background assumptions) actually *predicts* that Counterfactual should seem less plausible than parallel claims about the Vogons' moral beliefs.

2. Going social

In discussing the worry about objectionable subjectivism, we have so far assumed an individualistic version of relativism, according to which moral propositions are true or false relative to standards held by individuals.

Whilst we have seen that individualistic relativism has considerable resources available with which to meet this objection, there is one further move that the relativist can make in response (2.1). This is to “go social” – to move to a version of relativism which holds that moral propositions are true or false relative to standards endorsed by a society. This move, as we will see, can be motivated not only by the dialectical consideration that it is more robust in the face of the worry about objectionable subjectivism, but also by the experimental data, and by our best understanding of the evolutionary genealogy of moral discourse (2.2).

But societal relativism also faces objections which are not faced by individualistic relativism (2.3). Most prominent, perhaps, is the worry that ‘society’ is an unhelpful notion in this context, for societies overlap, and are not sharply bounded. Thus, depending on how we define ‘society’ it may be that every individual belongs to a very large number of societies (a workplace, a university, a sports club, a family, etc.). Which society, then, determines the correctness of their moral beliefs and assertions?¹²¹

This objection does not, in my view, undermine societal relativism. We can understand the notion of ‘society’ in a theoretically useful way, which preserves sufficient determinacy of correctness and incorrectness of moral judgements. We can therefore, happily, hold on to societal relativism’s gains in independent plausibility, and robustness against the worry from objectionable subjectivism.

2.1 How going societal assists with the problem of objectionable subjectivity

¹²¹ C.f. (Suikkanen, 2019)

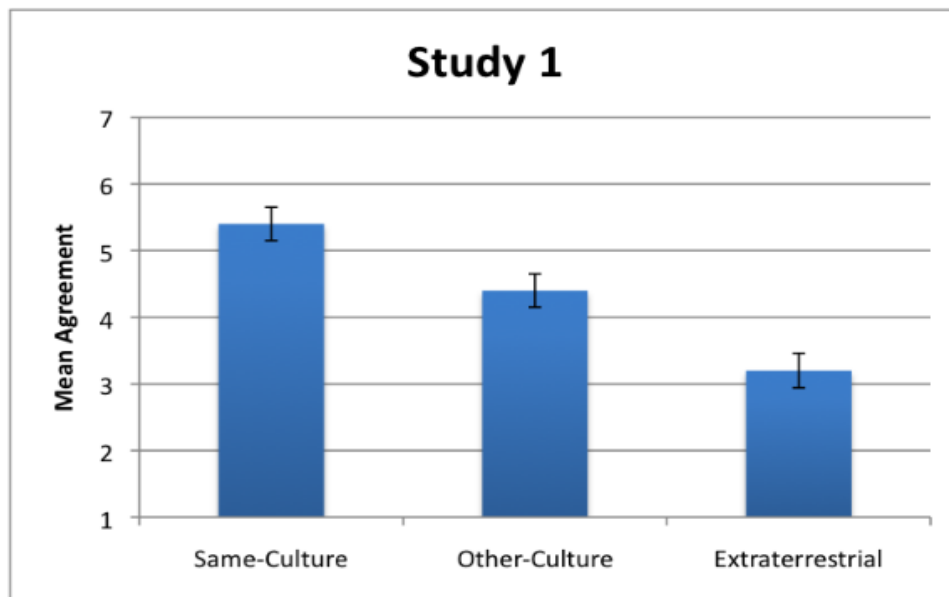
How does going social help with the problems discussed above? It helps because at least as far as actual judgements are concerned, the societal view provides an additional layer of pseudo-objectivity. This is because it is less likely that whole societies would endorse moral beliefs that we strongly disagree with, than it is that individuals would endorse such beliefs.

This is because it is unlikely that societies will endorse moral codes that are not even minimally socially viable, in the sense that these codes would directly vitiate this society's collective aims (Copp, 1995). For example, even if *ceteris paribus* prohibitions on theft, murder, etc. are not determined purely by the contents of moral concepts, it is plausible that no society could survive without such prohibitions. It is less obvious, in contrast, that we should expect no individuals to reject such prohibitions. Perhaps there are individuals whose values have become so warped that they might reject such principles, even in wide reflective equilibrium. If so, individual relativists would have to accept that their beliefs are correct. In contrast, societal relativists need not accept this, for it seems less likely that any whole society would accept such claims. Societal relativism is thus in a better position to explain the intuitive judgement (if it is intuitive) that no actual beliefs, of this form, are correct.

This admittedly is not to say that such societies are conceptually impossible. Such societies might not survive for very long, but they could conceivably exist momentarily. Nevertheless, the societal view, in this way, adds an additional layer of objectivity inasmuch as we have actual disagreements in mind.

2.2 Why going societal is independently motivated

Furthermore, going social is not merely an ad hoc move, with the aim of beefing up the objectivity in relativism. Rather, it is independently motivated by the experimental data. For as we saw in chapter 2, experimental evidence suggests that ordinary speakers' beliefs about the possibility of faultless moral disagreement are sensitive to the extent of cultural distance between parties to the disagreement. To briefly recap, in those experiments, participants were presented with a vignette involving a moral disagreement, and asked whether at least one party must be incorrect. A visual representation of the results:



What this shows is that participants in the experiments detailed in chapter 2 tended to view faultless moral disagreement as *impossible* in contexts where the parties to the disagreement were members of the same culture, but that this conviction tended to dim in cases of greater cultural separation. When the disagreement involved merely terrestrial cultural separation (i.e. when one party was a liberal westerner, and another was a member of an isolated tribe), participants tended to suspend judgement about whether it could be faultless. And when the disagreement involved extreme hypothetical cultural separation (i.e. when one party was a liberal westerner, while the other was a member of an extraterrestrial species with a radically different psychology) participants tended to endorse the possibility of faultless disagreement.

As we noted in chapter 2, these results, interestingly, *undermine* individualistic versions of relativism, even while supporting societal relativism. This is because individualistic versions of relativism straightforwardly entail that intra-societal moral disagreements could be faultless (just as societal versions entail that inter-societal disagreements could be). Only societal relativism, therefore, is consistent with the pattern of results recorded.

This claim might be questioned. It might be argued that if individuals are sufficiently similar in the standards they would have after reflective equilibrium, this would rule out intra-cultural faultless disagreement, consistent with individual relativism. However, even if all

actual individuals would converge in reflective equilibrium, this would not render intra-cultural disagreement *impossible* (all it would take would be for a mad scientist to put a sensibility-transforming pill in some individual's breakfast). Since the experimental materials are formulated in explicitly modal terms, this possibility undermines individualistic relativism.

So this is one way in which going social is independently motivated. But there is another way. Going social derives support from our best understanding of the evolutionary origins of morality.

