Twin Earth and the Normativity of Meaning
Jon Keyzer

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago, Dunedin,
New Zealand

November 2016
Abstract

In this dissertation, I attempt to provide some new evidence in favour of the claim that meaning is normative—specifically, for the claim that semantic judgments or ascriptions of meaning are action-guiding. I attempt to achieve this by developing an analogue of the Moral Twin Earth argument advanced by Horgan and Timmons (1992a) which I call the ‘Meaning Twin Earth’ argument. In the course of the dissertation, I outline Kripke’s 1982 sceptical argument for the thesis that there are no meaning facts in virtue of which ascriptions of meaning are true or false and highlight how the claim that meaning is normative is necessary for that argument to succeed with the kind of generality that Kripke intends. I then explain how one of Kripke’s main arguments against dispositionalist accounts of meaning can be viewed as a kind of open question argument. This argument is ultimately a failure, but nonetheless, I argue that the argument can be revised in a way that does pose a genuine threat to dispositionalist accounts of meaning by using an analogue of Horgan and Timmons’ Moral Twin Earth argument. I claim that a Meaning Twin Earth scenario yields linguistic intuitions that, in my view, are best explained by invoking the claim that meaning is normative. These intuitions constitute evidence against both reductive dispositionalism and anti-normativism about meaning. I forestall several potential objections to the Meaning Twin Earth argument drawn from Plunkett and Sundell (2013), Copp (2000), Merli (2002), and Baker (2016). In closing, I consider two broad options for a normativist account of semantic judgment: an expressivist form (Gibbard) and a rationalist form (McDowell). I argue that, provisionally, McDowell’s rationalist version of factualism about meaning faces a less serious challenge than Gibbard’s expressivist alternative.
Acknowledgments

First and foremost, I am indebted to my primary supervisor Alex Miller for helping develop many of the core ideas and arguments that comprise the dissertation. Alex is an incredibly diligent, patient, and encouraging person and I have learned and benefited immensely from his supervision. I couldn’t have done it without him and am very grateful for our many conversations over the years. I am also indebted to my secondary supervisor Charles Pigden for his comments on the final drafts of the dissertation.

I would like to thank the Philosophy Department at the University of Otago for its ongoing support. I have benefited from countless seminars and discussions with members of the department. In particular, I would like to thank Kirk Michaelian, Michael LeBuffe, and Daniel Wee for many chats around the department and at the pub.

I would like to thank my family and friends, who always listened and encouraged me, especially when the going was tough. My mother, Janet Fulton, has always been a source of inspiration for me and pursuing the Ph.D. is one of the greatest things she has inspired me to do. Emma Payton, Matt Connelly, Hazem Zohny, and many more friends have been supportive of me and I am both grateful and humbled to be in such good company.

Last but not least, I am indebted to the University of Otago for their financial support for the duration of the thesis by granting me the generous Doctoral Scholarship.
Contents

Introduction ........................................................................................................................................ 1
Chapter Outline ............................................................................................................................... 1
Overview ......................................................................................................................................... 4

Chapter One: Kripke’s Sceptical Argument ............................................................................... 5
§1.1 Introduction ............................................................................................................................. 5
§1.2 Kripke’s Wittgenstein and the Sceptical Paradox ................................................................. 5
§1.2.1 Ground Rules ...................................................................................................................... 7
  §1.2.2 Generalizing the Sceptical Challenge ............................................................................... 7
§1.3 Two Constraints on Any Account of Meaning ................................................................. 8
  §1.3.1 The Extensionality Constraint ....................................................................................... 8
  §1.3.2 The Normativity Constraint ......................................................................................... 9
§1.4 Scope of the Sceptical Challenge ...................................................................................... 9
§1.5 Two Ways to Respond to the Sceptical Challenge ............................................................ 10
§1.6 Candidate Facts .................................................................................................................. 11
  §1.6.1 Past Behaviour or Mental History ............................................................................... 11
  §1.6.2 Set of Instructions or General Rule ............................................................................. 11
  §1.6.3 Dispositionalism .......................................................................................................... 12
  §1.6.4 Simple Dispositionalism .............................................................................................. 12
  §1.6.5 KW’s Finitude Objection ............................................................................................. 13
  §1.6.6 KW’s Normativity Objection ....................................................................................... 14
  §1.6.7 Sophisticated Dispositionalism .................................................................................... 16
  §1.6.8 KW’s Revived Finitude Objection .............................................................................. 18
  §1.6.9 KW’s Circularity Objection ......................................................................................... 19
  §1.6.10 Boghossian’s Reply to KW’s Revived Finitude Objection ......................................... 20
  §1.6.11 Boghossian’s Belief-Holism Objection ..................................................................... 21
  §1.6.12 Simplicity ................................................................................................................... 22
  §1.6.13 Mental Images and other Mental States ................................................................. 22
  §1.6.14 Primitive Mental State ............................................................................................... 23
  §1.6.15 Platonism ................................................................................................................... 24
§1.7 What is the Normativity of Meaning? ................................................................................. 25
  §1.7.1 Hattiangadi: Normativity and Norm-Relativity ......................................................... 25
§1.8 Conclusion ............................................................................................................................ 28

Chapter Two: Twin Earth and the Open Question Argument .............................................. 30
§2.1 Introduction ............................................................................................................................ 30
§2.2 Moore: The Classical Open Question Argument ............................................................. 30
§2.3 Zalabardo: KW’s Normativity Argument .......................................................................... 32
§2.4 Horgan and Timmons: The Open Question Argument Revised .................................. 34
  §2.4.1 New Wave Moral Semantics ...................................................................................... 35
  §2.4.2 General Strategy .......................................................................................................... 37


§2.4.3 Putnam’s Twin Earth................................................................. 38
§2.4.4 Putnam and the Open Question Argument................................. 39
§2.5 Moral Twin Earth....................................................................... 39
§2.5.1 The Open Question Argument Revised...................................... 41
§2.5.2 Generalizing the Revised Open Question Argument...................... 42
§2.5.3 How to Explain Moral Twin Earth............................................. 42
§2.6 Meaning Twin Earth................................................................... 43
§2.6.1 Robinson Crusoe.................................................................... 44
§2.6.2 How to Explain Meaning Twin Earth......................................... 45
§2.6.3 Meaning and Normativity .......................................................... 46
§2.7 Conclusion.................................................................................. 46

Chapter Three: Plunkett and Sundell on ‘Metalinguistic Negotiation’................. 47
§3.1 Introduction .............................................................................. 47
§3.2 Plunkett and Sundell: Shared Meaning and Disagreement ................. 47
§3.3 Hare: The Missionary and the Cannibals...................................... 49
§3.4 Smith: Objectivity and the Pitfall of Moral Relativism....................... 50
§3.5 Horgan and Timmons: Moral Twin Earth....................................... 51
§3.6 Plunkett and Sundell: Disagreement and Literally Expressed Content .... 53
§3.6.1 Some Disagreements are Non-Canonical Disputes....................... 53
§3.6.2 Metalinguistic Disputes............................................................ 55
§3.6.3 Metalinguistic Negotiation......................................................... 57
§3.6.4 Content and Character, Sense and Reference............................. 58
§3.6.5 Negotiation Over Character, not Content .................................. 59
§3.6.6 Normative Disputes................................................................. 62
§3.6.7 What counts as ‘torture’?............................................................ 63
§3.7 Back on Twin Earth.................................................................... 65
§3.8 Meaning Twin Earth................................................................... 67
§3.9 Why Metalinguistic Negotiation Can’t Explain Meaning Twin Earth .... 69
§3.10 Conclusion................................................................................. 71

Chapter Four: Copp ‘Milk, Honey, and The Good Life on Moral Twin Earth’ ......... 72
§4.1 Introduction .............................................................................. 72
§4.2 Milk, Honey, and The Good Life on Moral Twin Earth .................... 72
§4.3 Brief Review ............................................................................. 72
§4.4 Copp’s First Reply: Accommodating Moral Twin Earth ................... 73
§4.4.1 First Response: Can We Dispose of the Principle that Meaning
Determine’s Reference? .................................................................... 75
§4.4.2 Second Response: Disagreeing for ‘Most Practical Purposes’? ......... 76
§4.5 Copp’s Second Reply: ‘A Theory of Error’ .................................... 77
§4.5.1 Fallback Conditions and Referential Intentions .......................... 78
§4.5.2 Back to Twin Earth ................................................................ 80
§4.5.3 Making Room for Error............................................................. 81
Chapter Five: Merli ‘Return to Moral Twin Earth’

§5.1 Introduction ................................................................. 93
§5.2 Describing Moral Twin Earth ................................................ 93
§5.3 Merli’s Dilemma for H&T ..................................................... 94
§5.4 Merli’s First Objection ....................................................... 96
  §5.4.1 Merli’s Target: H&T’s Univocity Concession ....................... 96
  §5.4.2 Radical Differences ...................................................... 96
  §5.4.3 Meeting the Second Condition ....................................... 97
§5.4.4 Moral Twin Earth is Still a Problem, Even if Based on Familiar Worries ... 98
§5.4.5 How Can Realism Meet H&T’s Challenge? ........................... 99
§5.4.6 H&T Do Not Need to Make the Univocity Concession to NMR .......... 99
§5.4.7 Disagreement on ‘Earth’ and Revising the MTE Scenario ......... 100
§5.4.8 A Holistic Strategy for Preserving Univocity ....................... 102
§5.4.9 Picking a Suitable Theory of Reference .............................. 102
§5.4.10 Convergence and an End-of-the-Day Moral Theory ............... 103
§5.4.11 Two Advantages of Merli’s Strategy ................................ 104
§5.4.12 Why Merli’s Strategy is the Best Defence for NMR ............ 106
§5.4.13 Responding to the Second Part of Merli’s Objection ......... 107
§5.5 Merli’s Second Objection: “Revising Our Intuitions” ................ 109
  §5.5.1 Merli’s Second Objection: Step One ................................ 109
  §5.5.2 What if Earthers and Twin Earthers Change their Views? ........ 111
  §5.5.3 What if Earthers and Twin Earthers Don’t Change their Views? ... 112
  §5.5.4 The First Step Summed Up ......................................... 112
  §5.5.5 Merli’s Second Objection: Step Two.............................. 114
  §5.5.6 Why Merli’s Second Objection Cannot Succeed ................. 115
§5.6 Merli’s Third Objection: “Preserving (Nonmoral) Disagreement” .... 116
  §5.6.1 Can Moving the Disagreement Help NMR Defend Against H&T? ... 117
  §5.6.2 Is H&T’s Scenario Better Understood as a Practical Disagreement? 118
  §5.6.3 The “Last Ought Before Action” .................................... 119
  §5.6.4 Can H&T’s Challenge Be Revised to Target Broader Evaluative ...
Introduction

One topic of recent debate in the philosophy of language is whether meaning is normative. If an ascription of meaning is a statement of the form ‘Smith means *pigeon* by ‘pigeon’’, then the claim that meaning is normative amounts (roughly) to the idea that ascriptions of meaning are evaluative, prescriptive, or action-guiding, in a way that is loosely analogous to the normative nature of moral judgment. The view that meaning is normative has been popular and even uncontroversial since Ludwig Wittgenstein’s discussion of rule-following and meaning in the *Philosophical Investigations* (1953/2009), but more recently this view has come under scrutiny in the wake of the sceptical argument that Saul Kripke extracts from Wittgenstein in his *Wittgenstein on Rules and Private Language* (1982). According to Kripke, we can extract from Wittgenstein a sceptical argument for the conclusion that there are no facts about meaning. Some critics have attempted to rebut this argument and its apparently intolerable conclusion by abandoning the thesis that meaning is normative, given the role this thesis plays in Kripke’s defence of the argument. We will call these critics ‘anti-normativists’. According to them, we can capture the intuitive notion of meaning without invoking the normativity of meaning and thus avoid Kripke’s unpalatable conclusion that there are no meaning facts. Normativists have argued that we cannot abandon the idea that meaning is normative without causing serious damage to the notion of meaning in general.

This dissertation attempts to make a modest contribution to the debate over the normativity of meaning by providing some new evidence in favour of the claim that meaning is normative. I do this by developing an analogue of the ‘Moral Twin Earth’ argument proposed by Terence Horgan and Mark Timmons against certain forms of ethical naturalism. I call this the ‘Meaning Twin Earth’ argument and claim that our intuitions about a so-called Meaning Twin Earth scenario are best explained by the thesis that meaning is normative. I frame this as a kind of explanatory challenge against the anti-normativist, to explain the intuitions elicited from the Meaning Twin Earth scenario without invoking the normativity of meaning.