To my knowledge, the most up-to-date rendering of this understanding is Stanford's (2018), which integrates a number of strands of empirical evidence. First, it notes the suggestion (due to (DeScioli and Kurzban, 2009)) that a central problem which our moral psychology seeks to resolve is the need for bystanders to take sides in cases of intra-group conflict. Bystanders, importantly, must normally all take the same side in such conflicts, for individual conflicts may otherwise easily escalate. But in addition, bystanders' dispositions to take sides must be independent of the identities of the particular individuals involved, for otherwise the power-advantages involved in having the rest of the group take a particular individual's side would allow particular individuals to achieve despotic status. Developing a practice of moral condemnation, objectively determined in relation to a cultural store of moral norms, allows groups to solve this problem.

Second, Stanford draws attention to a key challenge to the evolution of any altruistic system of motivation, namely that altruistic individuals (who provide benefits to others at some cost to themselves) are always at risk of being outcompeted or invaded by individuals who accept altruism from others, but do not behave altruistically themselves.¹²² According to Stanford, what is most fundamentally required for altruists to be protected against such competition is a way for altruists to ensure that they co-operate only with other altruists (that is, that individuals benefit others at cost to themselves only if they can expect those they benefit to respond in kind). The importance of this is especially pronounced, according to Stanford,

¹²² C.f. (Joyce, 2007).

given the extent to which human dispositions to co-operate are spontaneous and plastic (in other words, the extent to which human beings will automatically seek to cooperate, and will do so in creative ways, to solve novel problems). These spontaneous and plastic cooperation behaviours are likely a large part of what allowed early hominin species to outcompete other primates, and quickly radiate into new habitats. However, they also make humans much more vulnerable, than other primates, to invasion and outcompetition by non-altruistic individuals.

As Stanford points out, not only is it likely, from a game-theoretic perspective, that early humans would have needed *some* mechanism for achieving this (that is, allowing altruistic individuals to confine their altruistic behaviour to other altruists), there is also evidence that our contemporary moral psychology plays exactly this role. In particular, there is evidence (Skitka et al., 2005) that the desirability of social partners for cooperation is sensitive to substantive moral agreement between the co-operators. In other words, we are less likely to cooperate with people who do not share our moral beliefs. In particular, Skitka et al. found that awareness of moral disagreement made groups less likely to effectively cooperate to resolve disagreements, even when these disagreements were not themselves moral in character.¹²³

Stanford's key insight, finally, is that an *objective* moral psychology (in the sense of one which is authority-independent, and generalisable), is uniquely suited to solve this problem, by playing this role. For an objective psychology is uniquely suited to maintain a kind of dynamic alignment between one's altruistic motivations, and one's dispositions to exclude those who do not share these motivations. By *automatically* updating expectations about the behaviour of others, in line with one's own altruistic motivations, individuals are able to both maintain spontaneous and plastic moral dispositions, and, via a single cognitive mechanism, protect themselves from exploitation by those who lack such dispositions.

Now, there is no straightforward move from this account to any particular metaethical theory. For this is an empirical account of moral psychology, not a philosophically controversial

¹²³ (Bjornsson & McPherson, 2014) bears a number of striking resemblances to (Stanford, 2018), including emphasis on the convergence-seeking function of moral discourse.

account of moral semantics or metaphysics. All plausible metaethical theories are therefore strictly consistent with Stanford's account. However, some metaethical theories may fit better with Stanford's account than others. And in particular, it is quite clear that societal relativism fits better with this account than individualistic relativism does. For individualistic relativism would seem to undermine precisely the function that, on this account, morality essentially serves, namely to protect altruistic communities from invasion by defectors and free riders. This is because inasmuch as faultless intra-societal disagreements are possible, there would be no normative pressure on members of the same society (that is, individuals between whom cooperation is feasible) to converge in their moral beliefs, and this would be inefficient from the point of view of cooperation, given the account just outlined. Note that this is true *even if*, as a matter of contingent fact, all actual individuals would converge in reflective equilibrium. For the key evolutionary point is that as long as groups are *vulnerable* to such invasion, we should expect it to occur, because invader traits would be fitness enhancing. Admittedly, this account is also consistent with realism. However, full objectivity would arguably be superfluous to this account, since full objectivity applies in contexts where social interaction is not feasible.

2.3 *Objections to going societal*

Inasmuch as the societal relativist wants to claim that the societal theory is better supported by this evidence, it is a key question what exactly counts as a society in the sense relevant to the explication of the theory. In other words, exactly what kind of social unit is such that its standards determine a speaker's context of assessment? There is a loose use of the word 'society' such that very small units, nested many times over within larger units, count in a perfectly good sense as societies, the expressions 'debating society', 'chess society', etc. are obviously not oxymoronic (see chapter 4, section 3). If the societal relativist's sense of 'society' captured these uses, this would be not much improvement over individualistic relativism. For it would then imply that there could be faultless disagreement which is internal to mid-sized and large societies, and this would clash with both the experimental data and the evolutionary story just mentioned. Thus, we should not use the word 'society' in this loose sense. How, then, should we define our theoretical use of this word?

Methodologically speaking, we should use the word in whatever way allows the theory to fit the data best. In that spirit, here is a proposal. A thinker/speaker's context of assessment, on this proposal determines the standards of (i) the largest group of people between which (ii) cooperation is feasible and which (iii) stands to benefit from arbitrating disputes involving the speaker, and is capable of effectively taking sides in order to do so.

First of all, this should solve the problem, mentioned above, of the possibility of faultless disagreement between idiosyncratic "mini-societies" and wider society at large (as between, for example, a neo-nazi cell, or religious cult, and wider national society), since such mini-societies would not count as societies in the theoretically relevant sense. This solution, moreover, is non-ad hoc, because, given the evolutionary story just described, we should expect a mechanism for putting normative pressure on parties to converge whenever there is a reasonable prospect of cooperation, so that permitting faultless disagreement between smaller units across which there is such a reasonable prospect would not be adaptive.

Let me provide some examples to help flesh this proposal out. First, this proposal predicts the plausible result that it is more likely that actual faultless disagreements existed in the distant historical past than that they do today. This is because historical societies were less interconnected, and there was less scope for cooperation between them. In contrast, modern societies are almost all interconnected at some level, and there is significantly greater scope for cooperation. Thus, for example, this view makes room for the possibility that there would have actually been faultless moral disagreement between societies in pre-columbian South America and societies in 13th century Europe. (Although, it does not guarantee this.)