Chapter Outline

1. Kripke’s Sceptical Paradox

   This chapter outlines the sceptical argument developed by Kripke’s Wittgenstein (KW) that there are no meaning facts. KW is able to sustain this argument against any putative candidate for a meaning fact, by assuming that meaning is normative. I highlight the role that the normativity thesis plays in KW’s argument by outlining KW’s reasons for rejecting the various possible responses to the sceptical argument. I delineate two interpretations of the idea that meaning is
normative (as it appears in KW’s argument), which we will call ‘Normativity’ and ‘Norm-Relativity’, and I will explain how these relate to the views defended by normativists and antinormativists.

2. Twin Earth and the Open Question Argument

One of the most formidable responses to KW’s argument, and the one that KW spends the most time in addressing, is a dispositionalist account of meaning, according to which meaning facts can be reduced to dispositional facts, i.e. facts about how speakers are disposed to apply expressions, especially under idealized conditions. KW’s response to this argument can be viewed as an analogue of G.E. Moore’s open question argument applied in the case of meaning. Ultimately, however, this argument cannot be convincingly applied against modern reductive dispositionalist accounts of meaning, given that these accounts view the reduction of meaning facts to dispositional facts as being a posteriori in the same manner of the identification of water with H₂O. Nonetheless I argue that a revised open question argument may pose a genuine threat to these reductive dispositionalist accounts of meaning. I attempt to show this by developing an analogue of the Moral Twin Earth argument advanced by Horgan and Timmons (1992a). I claim that a Meaning Twin Earth argument yields intuitions that are apparently at odds with both reductive dispositionalism and anti-normativism about meaning.

3. Plunkett and Sundell on ‘Metalinguistic Negotiation’

In this chapter I motivate a recent objection to the Moral Twin Earth argument advanced by David Plunkett and Timothy Sundell (2013) and situate this objection with respect to my Meaning Twin Earth argument. Plunkett and Sundell claim that Horgan and Timmons’ argument depends on the faulty assumption that speakers who disagree must mean the same thing by the words they use to express that disagreement. According to Plunkett and Sundell, genuine disagreement is not always best explained in this way: according to them, some genuine disagreements are non-canonical disputes—disputes in which speakers do not mean the same thing by the words they use in expressing conflicting contents. Plunkett and Sundell introduce the notion of metalinguistic negotiation in order to explain a special subset of these disputes and claim that Moral Twin Earth is one such example. If so, this would enable an objector to explain the Moral Twin Earth scenario without abandoning a naturalist semantic view or saying anything about the normativity of moral terms such as ‘morally right’. I argue that even if Horgan and Timmons’ argument can be defused via the invocation of metalinguistic negotiation, this same line of objection cannot be successfully applied against the Meaning Twin Earth argument.

4. Copp ‘Milk, Honey, and the Good Life on Moral Twin Earth’

In this chapter I discuss two objections against the Moral Twin Earth argument proposed by David Copp (2000) and claim that neither objection poses a threat to the Meaning Twin Earth
argument. Copp’s first objection claims that we can view the Earther and Twin Earther of Horgan and Timmons’ scenario as if they disagree for all practical purposes, but this objection cannot succeed for two reasons. First it disposes of something like the principle that sense determines reference, something that prima facie any account of meaning ought to preserve; and second, it does not enable us to view the Earther and Twin Earther as actually disagreeing. Copp’s second objection claims that a broader interpretation of Putnam’s work enables the naturalist to escape Horgan and Timmons’ charge that naturalism cannot capture the idea that the Earther and Twin Earther are genuinely disagreeing. According to Copp, the Earther and Twin Earther can be viewed as meaning the same thing and as genuinely disagreeing on account of their having the same referential intentions in using their respective moral terminologies. I argue that even if Copp’s strategy can be deployed against the Moral Twin Earth argument, it cannot be used against the Meaning Twin Earth argument.

5. Merli ‘Return to Moral Twin Earth’

In this chapter I address three objections against the Moral Twin Earth argument from David Merli (2002) and argue that each of these objections fails. In the first objection, Merli claims that an underdescription of the Moral Twin Earth argument enables Horgan and Timmons to conclude that naturalistic moral realism cannot explain the intuition that the Earther and Twin Earther of the Moral Twin Earth scenario genuinely disagree. Merli argues that once we attempt to fill in these details, it turns out that the Earthers and Twin Earthers can be viewed as meaning the same thing and as genuinely disagreeing in spite of the fact that their moral terms refer to distinct natural (functional) properties. Merli claims that this second component is a key concession in Horgan and Timmons’ argument. According to him, this concession demonstrates that the Moral Twin Earth argument is not a new challenge to realism. I object that the challenge is still a formidable one even if it is familiar, and that further, Horgan and Timmons can subtly alter their scenario to avoid Merli’s objection or even drop the concession that realism can capture univocity and disagreement on Earth. In the second objection, Merli assumes that Horgan and Timmons do drop this concession, but nonetheless Merli says, realism has good prospects for preserving univocity by its own lights given that realism can address this issue holistically, i.e. by appealing to similar problems in moral epistemology. The key reason for optimism on this front, Merli says, is that realists can count on a convergence in moral theory and so disputes like the one between the Earther and Twin Earther can be viewed as genuine disagreements if we construe reference as determined by the judgments the Earther and Twin Earther converge on at the end of inquiry. I argue that this is unreasonably optimistic and so Merli’s second objection cannot succeed because there are some disputes in which we cannot reasonably expect convergence in the relevant sorts of judgment. Merli introduces the idea that we should defer to idealized moralisers in these cases as well, but I argue that we can only speculate about what these idealized moralisers would judge. As such, this suggestion is likewise no help to Merli’s second objection.
In the third objection, Merli argues that we can view the disagreement between the Earther and Twin Earther of the Moral Twin Earth scenario as a kind of practical disagreement about what to do, something that can be explained via an expressivist analysis of all-in endorsement. I claim that this objection cannot succeed on account of two issues: First, the disagreement between the two parties cannot be viewed as a practical disagreement about what to do rather than a strictly-speaking moral disagreement about what is right. Second, if naturalistic moral realism must bring in expressivism to explain the disagreement between the Earther and Twin Earther, this leaves naturalistic moral realism facing a dilemma: it is either unmotivated or unable to meet the Moral Twin Earth argument. Either way, Merli fails to make a convincing case against the Moral Twin Earth argument and by proxy, none of these objections are available to damage the Meaning Twin Earth argument.

6. In What Sense is Meaning Normative?

I outline in broad terms two candidates for a normativist account of semantic judgment that would explain the intuitions elicited from the Meaning Twin Earth scenario. First, I discuss how a theory of this sort could be developed in the moral case by discussing one of R.M. Hare’s central arguments for a prescriptivist metaethical theory in his 1952. I then sketch out the two main options for a theory of normative judgment as outlined by Smith (1994), which are an expressivist form and a rationalist form. I then discuss how a theory such as Hare’s could be applied in the semantic case, as Gibbard (2012) attempts. I respond to two objections by Baker (2016) to this approach, including a defence of the application of the Moral Twin Earth argument to the case of meaning. I consider two candidates for a robust normativist theory of semantic judgment by looking to Gibbard’s non-factualist account of meaning and McDowell’s non-reductive realist account of meaning. I come to the provisional conclusion that McDowell’s non-reductive realist account stands to fare better in going forward. Granted, McDowell’s account does face formidable epistemological and metaphysical challenges, but Gibbard’s account faces a debunking objection from a generalised version of Kripke’s sceptical argument.

Overview

Overall, the thesis attempts to provide some new evidence in favour of the claim that meaning is normative, and suggests, provisionally, that a form of non-reductive normative realism offers the best hope of capturing this claim. However, this is a large and highly controversial area of inquiry, spanning as it does issues in both metaethics and the philosophy of language. In the conclusion of the thesis, I reiterate the modest nature of its aspirations and some of the limitations to which it is inevitably subject.
Chapter One: Kripke’s Sceptical Argument

§1.1 Introduction

In this chapter I outline the sceptical argument from Saul Kripke’s 1982, according to which there are no facts in virtue of which ascriptions of meaning, such as “Ingrid means bellbird by ‘bellbird’” are true or false. This argument proceeds in three stages: The first stage introduces a challenge via a so-called ‘sceptical paradox’ against the intuitive assumption that there is some fact about me that determines whether I meant, for example, the function of addition or some other non-standard function in my past use of the ‘plus’ sign. The second stage places a set of constraints on any adequate account of a meaning fact that would secure my meaning addition by ‘plus’ and thus block the hypothesis that I really meant some non-standard function. The third stage enumerates all of the potential candidates for a meaning fact that could potentially meet these constraints, but finds all of these candidates wanting. As such, it is concluded that there are no facts of the relevant kind, i.e. there are no facts in virtue of which ascriptions of meaning are true or false.

In the following sections I detail each stage of this argument and emphasize the importance of one key premise—the thesis that ‘meaning is normative’. Kripke’s original exposition is unclear about what this thesis amounts to. I draw on Hattiangadi (2007) to tease apart at least two conceptions of the normativity of meaning. I then explain how Kripke’s defence in the third stage of the argument hinges on the strongest conception of that thesis in order to rule out any putative meaning fact offered to meet the sceptical challenge. The primary line of objection against Kripke’s argument that I discuss at length here, as Kripke does in his exposition, is suggested by the proponent of semantic dispositionalism—the view that there are facts about meaning, constituted by a speaker’s dispositions to apply his or her expressions in a determinate pattern.

§1.2 Kripke’s Wittgenstein and the Sceptical Paradox

In Wittgenstein on Rules and Private Language (1982), Kripke develops an argument that aims to show that there are no facts in virtue of which ascriptions of meaning, such as “Jemma means addition by ‘plus’”, are true or false. Kripke takes this argument to be one possible upshot of Wittgenstein’s reflections on rule-following in the Philosophical Investigations. This version
of Wittgenstein, of “Wittgenstein as it struck Kripke”, we will call ‘Kripke’s Wittgenstein’ given that this view is at least inspired by Wittgenstein’s remarks on the topic.¹

Kripke’s Wittgenstein—hereafter referred to as ‘KW’—motivates his sceptical argument by using a mathematical example. Suppose that in the past I have never computed the sum for an arithmetical query involving numbers greater than or equal to 57.² Though I have never performed this particular operation, I know what to do with the two integers flanking the symbol ‘+’ given my grasp of the meaning of that symbol, i.e. as the symbol that denotes the function addition. To know what the symbol ‘+’ or the term ‘plus’ means is to know that addition is the correct function to perform for any two integers conjoined with the plus symbol.³ So when queried with ‘68 + 57 = ?’, I add 68 and 57, double check my work, and report the sum—‘125’.⁴

Suppose that a bizarre sceptic approaches and challenges my answer ‘125’ to the query ‘68 + 57 = ?’. The sceptic claims that I now misunderstand my past usage of ‘plus’ and that I should have answered ‘5’. I should have answered this way, the sceptic explains, because ‘plus’ as I used it in the past actually referred to a non-standard function that he refers to as ‘quaddition’ where quaddition is an arithmetical function continuous with addition for any two integers, unless one or both of those integers are greater than or equal to 57, in which case performing quaddition yields the answer ‘5’. So, for any x and y less than 57, performing addition and quaddition will yield the same answer. That is, the sum and the quum for any such x and y will be the same. For any x and y greater than or equal to 57, performing quaddition (unlike addition) will always yield the answer ‘5’. This function can be defined more formally, using the sign ‘⊕’ to denote quaddition in the following way:

\[
x ⊕ y = \begin{cases} x + y & \text{if } x, y < 57 \\ 5 & \text{otherwise} \end{cases}
\]

The sceptic’s hypothesis is admittedly bizarre, but not logically impossible—for if we assume that my past computations involved only integers less than 57, it is compatible with my past use of the ‘plus’ sign that I actually meant quaddition all along: “Perhaps when I used the term ‘plus’ in the past, I always meant quus: by hypothesis I never gave myself any explicit directions that were incompatible with such a supposition” (1982, p.13). This is possible, according to KW’s sceptic, because “no fact about my past history—nothing that was ever in my mind, or in my external behaviour—establishes that I meant plus rather than quus” (1982, p.13).