Nevertheless, even today's societies are not perfectly integrated, and although cooperation is feasible on some issues, it is infeasible on others. Thus, it is attractive to interpret this proposal as making room for a kind of semantic flexibility. In one context, it may be that the standards relevant to a thinker or speaker are highly cross-cultural. For example, at a meeting of the United Nations, the relevant standards at the context may be determined by some compromise between the standards of all nation states. This is because the issues discussed at

a meeting of the United Nations will tend to be those on which it is feasible and beneficial for all major societies to cooperate, and with respect to which it is feasible for the international community to take sides in arbitration. When moral discourse focuses on questions such as these, this proposal predicts that the discourse will function under the assumption that faultless disagreement is globally ruled out, and parties will feel normative pressure to converge. On the other hand, there will be many issues on which it is neither feasible nor beneficial for the international community to cooperate/arbitrate. Thus if moral discourse focuses on more local questions (such as whether wealthy individuals should pay more or less tax) we should expect the assumption to be that faultless disagreement is only locally ruled out, so that individuals from different nation states may not feel any normative pressure to converge.

This does not necessarily mean that there are only two levels from which standards can be drawn – that of the nation state and the global society. However, it does seem plausible that at least the default assumption should be that the relevant standard is either that of an individual's national group or that of global society. Indeed, the concept of the nation seems quite generally to play roughly the role specified by (i) – (iii) above, as responsible for the highest-level arbitration of disputes (or in other words, the monopoliser of legitimate use of force¹²⁴) and promoter of cooperation. The proposal should nevertheless leave room (again, without guaranteeing) the possibility of some more local faultless disagreement, for it is not always feasible or desirable for national groups to arbitrate highly local matters. Of course, this is not to equate national groups with nation-states, or arbitration with legal sanction. National groups may have the capacity to arbitrate disputes through sub-legal means (e.g. via social sanction), and these means obviously need not recruit the machinery of the state. Nor does the assumption that national groups are the default standard-setting units require that arbitration normally *actually involves* the taking of sides by the entire national group (which, obviously, it normally does not), it merely requires that the standards relevant to the correct resolution of a dispute are determined by the largest group which can feasibly take sides in resolving it.

¹²⁴ (Weber, 1946).

This proposal may also provide an explanation of an apparent historical trend towards moral convergence, concurrent with globalisation (Huemer, 2016). This trend is admittedly predicted already by the empirical account, independently of the relativist's metaethical theory. For the empirical account predicts that people will seek to converge in their moral beliefs whenever cooperation between them is feasible. The relativist's metaethical theory, however, offers an explanation of why such convergence is the result of normative, not merely causal, factors. For on this theory, whenever new possibilities for cooperation arise, new "societies" (in the sense of (i) – (iii)) will spring into existence, and these societies will determine standards relative to which both parties to disputes over the rules for cooperation will be subject.

To be clear, none of this guarantees the actual existence of faultless disagreement at any level, for this will depend also on the extent of actual divergence in wide reflective equilibrium, which as we have seen is non-obvious. It does, however, serve to specify, in a relatively determinate way, a standard relevant at a given context.

Admittedly, this proposal does not render a speaker/thinker's context of assessment completely determinate, for indeterminacy may remain in terms of the contexts in which a given group is able to effectively arbitrate disputes or cooperate. However, the relativist can here appeal to an analogy with the norms governing correct use of language. Although correct use of language is relative to the standards of a linguistic community, linguistic communities are not always sharply bounded. Although the borders of modern nation states have made the relevant divisions clearer than they have been in the past, it is still often the case that there are no bright lines separating contexts in which different norms are in force. Exactly under what conditions, for example, uses of language is governed relative to the rules of Scots rather than Northern-Dialect English; Swiss German rather than Lichtensteinian German; Modern Standard Arabic rather than vernacular Arabic; and so on, is a vague matter. Socio-political considerations may come into play to resolve these indeterminacies (it has been said that a language is a dialect with an army and navy) (Harman, 2015). None of this, however, prevents the norms governing the use of language from being determinate enough for our purposes. Nor does it undermine the assumption, on the part of many ordinary thinkers, that there are

correct and incorrect ways of speaking. Analogously, there is no reason to think that the proposal advanced here should render moral beliefs and assertions indeterminate to any problematic degree.

Finally, there is an important question how the standards associated with any given society are determined. Above, I suggested that relativists can help themselves to Björnsson & McPherson's (2014) account of characteristically moral non-cognitive states, holding that an individual's standards are the Björnsson & McPherson-style non-cognitive states which would be associated with the moral beliefs they would hold in wide reflective equilibrium. But there is no straightforward move from claims about an individual's standards to claims about a society's standards. For it might be reasonably doubted whether societies are the kinds of things which can literally be said to have non-cognitive mental states.

Some have recently argued that certain social groups literally qualify as agents (Dyke, 2020). According to (Björnsson & Hess, 2017), furthermore, inasmuch as these groups qualify as agents, they would also qualify as *moral agents*, for they would be capable of literally possessing non-cognitive states which parallel the states that (according to Björnsson & McPherson, 2014) constitute moral judgements. If societal relativists could help themselves to this view, this would make things simpler. For there *would* then be a relatively straightforward move from an account of an individual's standards to an account of a society's standards.

However, although some societies (in the sense of the theory) may qualify as agents, it is unlikely that all such societies do. In particular, part of the motivation of the theory is to explain how very large groups can, in some contexts, determine relevant standards, and it seems unlikely that such groups qualify as agents. Instead, then, the relativist should opt for a model where the standards of a society are understood as aggregations of the standards of its members. This will obviously require some spelling out, and a number of distinct questions arise. What level of consensus among members is required, for example, if a society is to count as having a certain functional state (e.g. if 51% of the members oppose phi-ing, does this make it correct for all members to believe that phi-ing is wrong?). Further, how is the strength of relevant judgements determined (e.g. if 99% of members are only 51% in favour of phi-ing, is

it correct for members to have a very high credence in the view that phi-ing is required, or a marginal one?).

The already substantial length of this chapter prohibits detailed discussion of these issues, but here are some preliminary suggestions. First, it is plausible, given the overall account defended here, that we should expect substantial convergence between societies, at least in reflective equilibrium. Therefore, cases where societies are sharply divided (in reflective equilibrium) on relevant matters, should be rare. Second, it is independently plausible that some moral questions lack determinate answers¹²⁵. And therefore, inasmuch as at least some indeterminacy *does* result from intra-societal disagreement, this may be an explanatory payoff of the theory, rather than any kind of defect.

Furthermore, in advance of the admittedly important working out of these details, it is plausible that answers to these questions should be available. The analogy with the normativity of meaning is again helpful here. It is widely accepted that the norms governing language use are governed by patterns of linguistic behaviour (i.e. language use) within linguistic communities¹²⁶. But structurally similar questions should arise as to how these patterns aggregate to form rules governing entire communities (since the relevant linguistic behaviours are not completely uniform across communities, and linguistic communities are, equally, not to be construed as agents in any literal sense). Since, again, it is very plausible that the norms governing correct language use, relative to linguistic communities, are determined by aggregations of individual behaviour, it is reasonable to expect that coherent answers should be available to the structurally similar questions that arise in the moral case.