---

¹ I expand on how KW’s reading departs from the standard interpretation of Wittgenstein in discussing John McDowell’s account of meaning in Chapter Five.
² There must be some such example given the finitude of computations I have performed in the past.
³ KW uses ‘+’ and ‘plus’ interchangeably. Hereafter I stick to the term ‘plus’ for clarity.
⁴ It is important to note that ‘125’ is the correct answer to ‘68 + 57 = ?’ in two senses: First, it is correct in the arithmetical sense that 125, as a matter of mathematical fact, is the sum of 68 and 57. Second, it is correct in the metalinguistic sense that the ‘plus’ symbol or ‘+’ as it appears in ‘68 + 57 = ?’ actually refers to the function of addition (Miller 2002, p.3). Miller points out that these two senses of correctness could in principle come apart—e.g. “if the ‘+’ sign really stood for the subtraction function, 125 would still be the sum of 68 and 57, but the correct answer to the question ‘68 + 57 = ?’ would now be ‘11’” (Miller 2002, p.3). KW is only concerned with the latter.
If there is no fact that secures *addition* as the unique function governing my past use of the term ‘plus’, the sceptic argues, there is likewise no fact that secures *addition* as the unique function governing my present use of that term (1982, p.13). Consequently, there is nothing that determines that ‘125’ and not ‘5’ is the answer I ought to give to the query ‘68 + 57 = ?’. So given that nothing in my past dictates the answer ‘125’ as the one I ought to give presently, then according to KW, we are forced toward a devastating sceptical conclusion: “When I respond in one way rather than another to such a problem as ‘68 + 57 = ?’, I can have no justification for one response rather than another . . . There can be no fact as to what I mean by ‘plus’, or any word at any time” (1982, p.21).

§1.2.1 Ground Rules

The sceptical challenge can only get off the ground, KW says, on account of a few basic rules. First, the sceptic must share a language with me, for otherwise he could not converse with me at all (1982, p.11-12). Second, the sceptic cannot challenge my present use of ‘plus’: “he agrees that, according to my present usage, ‘68 plus 57’ denotes 125” (1982, p.12). In fact, the sceptic, we shall suppose, “conducts the entire debate with me in my language as I presently use it” (1982, p.12). Third, the sceptic grants that “the accuracy of my computation nor of my memory is under dispute” (1982, p.11). The sceptic does not deny the *mathematical fact* that 125 is the sum of 68 and 57; instead he denies the existence of any *metalinguistic fact* that would determine whether my present use of the ‘plus’ sign conforms to my past use of that sign. If ‘plus’ as I used it in the past actually refers to *quaddition*, then ‘5’ would be the correct answer to ‘68 + 57 = ?’ in light of that past linguistic intention. If this hypothesis is correct, then according to KW, “the concepts of meaning and of intending one function rather than another will make no sense” (1982, p.13). In order to block the sceptic’s hypothesis, we need some account of my meaning, in terms of some fact about my past self, that justifies my response ‘125’: “Ordinarily, I suppose that, in computing ‘68 + 57’ as I do, I do not simply make an unjustified leap in the dark. I follow directions I previously gave myself that uniquely determine that in this new instance I should say ‘125’. What are these directions?” (1982, p.10).

§1.2.2 Generalizing the Sceptical Challenge

Though the challenge advanced by KW’s sceptic is directed against my past use of the ‘plus’ sign, that challenge extrapolates to language generally—a potentially infinite number of sceptical challenges could be drawn against the past use of any expression. The sceptic can, for example, question my past usage of the word ‘green’ as it applies to green-coloured things. Perhaps by ‘green’, I always meant grue, where x is grue if and only if t < t* and x is green or t ≥ t* and x is blue. If by ‘green’ I meant grue, then I ought to answer ‘green’ to the query “What colour is the sky?” if the query is made at time t ≥ t*. If there is no fact that secures my meaning grue or green by ‘green’, then we are left at the same impasse as with ‘plus’. If no answer can be
given to the sceptic for any expression, then there is no fact as to what anyone means by anything. The whole enterprise of language seems to “vanish into thin air” (1982, p.22).

§1.3 Two Constraints on Any Account of Meaning

Any suitable response to the sceptic’s challenge, as an account of my meaning *addition* by ‘plus’, must satisfy two constraints according to KW. First, it must provide an account of what *constitutes*, in terms of some fact about my past self, what I meant by the ‘plus’ sign. Second, it must explain how that fact about my past meaning informs my present use of that sign, i.e. *justifies* the answer ‘125’ to the query ‘68 + 57 = ?’. The first constraint demands some fact that would isolate *addition* as the unique function governing my past use of the ‘plus’ sign. At minimum, this would block the sceptical hypothesis that I meant quaddition all along. We can call this constraint the ‘extensionality constraint’ given that it is grounded in the principle that the meaning of an expression determines the class of things to which that expression can be correctly applied (and those to which it is incorrectly applied). The second constraint makes a further demand on any candidate fact that apparently meets the first: any meaning fact must explain how my past linguistic intention with respect to the ‘plus’ sign dictates or normatively compels the present answer ‘125’ as the only answer I ought to give when prompted with ‘68 + 57 = ?’. This constraint serves as a kind of normative requirement insofar as it calls for an account of how some fact about my past meaning bears a *normative relation* to my future behaviour in this way, i.e. any fact there is about my past use of the term ‘plus’ must prescribe how I ought to go on in using that expression, i.e. We will refer to this constraint as the ‘normativity constraint’ given that this constraint is grounded in the principle that the meaning of an expression prescribes how a speaker ought to use that expression. In the next two sections, I spell out the above two constraints in more detail and explain how they bear on any account of meaning offered in response to KW’s sceptical argument.

§1.3.1 The Extensionality Constraint

The extensionality constraint depends on the platitude that the meaning of a term determines how that term is correctly applied, i.e. by setting a standard according to which uses of that expression count as correct and incorrect. The meaning of the term ‘plus’ for instance dictates that certain answers to ‘plus’ queries are correct and others incorrect: e.g. ‘125’ is the correct answer to ‘57 + 68 = ?’ since the ordered triple <57, 68, 125> belongs to the extension of the function denoted by ‘plus’, while the answer ‘5’ is incorrect since the ordered triple <57, 68, 5> does not so belong. Moreover, if I were to use ‘plus’ to apply to *addition* I would use that term correctly while if I used it to apply to *quaddition* I would use it incorrectly. Likewise, if I were to apply ‘plus’ to *division* or *subtraction*, I would do something incorrect—*division* and *subtraction* are not referred to by ‘plus’ either. The sceptic’s challenge puts this intuitive thesis about meaning in danger. This can be brought home by subtly altering the way the sceptic’s charge is delivered.
The sceptic effectively argues that there is no fact that can be cited about my past self that determines what standard governs my use of the term ‘plus’ and thus my responses to arithmetical queries in which ‘plus’ appears. Given that there is no such fact, the sceptic argues, there is no sense to be made of the idea that I can accord or fail to accord with my past meaning. No standard dictates that answering ‘125’ would be the correct way for me to respond to the query ‘68 + 57 = ?’ and so, as per the sceptic’s hypothesis, the answer ‘5’ would be just as “correct” as ‘125’. Nothing secures the ‘addition standard’ rather than a deviant ‘quaddition standard’ as the one that determines the extension of the term ‘plus’ as I understand it (and therefore the correctness of answers to queries in which ‘plus’ appears). If the sceptic is right, then the very idea of my making a mistake is ruled out. In order to head off the sceptic’s challenge, there must be some fact that rules out the bizarre hypothesis that my use of the term ‘plus’ in the past referred to quaddition. After all, it is because I intend to conform to my past use of the term ‘plus’, as it is used to refer to the addition function, that I take ‘125’ to be the only correct answer to ‘68 + 57 = ?’. This first constraint is closely connected with the second in terms of how they each bear on any adequate account of meaning as a whole.

§1.3.2 The Normativity Constraint

Suppose for a moment that some fact is cited which apparently meets the extensionality constraint. That is, there is a fact in virtue of which the judgement ‘Jon means addition by ‘plus’’ would be true. According to KW, any fact specified on account of the first constraint must answer to a second and further constraint in order to block the sceptic’s hypothesis: Any account of a fact about my past meaning must explain how that fact bears a normative relation to my present behaviour in the sense that it justifies or normatively compels the answer ‘125’ to ‘68 + 57 = ?’. Otherwise, if my past meaning does not prescribe ‘125’ as the only answer I ought to give, then I might as well say ‘5’ just the same—the sceptic’s claim that I should have said ‘5’ would be equally justified! In a nutshell, any fact about my meaning with respect to the ‘plus’ sign must be essentially ‘normative’ in the action-guiding sense of that term, i.e. it must dictate how I ought to go on in using that sign or expression. So supposing that I intended to use the ‘plus’ sign to mean addition, there is only one thing I ought to do when prompted with ‘68 + 57 = ?’: to respond with the answer ‘125’. This rules out the sceptic’s suggestion that ‘5’ would have been equally appropriate in light of my past meaning. In summary, then, any adequate account of my meaning addition by ‘plus’ must guarantee the truth of two judgments: First, it must be true that ‘125’ is the only answer that would be correct in virtue of my past meaning. Second, it must be true that ‘125’ is the only answer I ought to give. If both of these constraints can be met, then we can block the sceptic’s hypothesis about a change in my use of the ‘plus’ sign.

§1.4 Scope of the Sceptical Challenge

According to KW, there are no limitations to the kind of fact that can be cited to answer the sceptical challenge. In particular, KW says, there are no behaviourist limitations on such a
fact (14). That is, any fact cited about my meaning plus rather than quus by the ‘plus’ symbol need not be something available to an external observer. KW allows that there could be some fact about my internal mental state that would satisfy the sceptic. So even if nothing in my external behaviour can be shown to demonstrate that I meant plus rather than quus in using the ‘plus’ sign, something in my inner state might nonetheless do so (1982, p.14). As KW points out, this puts some distance between his challenge and Quine’s argument concerning the ‘indeterminacy of translation’ because the evidence permitted to satisfy KW’s sceptical challenge is, unlike Quine’s case, not limited to behavioural evidence (1982, p.14). The sceptical challenge does not consist in a behaviouristic doubt about whether there is something about me in virtue of which an outside observer could judge what I meant by the ‘plus’ sign. Instead, KW explains, the doubt is delivered from the “inside” (1982, p.15):

Whereas Quine presents the problem about meaning in terms of a linguist, trying to guess what someone else means by his words on the basis of his behavior, Wittgenstein's challenge can be presented to me as a question about myself: was there some past fact about me - what I ‘meant’ by plus - that mandates what I should do now? (1982, p.15)

KW argues that there is no such ‘solution’ to the paradox in terms of some fact—even one concerning a mental state of mine—that would head off the sceptic’s challenge.

§1.5 Two Ways to Respond to the Sceptical Challenge

KW argues that there are two broad ways to respond to the sceptical challenge. First, one could offer a ‘straight solution’ to the argument by pointing out a flaw in the line of reasoning that led to its conclusion that there are no meaning facts or by producing a fact that satisfies the two constraints imposed by KW. Alternatively, one could offer a ‘sceptical solution’ to the argument by conceding its conclusion and then rekindling the notion of meaning in a way that does not depend on the existence of meaning facts. KW takes his sceptical argument to succeed, naturally, and so he concludes that there are no meaning facts of the sort that would meet both sceptical constraints. As such, KW claims that we should abandon meaning factualism and turn to a sceptical solution, built on some non-factualist account of meaning. KW endorses a solution of this kind—one that he takes Wittgenstein to adopt in the wake of the sceptical paradox of Investigations §201—according to which we can preserve the notion of meaning by viewing ascriptions of meaning, such as “Simon means bellbird by ‘bellbird’” as assertable under certain conditions:

All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives. No supposition that 'facts correspond' to those assertions is needed. (Kripke 1982, p.77-78)

5 See Quine Word and Object (1960).
6 In this respect, a straight solution to the paradox is a form of meaning factualism.
In Chapter Six, I argue that non-factualist ‘solutions’ of this sort, including KW’s, are unlikely to have any more traction than the factualist alternatives that we will consider later in this chapter. For now, I set this topic aside and turn to KW’s defence of the sceptical challenge against any putative straight solution to that challenge.

§1.6 Candidate Facts

In the following sections, I enumerate several candidates for a meaning fact, as they are discussed in KW, and show how each candidate apparently fails to meet the sceptical challenge. Perhaps the most formidable candidate and the one that we will discuss at length in this dissertation is dispositionalism.