3. Comparison with other metaethical views

Since we have been able, here, to flesh out the substantive metaethical commitments of the version of relativism defended in this thesis, it may now be helpful to draw some brief but

¹²⁵ (Parfit, 2011).

¹²⁶ (Larson & Segal, 1995).

broad comparisons between this version of relativism and some of the other main metaethical contenders.

In comparison to *non-naturalist realism*, relativism looks attractive. This is because relativism avoids the epistemic and metaphysical burdens of robust non-naturalist realism, while retaining its commitment to cognitivism. Furthermore, as we have seen, the apparent advantages of the full objectivism which attach to non-naturalist realism are illusory. Rather, theories should *aim* to secure a certain degree of “subjectivism”, because there is evidence that faultless disagreement is possible. Non-naturalist realism is unable to account for this evidence.

Naturalist realism also fails to account for faultless moral disagreement. And, naturalist realism is vulnerable to familiar worries involving “moral twin-earth” style arguments from disagreement. Relativism avoids these worries, and therefore looks attractive relative to naturalist realism.

In comparison to *non-realist versions of objectivism*, relativism is attractive for two reasons. First, it does not require the success of dubious arguments to the effect that substantive moral requirements can be derived from practical reason alone. Second, again, these arguments are motivated by the assumption that morality is fully objective. And again, as we have seen, this assumption is undermined by experimental evidence. It is therefore not clear that the success of such arguments is even desirable.

Moving on to anti-objectivist theories. The arguments of this thesis have interesting upshots for the dialectic between *error theorists* and relativists. Error theory is normally motivated by a pair of claims. First, the claim that ordinary thought and talk is committed to objectivity, and second, the claim that this commitment is not satisfied. But although this thesis has not discussed the specific form of objectivity which error theorists have focused on most centrally (that is, rational categoricity of moral requirements), it has shown that in at least one important respect, ordinary moral discourse does not appear to be committed to full objectivity. Furthermore, when error theorists and relativists have engaged with one another (Finlay,

2008) (Joyce, 2011) the error theorist's arguments have often focused on features specific to content, rather than truth relativism. The form of truth relativism defended here is immune to many of the concerns, about a relativist analysis of ordinary thought and talk, put forward by error theorists.

Finally, this thesis has largely avoided discussion of expressivist views. This is because there is a huge literature on this topic, containing a dizzying variety of proposals. It would thus have been impossible to do justice to the vibrant research programme of expressivism here. However, at least *prima facie*, relativism does seem to have some significant advantages over expressivism. First, it is compatible with a standard truth-conditional semantics, which many find independently attractive. Second, and as a result, it has no need to grapple with the infamous Frege-Geach problem, which is an outstanding pain point for expressivists. Furthermore, given the account of moral standards defended here, relativism would seem to capture many of the same theoretical desiderata, in terms of connection between moral judgement and motivation, that motivate expressivism. At the same time, we have seen that relativism avoids any implausibly subjectivist upshots, where the attribution of such upshots to relativism has long been taken to be a key motivation for preferring an expressivist approach over a cognitivist anti-realist one. To repeat a point once again, we have seen that it is in fact a desideratum of a metaethical theory that it should account for at least some degree of explicit subjectivity, in the form of faultless disagreement. Many expressivist theories, however, are explicitly designed to avoid such implications. Schroeder (2014) for example, in an attempt to fend off allegations of subjectivism, claims that *all that can be said* in terms of norms governing moral speech acts and judgements is that they are sincere (or that they are not). It is not even coherent, according to Schroeder, to level objections against expressivism which take the form of claims about the correctness of assertions. If this is right, faultless disagreement would seem to be ruled out not on substantive metaethical grounds, but rather trivially, purely in virtue of the machinery of the theory. Relativism, by contrast, makes the attribution of faultless disagreement a substantive metaethical matter. In light of the evidence discussed here, this seems the preferable approach.

Metaethical truth-relativism thus appears to have many attractive features, in comparison with some of the other main theoretical options. It thus appears to offer an exciting new avenue for further research. This thesis has only skimmed the surface of such a research programme, and many open questions remain. I sketch some of these questions in the concluding chapter.

CONCLUSION

ABSTRACT: This conclusion recapitulates the key claims and themes of the thesis. Outstanding issues, and directions for future research are also discussed.

It is perhaps a telling measure of the extent to which relativism is viewed as unworthy of consideration, among professional metaethicists, that it is not awarded a chapter in the most widely used advanced textbook: Miller's excellent (2013).¹²⁷ Against this view, this thesis has attempted to show that metaethical relativism is an attractive and plausible theory, which deserves to be taken seriously as a main contender in the metaethical space. I do not claim to have shown that metaethical relativism is immune to criticism. But I do claim to have shown that the simple objections that are often taken to straightforwardly refute it, do not do so. Whether or not the thesis has succeeded in convincing readers that relativism is true, I will consider it a success if it has convinced readers that relativism is as plausible as many other mainstream theories, and far more plausible than its lack of representation in standard textbooks and the literature suggests.

At the outset of the project, I expected that an argument from disagreement would provide the crucial evidence for relativism. My experiences of disagreement in normative ethics (specifically, the apparent unresponsiveness of such disagreements to resolution by rational means) had undermined my earlier confidence in moral realism. And, I had expected that this

¹²⁷ In fact, relativism receives *no* significant attention in Miller (2013).

experience would ground an argument for moral relativism, involving claims about moral epistemology. Chapter 1 grew out of this expectation. But as I ultimately argued, in that chapter, it is unclear that such considerations undermine moral objectivism, let alone (what would in any case be a further step) that they support relativism.

Fortunately, this thesis was written concurrent with a wider surge in interest in the phenomenon of so-called faultless disagreement, and its connections with a new form of relativism. The key idea of this research is that a much stronger claim about disagreement is true: that both parties can be free not only of epistemic fault, but of any fault whatsoever. Not only, proponents of this research have argued, does this claim support relativism broadly construed, it supports, in particular, a new form of relativism, involving new semantic notions. This form of relativism also, happily, seems to avoid many of the objections which had historically been levelled against traditional forms of relativism in metaethics. Although the mainstream of this debate focuses on paradigmatically subjective topics, such as taste and aesthetics, I have argued (in chapter 2 through 5) that these ideas can be effectively applied to ethics.

Although this argument, and this view, seems attractive, there are points of detail which need to be worked out if it is to be applied to ethics. For one, metaethicists need to be convinced (what, in my impression, they often doubt) that moral disagreement can be faultless. I attempted to do this in chapter 2. Furthermore, the claim that truth relativism offers a superior account of disagreement (compared with content relativism) cannot be lazily assumed. For critics of truth relativism, and proponents of content relativism, have argued at some length that, despite initial appearances, this is not in fact the case. Chapters 3 and 4 therefore developed in detail this key move in the argument of chapter 2, that truth relativism offers a more attractive account of disagreement than content relativism. This dialectical point has both a positive and a negative side. On the positive side, I argued (chapter 3) that truth relativism has access to a positive account of disagreement which is not available to content relativism. And on the negative side, I argued (chapter 4) that content relativism suffers from objections to which truth relativism is immune, and that content relativists are unlikely to be able to deflect such objections.

I therefore concluded, at the close of chapter 4, that truth relativism is indeed evidenced by an argument from disagreement, and that it does indeed solve at least some of the most salient historical objections to metaethical relativism. Two questions, at this stage, remained. First, although this view seems to solve the objection from disagreement, the worry that relativism attributes an implausible degree of subjectivity to ethics may remain. Second, the thesis had, until this point, said very little about the substantive metaethical commitments of relativism, focusing instead on semantic questions. Happily, dealing with the first objection provided opportunities to comment on the second point also. Thus, in chapter 5, I argued that the objection just mentioned is less threatening than it is often claimed, while also fleshing out the details of the metaethical theory associated with the semantics defended in earlier chapters. In conclusion, I argued that when this theory is compared with the other main metaethical options, it seems, *prima facie*, a serious contender.

This thesis has more or less taken the semantic framework of truth relativism for granted. This seemed justified, since much excellent work has recently been done in an attempt to show that this framework is in general viable. But although this claim has gained much ground in recent years, interesting outstanding questions (and on the part of some, doubts) remain (Evers, 2020). Perhaps most interesting, to my mind, are questions about the fundamental nature of representation. Relativists like to point out that relativity at the level of proposition truth is utterly orthodox – for it is utterly orthodox to hold that contents have their truth values only relative to possible worlds. However, there is a clear sense in which relativism also departs from the orthodox semantic/pragmatic picture. For on the orthodox picture, there is a straightforward sense in which a particular possible world, namely the actual one, is objectively privileged. And arguably, this component of the picture is crucial to our understanding of representation, and of successful communication. Relativism does away with this component, for relativists are committed to denying that there is any objectively privileged moral standard.

So this is a key outstanding puzzle for truth relativism as a domain-independent research programme. Truth relativists have said much about why a relativistic notion of *truth* makes

sense, but they have said less about why a relativistic notion of *representation* makes sense (Einheuser, 2012) (Kolbel, 2018) (2015a) (2015b). Although the focus of this thesis has been the metaethical application of this framework, rather than its domain-neutral elucidation, metaethical relativists should take a keen interest in ongoing research on this question.

Another outstanding puzzle concerns the fundamental nature of disagreement itself. Truth relativist perspectives on disagreement have reinvigorated interest in this question, which had previously seemed straightforward (see ch3 sct 1). The discussion of relativism in this thesis made the simplifying move of seeking simply to show that relativism can account for our intuitive attributions of disagreement. But it is a further question what, on a deeper level, these intuitions are tracking. Given the special importance of disagreement in the domain of ethics, this question is one in which metaethical relativists have reason to take an interest going forward.¹²⁸

Finally, there are two important respects in which the arguments of this thesis have turned on empirical considerations. The first, most obviously, is that the key thesis (chapter 2) that there can be faultless moral disagreements turns on experimental data, which would appear to cut against many (although not all) metaethicists' intuitions. Since these results are at this stage limited (both in terms of replication, and in terms of variation), the case for relativism would be strengthened by additional experimental research in this area. For instance, if Sarkissian et al.'s pioneering (2011) work on cross-cultural disagreement can be replicated as robustly as Goodwin & Darley's (2008) equally pioneering work on disagreement (*simpliciter*) has been, it would then become increasingly difficult to deny that, at least, there is an important data point to be explained, in the region of faultless moral disagreement.

The second empirical consideration has been the claim (chapter 5) that our best understanding of the evolutionary explanation of the development of our moral concepts primarily involves the functioning of these concepts in contexts where social interaction between moral judges is feasible. This is an important point in support of relativism, for it would seem to provide a satisfying explanation, in evolutionary terms, of the experimental data just discussed. And, it

¹²⁸ (Baker, 2013) (Marques, 2014) (Beddor, 2019), (Zeman, 2020).

would seem to suggest that a fully objectivist understanding of such concepts (where such objectivity holds even in contexts, like Hyperspace Bypass, where social interaction is impossible) would be unexpected. Taken together, these two empirical claims seem to undermine a long-standing sense, among metaethicists, that there are only three plausible options in the space of metaethical theory: first, a realist view which straightforwardly and robustly vindicates the full objectivity putatively inherent in the discourse (via the postulation of a distinctively moral stratum in objective reality); second, an expressivist view which non-straightforwardly vindicates this same putative objectivity (via a global deflationism); or third, an error-theoretic view which throws its hands in the air and relegates the discourse to the historical rubbish heap of myth and superstition. Against this framing, the two empirical strands discussed herein suggest that we might be, with caution, more optimistic. Perhaps, these considerations suggest, some of the deep puzzles of metaethics may turn simply on an overestimation of the degree of objectivity required to avoid error. Further research in these two areas is thus of great interest.

References

- Abarbanell, L. & Hauser, M. D. (2010), 'Mayan Morality: An exploration of permissible harms', *Cognition* 115 (2): 207-224.
- Adams, D. (2007) *The Hitch-Hiker's Guide to the Galaxy*, New York: Penguin Random House.
- Anscombe, G. E. (1958) 'Modern Moral Philosophy', *Philosophy*, 33 (124), 1-19.
- Baker, C. & Robson, J. (2017), 'An Absolutist Theory of Faultless Disagreement in Aesthetics', *Pacific Philosophical Quarterly*, 98(3).
- Baker, C. (2013) 'The role of disagreement in semantic theory', *Australasian Journal of Philosophy*, (1): 1-18.
- Beddor, B. (2019) 'Subjective Disagreement', *Nôus*, 53 (4), 819-851.
- Beebe, J. R. & Sackris, D. (2016) 'Moral objectivism across the lifespan', *Philosophical Psychology*, 29 (6): 912-929.
- Beebe, J. R., Qiaoan, R., Wysocki, T. & Endara, A. M. (2015) 'Moral Objectivism in Cross-Cultural Perspective', *Journal of Cognition and Culture*, 15 (3-4), 386-401.
- Bennigson, T. (1996) 'Irresolvable Disagreement and the Case Against Moral Realism', *The Southern Journal of Philosophy*, 34 (4): 411-437.
- Björnsson, G. & Hess, K. (2017), *Philosophy and Phenomenological Research*, 94 (2): 273-298.
- Björnsson, G. & McPherson, T. (2014) 'Moral Attitudes for Non-Cognitivists: Solving the Specification Problem', *Mind*, 123 (428): 1-38.
- Björnsson, G. (2012) 'Do "Objectivist" Features of Moral Discourse and Thinking Support Moral Objectivism?', *The Journal of Ethics*, 16 (4): 376-393.
- Björnsson, G. (2015) 'Disagreement, correctness, and the evidence for metaethical absolutism', in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* 8.
- Blackburn, S. (1984) *Spreading the Word: Groundings in the Philosophy of Language*, Oxford: Oxford University Press.
- Blackburn, S. (1998) *Ruling Passions: A Theory of Practical Reasoning*, Oxford: Oxford University Press.
- Brandom, R. B. (1994) *Making it Explicit: Reasoning, Representing, and Discursive Commitment*, MA: Harvard University Press.