§1.6.1 Past Behaviour or Mental History

The first candidate that KW considers is the suggestion that some fact about my past behaviour would meet the sceptical challenge. KW rejects this response because: (1) by stipulation, I have never computed the sum for any integers greater than or equal to 57; and (2) ‘plus’ and ‘quus’ have the same extension for any two integers less than 57. Therefore, KW says, my past behaviour is perfectly consistent with the hypothesis that I meant quaddition by the ‘plus’ sign. Even if we were to enlarge the pool of past behaviour this would not make any difference:

... no matter how [the pool of behaviour] is enlarged, a “deviant” interpretation of “+”, such as that which takes it as standing for the quaddition function, will always be possible; even if we enlarge the pool of previous behaviour so that we have encountered numbers larger than 57, there will always be some number which is larger than those we have previously encountered, and the skeptic can use this to construct an analogue of the quaddition interpretation. (Miller 2002, p.5)

§1.6.2 Set of Instructions or General Rule

The second candidate KW considers is the suggestion that a set of instructions or a general rule that I gave myself in the past would meet the challenge:

Many readers . . . protest that our problem arises only because of a ridiculous model of the instruction I gave myself regarding ‘addition’. Surely I did not merely give myself some finite number of examples, from which I am supposed to extrapolate the whole table (“Let ‘+’ be the function instantiated by the following examples: . . . “). No doubt infinitely many functions are compatible with that. Rather I learned - and internalized instructions for - a rule which determines how addition is to be continued. What was the rule? Well, say, to take it in its most primitive form: suppose we wish to add x and y. Take a huge bunch of marbles. First count out x marbles in one heap. Then count out y marbles in another. Put the two heaps together and count out the number of marbles in the union thus formed. The result is x + y. This set of directions, I may suppose, I explicitly gave myself at some earlier time. It is engraved on my mind as on a slate. It is incompatible with the hypothesis that I meant quus. It is this set of directions, not the finite list of particular additions I performed in the past, that justifies and determines my present response. (1982, p.15)

KW rejects this response because the sceptical challenge can be readily reconfigured to target my understanding of the word ‘count’. Given that I have applied ‘count’ on finitely many occasions, the sceptic can argue that by ‘count’ I really meant quount where to quount a heap “is to count it
in the ordinary sense, unless the heap was formed as the union of two heaps, one of which has 57 or more items, in which case one must automatically give the answer ‘5’ (1982, p.16). So if ‘counting’ really amounted to quounting, then even if I were to follow the general rule above with respect to ‘plus’, I would be forced to admit that ‘5’ would be the correct answer to ‘68 + 57 = ?’ (1982, p.16). We have no further traction here in citing a general rule or set of instructions because “the sceptic can always respond by giving a deviant interpretation of the symbols of the general thought or instruction itself” (Miller 2002, p.5). Moreover, this point can be generalized: “any set of instructions that come before the mind require interpretation as much as the linguistic expression whose understanding they are supposed to facilitate, and are thus as susceptible to deviant interpretation as that original expression” (2002, p.5-6). At this stage, we might be tempted to bring in a set of instructions for interpreting the instructions here, but this will push us toward an infinite and fruitless regress (2002, p.6). KW takes this to be the lesson of Wittgenstein’s discussion of a “rule for interpreting a rule” (1982, p.17):

It is tempting to answer the sceptic by appealing from one rule to another more ‘basic’ rule. But the skeptical move can be repeated at the more ‘basic’ level also. Eventually the process must stop - “justifications come to an end somewhere” - and I am left with a rule which is completely unreduced to any other. How can I justify my present application of such a rule, when a sceptic could easily interpret it so as to yield any of an indefinite number of other results? (1982, p.17)

In other words, KW says:

To say that there is a general rule in my mind that tells me how to add in the future is only to throw the problem back on to other rules that also seem to be given only in terms of finitely many cases. (1982, p.22)

Consequently, KW concludes that any conception of a set of instructions or some general rule cannot meet the sceptical challenge.

§1.6.3 Dispositionalism

One way to avoid the above problems would be to abandon the idea that a meaning intention consists in an *occurrent* mental state (one that ‘comes before the mind’) and opt instead for a *dispositional* account of my meaning something by an expression such as ‘plus’. KW replies to two dispositionalist responses of this sort. The first is embodied by a view that we will call ‘simple dispositionalism’ and the second, a more nuanced version of this view, ‘sophisticated dispositionalism’.

§1.6.4 Simple Dispositionalism

Dispositionalism is the thesis that what a speaker means by an expression is constituted by his or her disposition to apply that expression in a determinate pattern, e.g. to apply a term to some objects and not others, to apply ‘penguin’ only to penguins and not sharks. According to the dispositionalist, what determines that ‘plus’ as I used it in the past means *addition* is a disposition to respond with the unique sum for any x and y conjoined with the addition symbol:
To mean addition by ‘+’ is to be disposed, when asked for any sum ‘x + y’ to give the sum of x and y as the answer (in particular, to say ‘125’ when queried about ‘68 + 57’).

(1982, p.22)

The dispositionalist argues that it is the possession of this disposition that rules out the sceptical hypothesis that reporting the answer ‘5’ would be the correct way for me to respond when prompted with ‘68 + 57 = ?’. The only correct way for me to respond, if indeed I possess the disposition cited above, is to report the answer ‘125’. Given that I possessed this disposition in the past I would likewise have answered ‘125’ if previously queried with ‘68 + 57 = ?’:

True, my actual thoughts and responses in the past do not differentiate between the plus and the quus hypotheses; but, even in the past, there were dispositional facts about me that did make such a differentiation. To say that in fact I meant plus in the past is to say - as surely was the case! - that had I been queried about ‘68 + 57’, I would have answered ‘125’. By hypothesis I was not in fact asked, but the disposition was present none the less. (1982, p.23)

Any like query involving two integers conjoined with the ‘plus’ symbol will, given my disposition, be met by my reporting their sum and not their quum.