- Brandt, R. (1954) *Hopi Ethics*, Chicago: Chicago University Press.
- Brink, D. (1988) 'How to be a Moral Realist', in G. Sayre-McCord (ed.), *Essays on Moral Realism*. London: Cornell University Press.
- Brink, D. O. (1984) 'Moral realism and the sceptical arguments from disagreement and queerness', *Australasian Journal of Philosophy*, 62 (2): 111-125.
- Brogaard, B. (2008) 'Moral Contextualism and Moral Relativism', *Philosophical Quarterly*, 58 (232): 385-409.
- Brogaard, B. (2012) 'Moral Relativism and Moral Expressivism', *Southern Journal of Philosophy*, 50 (4): 538-556.
- Cappelen, H. & Hawthorne, J. (2009), *Relativism and Monadic Truth*, Oxford: Oxford University Press.
- Cath, Y. (2016) 'Reflective Equilibrium', in Cappelen H., Gendler, T. & Hawthorne, J. (eds.), *The Oxford Handbook of Philosophical Methodology*, Oxford: Oxford University Press.
- Chagnon, N. (1967) *Yanomamö: The Fierce People*, New York: Hold, Rinehart & Winston.
- Chan, A. (2013) (ed.) *The Aim of Belief*, Oxford: Oxford University Press.
- Clarke-Doane, J. (2013) 'Moral Epistemology: The Mathematics Analogy', *Noûs*, 48 (2): 238-255.
- Copp, D. (1995) *Morality, Normativity, and Society*, Oxford: Oxford University Press.
- Cuneo, T. & Shafer-Landau, R. (2003), 'The Moral Fixed Points: new directions for moral nonnaturalism', *Philosophical Studies*, 171 (3): 399-443.
- Daniels, N. (1979) 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy*, 76 (5): 256-282.
- De Sélincourt, Aubrey, trans. 1972 (1954). Herodotus, *The Histories*. London, New York: Penguin Books.
- DePaul, R. M. (1986) 'Reflective Equilibrium and Foundationalism', *American Philosophical Quarterly* 23 (1): 59-69.
- Doris, J. & Plakias, A. (2008) 'How to Argue About Disagreement: Evaluative diversity and moral realism', In Sinnott-Armstrong, W. (ed.), *Moral Psychology Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, MA: MIT Press.
- Dowell, J. L. (2016) 'The Metaethical Insignificance of Moral Twin Earth', in Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics XI*, Oxford: Oxford University Press.

- Dreier, J. (2009) 'Relativism (and Expressivism) and the Problem of Disagreement', *Philosophical Perspectives*, 23 (1), 79-110.
- Dwyer, S. (2009) 'Moral Dumbfounding and the Linguistic Analogy: Methodological Implications for the Study of Moral Judgement', *Mind and Language*, 24 (3): 274-296.
- Dwyer, S. (2009) 'Moral Dumbfounding and the Linguistic Analogy', *Mind & Language*, 24 (3): 274-296
- Dyke, M. M. (2020) 'Group Agency Meets Metaethics: How to Craft a More Compelling Form of Normative Relativism', in Shafer-Landau (ed.), *Oxford Studies in Metaethics* 15.
- Egan, A. (2014) 'There's Something Funny About Comedy: A Case Study in Faultless Disagreement', *Erkenntnis*, 79 (S1), 73-100.
- Einheuser, I. (2006) 'Counterconventional Conditionals', *Philosophical Studies*, 127 (3): 459-482.
- Einheuser, I. (2012) 'Relativized Propositions and the Fregean Orthodoxy', *Philosophy and Phenomenological Research*, 84 (3): 590-603.
- Enoch, D. (2009) 'How is Moral Disagreement a Problem for Realism', *The Journal of Ethics*, 13 (1): 15-50.
- Enoch, D. (2011) *Taking Morality Seriously: A Defence of Robust Realism*, Oxford: Oxford University Press.
- Eriksson, J. & Tiozzo, M. (2016) 'Matters of Ambiguity: Faultless Disagreement, Relativism and Realism', *Philosophical Studies*, 173 (6), 1517-1536.
- Eriksson, J. (2016) 'Expressivism, Attitudinal Complexity and Two Senses of Disagreement in Attitude', *Erkenntnis*, 81 (4): 775-794
- Evers, D. & Streumer, B. (2016) 'Are the Moral Fixed Points Conceptual Truths?', *Journal of Ethics and Social Philosophy*, (1): 1-9.
- Evers, D. (2021) 'Relativism and the Metaphysics of Value', *British Journal of Aesthetics*, 61 (1).
- Finlay, (2008) 'The Error in Error Theory', *Australasian Journal of Philosophy*, 86 (3): 347-369.
- Finlay, S. & Björnsson, G. (2010) 'Metaethical Contextualism Defended', *Ethics*, 121 (1): 7-36.
- Finlay, S. (2011) 'Errors Upon Errors: A Reply to Joyce', *Australasian Journal of Philosophy*, 89 (3): 535-547.
- Finlay, S. (2014) *Confusion of Tongues*, Oxford: Oxford University Press.
- Fogal, D. (2016) 'Confusion of Tongues: A Theory of Normative Language, by Stephen Finlay', *Ethics*, 127 (1): 281-288.

- Francén, R. (2007) *Metaethical Relativism: Against the Single Analysis Assumption*, Dissertation, University of Gothenburg.
- Francén, R. (2010) 'No Deep Disagreement for New Relativists', *Philosophical Studies*, 151 (1), 19-37.
- Francén, R. (2016) 'Some Varieties of Metaethical Relativism', *Philosophy Compass*, 11 (10): 529-540.
- Fraser, B. & Hauser, M. D. (2010) 'The Argument from Disagreement and the Role of Cross-Cultural Empirical Data', *Mind and Language*, 25 (5): 541-560.
- Gewirth, A. (1968) *Reasons and Morality*, London: University of Chicago Press
- Gibbard, A. (2003) *Thinking How to Live*, Oxford: Oxford University Press.
- Goodwin, G. P. & Darley, J. M. (2008) 'The Psychology of Meta-ethics: Exploring Objectivism', *Cognition*, 106 (3): 1339-1366.
- Gowans, C. W. (ed.) (2000) *Moral Disagreements: Classic and Contemporary Readings*, NY: Routledge.
- Haidt, J. (2012) *The Righteous Mind: Why good people are divided by politics and religion*, NY: Vintage.
- Hallam, S., Cross, I. & Thaut, M (2009) *The Oxford Handbook of Music Psychology*, (2nd ed), Oxford: Oxford University Press.
- Hallam, S., Cross, I. & Thaut, M (2009) *The Oxford Handbook of Music Psychology*, (2nd ed), Oxford: Oxford University Press.
- Harman, G. (1975) 'Moral Relativism Defended', *Philosophical Review*, 84 (1): 3-22.
- Harman, G. (1977) *The Nature of Morality*, Oxford University Press: Oxford.
- Harman, G. (2000) *Explaining Value and Other Essays in Moral Philosophy*, Oxford: Oxford University Press.
- Harman, G. (2015) 'Moral Relativism is Moral Realism', *Philosophical Studies*, 172 (4): 855-863.
- Hills, A. (2013) 'Faultless Moral Disagreement', *Ratio*, 26 (4): 410-427.
- Horgan, M. & Timmons, M. (1991) 'New Wave Moral Realism Meets Moral Twin Earth', *Journal of Philosophical Research*, 16: 447-465.
- Huemer, M. (2005) *Ethical Intuitionism*, Hampshire: Palgrave Macmillan.
- Hume, D. (1910) 'Of the Standard of taste', in C.W. Elliot (ed.), *English Essays from Sir Philip Sidney to Macaulay*, 215-236.

- Incurvati, L. & Schlöder, J. J. (2017), 'Weak Rejection', *Australasian Journal of Philosophy*, 95 (4): 741-760.
- Jackson, F. & Pettit, P. (1998) 'A Problem for Expressivism' *Analysis*, 58 (4): 239-251.
- Joyce, R. (2001) *The Myth of Morality*, Cambridge: Cambridge University Press.
- Joyce, R. (2007) *The Evolution of Morality*, London: MIT Press.
- Joyce, R. (2011) 'The Error in 'The Error in Error Theory'', *Australasian Journal of Philosophy*, 89 (3): 519-534.
- Kaplan, D. (1989) 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals', In Almog, J., Perry, J. and Wettstein, H. (eds), *Themes From Kaplan*, Oxford: Oxford University Press.
- Kelly, T. & McGrath, S. (2010) 'Is Reflective Equilibrium Enough?', *Philosophical Perspectives*, 24 (1): 325-359.
- Kelly, T. (2016) "Evidence", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/win2016/entries/evidence/>](https://plato.stanford.edu/archives/win2016/entries/evidence/).
- Khoo, J. & Knobe, J. (2016) 'Moral Disagreement and Moral Semantics', *Noûs*, 52 (1), 109-143.
- Kolbel, M. (2004a) 'Faultless Disagreement', *Proceedings of the Aristotelian Society*, 104 (1), 53-73.
- Kolbel, M. (2004b) 'Indexical Relativism Versus Genuine Relativism', *International Journal of Philosophical Studies*, 12 (3), 297-313.
- Kolbel, M. (2015a) 'Relativism 1: Representational Content', *Philosophy Compass*, 10 (1): 38-51.
- Kolbel, M. (2015b) 'Relativism 2: Semantic Content', *Philosophy Compass*, 10 (1): 52-67.
- Korman, D. Z. & Locke, D. (2020) 'Evolutionary Debunking and Moral Relativism', in Kusch, M. (ed.), *The Routledge of Handbook of Relativism*, NY: Routledge.
- Korsgaard, C. M. (1996) *The Sources of Normativity*, Cambridge: Cambridge University Press
- Lasersohn, P. (2005) 'Context dependence, disagreement, and predicates of personal taste', *Linguistics and Philosophy*, 28 (6), 643-686.
- Lewis, D. (1980) 'Index, Context, and Content', In Kanger, S. & Ohman, S (eds), *Philosophy and Grammar*.
- Loeb, D. (1998) 'Moral Realism and the Argument From Disagreement', *Philosophical Studies*, 90: 281-303.

- López de Sa, D. (2008) 'Presuppositions of Commonality: An Indexical Relativist Account of Disagreement', In Garcia-Carpintero, M & Kolbel, M. (eds.), *Relative Truth*, Oxford: Oxford University Press, 297-310.
- MacFarlane, J. (2007) 'Relativism and Disagreement', *Philosophical Studies*, 132 (1), 17-31.
- MacFarlane, J. (2014) *Assessment Sensitivity*, Oxford: Oxford University Press.
- MacFarlane, J. (2020) 'Indeterminacy as Indecision, Lecture II: Seeing Through the Clouds' *Journal of Philosophy*, 117 (11): 617-642
- Machery, E., Kelly, D., Stich, S. (2005), 'Moral Realism and Cross Cultural Normative Diversity', *Behavioural and Brain Sciences*, 28 (6): 830.
- Mackie, J. (1977) *Ethics: Inventing Right And Wrong*, London: Penguin.
- Marques, T. & Garcia-Carpintero, M. (2014) 'Disagreement about Taste: Commonality Presuppositions and Coordination', 92 (4), 701-723.
- Marques, T. (2014) 'Doxastic Disagreement', *Erkenntnis*, 79 (1): 121-142.
- McGrath, S. (2007) 'Moral Disagreement and Moral Expertise', in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics 4*, Oxford: Oxford University Press.
- McGrath, S. (2010) 'Moral Realism Without Convergence', *Philosophical Topics*, 38 (2), 59-90.
- McHugh, C. (2012) 'The Truth Norm of Belief', *Pacific Philosophical Quarterly*, 93 (1): 8-30.
- Miller, A. (2013) *Contemporary Metaethics: An Introduction (2nd ed)*, Cambridge: Polity.
- Moody-Adams, M. (1997) *Fieldwork in Familiar Places*, Cambridge, MA: Harvard University Press.
- Nisbett, R. E. & Cohen, D. (1996) *Culture of Honor: The psychology of violence in the south*.
- Olson, J. (2014) *Moral Error Theory: History, Critique, Defence*, Oxford: Oxford University Press.
- Palmira, M. (2015) 'The Semantic Significance of Faultless Disagreement', *Pacific Philosophical Quarterly*, 96 (3), 349-371.
- Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press.
- Parfit, D. (2011) *On What Matters (Vol 2)*, Oxford University Press: Oxford.
- Pietroiuști, G. (2020) 'Disagreement and Conflict: How Moral and Taste Judgements Do Not Differ', *Theoria*.
- Plantinga, A. (1993) *Warrant and Proper Function*, Oxford: Oxford University Press.
- Plunkett, D. & Sundell, T. (2013) 'Disagreement and the Semantics of Normative and Evaluative Terms', *Philosophers Imprint*, 13 (23), 1-37.

- Prinz, J. (2007) *The Emotional Construction of Morals*, Oxford: Oxford University Press.
- Putnam, H. (1975) 'The Meaning of 'Meaning'', in *Mind, Language and Reality: Philosophical Papers, Volume 2*, Cambridge: Cambridge University Press: 215-272.
- Rawls, J. (1951) 'Outline of a Decision Procedure for Ethics', *Philosophical Review*, 60 (2): 177-197.
- Rawls, J. (1974) 'The Independence of Moral Theory', In Freeman, S. (ed.), *John Rawls: Collected Papers*. MA: Harvard University Press.
- Ridge, M. (2013) 'Disagreement', *Philosophy and Phenomenological Research*, 86 (1): 41-63.
- Ridge, M. (2014) *Impassioned Belief*, Oxford: Oxford University Press.
- Ripley, D. (2011) 'Negation, Denial, Rejection', *Philosophy Compass*, 6 (9): 622-629.
- Risberg, O. & Tersman, F. (2020) 'A New Route from Moral Disagreement to Moral Skepticism', *Journal of the American Philosophical Association*, 5 (2): 189-207
- Rowland, R. (2017) 'The Significance of Significant Fundamental Moral Disagreement', *Noûs*, 51 (4), 802-831.
- Sarkissian, H., Park, J., Tien, D., Wright, J., & Knobe, J. (2011) 'Folk Moral Relativism', *Mind and Language*, 26 (4), 482-505.
- Sauer, H. (2019) 'The Argument From Agreement: How universal values undermine moral realism', *Ratio*, 32 (4): 339-352.
- Sayre-McCord, G. (1996) 'Coherentist Epistemology and Moral Theory', In Sinnott-Armstrong, W. & Timmons, M. (eds.), *Coherentist Epistemology and Moral Theory*.
- Schafer, K. (2011) 'Faultless Disagreement and Aesthetic Realism', *Philosophy and Phenomenological Research*, 82 (2), 265-286.
- Schafer, K. (2012) 'Assessor Relativism and the Problem of Moral Disagreement', *Southern Journal of Philosophy*, 50 (4), 602-620.
- Schafer, K. (2014) 'Constructivism and Three Forms of Perspective-Dependence in Metaethics', *Philosophy and Phenomenological Research*, 89 (1), 68-101.
- Schroeder, M. (2008) *Being For: Evaluating the Semantic Program of Expressivism*, Oxford: Oxford University Press.
- Schroeder, M. (2009) 'Hybrid Expressivism: Virtues and Vices', *Ethics*, 119 (2): 257-309.
- Schroeder, M. (2014) 'Does Expressivism Have Subjectivist Consequences', *Philosophical Perspectives*, 28 (1): 278-290.

- Segal, G. & Larson, R. (1995) *Knowledge of Meaning: An Introduction to Semantic Theory*, MA: MIT Press.
- Seipel, P. (2020) 'Famine, Affluence, and Philosophers' Biases', *Philosophical Studies*, 177 (10): 2907-2926.
- Shafer-Landau, R. (2003) *Moral Realism*, Oxford: Oxford University Press
- Silk, A. (2017) 'Metaethical Contextualism', In McPherson, T. & Plunkett, D. (eds), *The Routledge Handbook of Metaethics*, 102-112.
- Singer, P. (1972) 'Famine, Affluence, and Morality', *Philosophy and Public Affairs*, 1 (3): 229-243.
- Smith, M. (1994) *The Moral Problem*, Oxford: Oxford University Press.
- Speaks, J. (2019) 'Theories of Meaning', *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/spr2021/meaning/>](https://plato.stanford.edu/archives/spr2021/meaning/).
- Stalnaker, R. (1978) *Syntax and Semantics*
- Stanford, K. P. (2018) 'The difference between ice cream and Nazis: Moral externalisation and the evolution of human cooperation', *Behavioural and Brain Sciences*, 41.
- Stevenson, C. L. (1937) 'The Emotive Meaning of Ethical Terms', *Mind*, 46 (181).
- Stojanovic, I. (2007) 'Talking About Taste: Disagreement, Implicit Arguments, and Relative Truth', *Linguistics and Philosophy*, 30 (6): 691-706.
- Stojanovic, I. (2019) 'Disagreements about Taste vs. Disagreements about Moral Issues', *American Philosophical Quarterly*, 56 (1): 29-42
- Street, S. (2006) 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies*, 127 (1): 109-166.
- Street, S. (2016) 'Objectivity and Truth: You'd Better Rethink It', in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics 11*, Oxford: Oxford University Press.
- Sturgeon, N. L. (1994) 'Moral Disagreement and Moral Relativism', *Social Philosophy and Policy*, 11 (1): 80-115.
- Suikkanen, J. (2009) 'The Subjectivist Consequences of Expressivism', *Pacific Philosophical Quarterly*, 90 (3): 364-387.
- Suikkanen, J. (2019) 'Contextualism, Moral Disagreement, and Propositions Clouds', in Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics XIV*, Oxford: Oxford University Press.

- Tersman, F. (2006) *Moral Disagreement*, Cambridge: Cambridge University Press.
- Tolhurst, W. (1987) 'The Argument From Moral Disagreement', *Ethics*, 97 (3): 601-621.
- Turnbull, M. G. (2018) 'Underdetermination in Science: What it is and Why We Should Care', *Philosophy Compass*, 13 (2).
- von Fintel, K. & Gillies, A. (2011) "'Might" Made right', in Egan, A. & Weatherson, B. (eds.) *Epistemic Modality*, Oxford: Oxford University Press.
- Weber, M. (1946) *Politics as a Vocation*, NY: Oxford University Press.
- Westermarck, E. (1906) *The Origin and Development of the Moral Ideas*, NY: Books for Libraries Press.
- Westermarck, E. (1932) *Ethical Relativity*, NY: Routledge.
- Williams, B. (1981) 'The Truth in Relativism', in *Moral Luck: Philosophical Papers*, Cambridge: Cambridge university Press: 132-142.
- Williamson, T. (2000) *Knowledge and its Limits*, Oxford: Oxford University Press.
- Wong, D. (1984) *Moral Relativity*, California: University of California Press.
- Woods, J. (2020) 'Footing the Cost (of Normative Subjectivism)', in Suikkanen, J. & Kauppinen, A. (eds.), *Methodology of Moral Philosophy*: Forthcoming: Routledge.
- Wright, C. (1992) *Truth and Objectivity*, MA: Harvard University Press.
- Wright, C. (1995) 'Truth in Ethics', *Ratio*, 8 (3): 209-226.
- Zeman, D. (2020) 'Faultless Disagreement', in Martin Kusch (ed.), *The Routledge Handbook of Philosophy of Relativism*, Routledge: Oxford.
- Zeman, D. (2020) 'Minimal Disagreement', *Philosophia*, 48 (4): 1649-1670.