§1.6.5 KW’s Finitude Objection

KW’s first objection against simple dispositionalism states that dispositionalism cannot meet the extensionality constraint on any adequate account of a meaning fact, given that the finite disposition of a speaker cannot determine the correct application of an expression across a potentially infinite number of future cases. This objection can be called the ‘finitude objection’, given that this argument is based on the principle that the meaning of an expression should determine its extension in a potentially infinite number of cases conjoined with the fact that a speaker, as a finite being, has a finite disposition for applying a given expression in the future. The former premise, which we can call the ‘principle of extension’, states that the meaning of an expression determines its extension in the sense that it determines the class of things to which that expression correctly applies. For example, the word ‘turtle’ is correctly applied to turtles, and not crocodiles, seagulls, etc. Consequently, ‘turtle’ is applied incorrectly when it is applied to things that do not fall under the extension of that word. The problem for dispositionalism is that a speaker’s dispositions only determine how an expression is to be used in a finite number of cases, and so a dispositional analysis of meaning looks inadequate at first pass. As Boghossian puts it:

If I mean horse by ‘horse’, then there are literally no end of truths about how it would be correct for me to apply the term—to horses on Alpha Centauri, to horses in Imperial Armenia, and so on, but not to cows or cats wherever they may be—if I am to use it in accord with its meaning. But, Kripke argues, the totality of my dispositions is finite, being the dispositions of a finite being that exists for a finite time. And so, facts about dispositions cannot capture what it is for me to mean addition by ‘+’. (1989, p.509)

The meaning of an expression, as Boghossian explains, seems to dictate the correct application of that expression in an infinite number of cases, i.e. there is “literally no end of truths” in virtue of which an expression is correctly applied (1989, p.509). One way to bring out the inadequacy of the simple dispositionalist view is to note that all of the facts about my actual dispositions are
consistent with the hypothesis that I mean *skaddition* by ‘+’, where *x skadd y* is equal to *x plus y* for numbers *x* and *y* that are small enough for me to comprehend during my lifespan, but equal to 5 otherwise. Given this, a dispositional account of a meaning fact looks doomed to fail insofar as a speaker’s dispositions are always in this sense finite.

§1.6.6 KW’s Normativity Objection

KW’s second objection against a simple dispositionalist response to the sceptical argument states that dispositionalism fails the normativity constraint: Any adequate fact cited about my past meaning must *justify* the answer ‘125’ as the only answer I ought to give to the present query ‘68 + 57 = ?’. KW’s sceptic argues that my response ‘125’ “is no better than a stab in the dark” in the sense that ‘125’ is no more justified than ‘5’ (1982, p.23). According to him, nothing about my past meaning justifies ‘125’ over ‘5’ as the answer I should give to ‘68 + 57 = ?’. As such, the sceptic argues that if I were to report the answer ‘5’, this would be equally compatible with my past meaning. The way that the sceptic brings this out, of course, is to hypothesize that whenever I used the ‘plus’ symbol in the past I really meant *quaddition* and so I ought to answer ‘5’ to the current prompt ‘68 + 57 = ?’. Any adequate answer to the sceptic must block this hypothesis by giving an account of my meaning, in terms of some fact, that justifies the answer ‘125’. A dispositionalist analysis of my meaning, in terms of some fact about how I am disposed to use the ‘plus’ symbol, appears inadequate for this purpose; how I ought to respond for any arithmetical query involving the ‘plus’ symbol, on a dispositional analysis, simply collapses into however I am disposed to or will respond to any like query involving the ‘plus’ symbol (Miller 2007, p.172).

KW argues that dispositionalism is not equipped to meet the sceptic’s hypothesis. All that the dispositionalist has said is that at present I am disposed to answer ‘125’ and that perhaps I would have answered ‘125’ to the same query in the past. Therefore, it is unclear how a strictly dispositional account of my meaning could do anything more than, at best, describe how I am disposed to answer arithmetical queries, when what is required by KW’s sceptic is a fact that determines how I *ought* to answer such queries:

“‘125’ is the response you are disposed to give, and (perhaps the reply adds) it would also have been your response in the past.” Well and good, I know that ‘125’ is the response I am disposed to give (I am actually giving it!), and maybe it is helpful to be told - as a matter of brute fact - that I would have given the same response in the past. How does any of this indicate that - now or in the past - ‘125’ was an answer justified in terms of instructions I gave myself, rather than a mere jack-in-the-box unjustified and arbitrary response? Am I supposed to justify my present belief that I meant addition, not quaddition, and hence should answer ‘125’, in terms of a hypothesis about my past dispositions? (Do I record and investigate the past physiology of my brain?) Why am I so sure that one particular hypothesis of this kind is correct, when all my past thoughts can be construed either so that I meant plus or so that I meant quus? Alternatively, is the hypothesis to refer to my present dispositions alone, which would hence give the right answer by definition? (Kripke 1982, p.23)
As KW puts it, the dispositionalist looks committed to the idea that my present disposition with respect to the ‘plus’ sign is, by its very definition, a disposition to answer correctly with the sum for any arithmetic query involving that symbol. Consequently, KW argues, this dispositionalist response misses the mark by simply running together performance with correctness, where KW’s sceptic is precisely worried about whether my performance really does count as correct—whether ‘125’ really is justified by my past meaning.

Any plausible account of meaning must accommodate the idea that a speaker could fail to use an expression in accordance with how that expression ought to be applied. Dispositionalism fails this requirement insofar as it offers a descriptive account of a normative relation. Describing my disposition to report the sum for any arithmetical query involving ‘plus’ suffices as an account of what I will do, that I will answer ‘125’ to the query ‘68 + 57 = ?’, but that description does not suffice as an account of what I should do, that I ought to report the sum ‘125’ in virtue of the fact that it is the only answer justified by my past meaning:

Suppose I do mean addition by ‘+’. What is the relation of that supposition to the question of how I will respond to the problem ‘68 + 57’? The dispositionalist gives a descriptive account of this relation: if ‘+’ meant addition, then I will answer ‘125’. But this is not the proper account of the relation, which is normative, not descriptive. The point is not that, if I meant addition by ‘+’ I will answer ‘125’, but that I should answer ‘125’. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be disposed to respond as I should, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is normative and not descriptive. (Kripke 1982, p.37)

Dispositionalism fails to capture the normative relation between meaning and use. For this reason, dispositionalism fails as a response to the sceptical challenge.

The key problem for dispositionalism is that, on strictly dispositional terms, there is no room to explain how a speaker’s use of an expression could come apart from how that speaker ought to use that expression, in accordance with its meaning and the conditions under which it can be correctly applied. Dispositionalism is committed to running these two together and simply equates performance with competence (Kripke, 1982, p.24). This is problematic because the fact that a speaker is disposed to apply an expression in a certain pattern cannot itself guarantee that that pattern would be correct or uniquely justified in virtue of the meaning of that expression. Boghossian highlights this point in KW’s reply:

The point is that, if I mean something by an expression, then the potential infinity of truths that are generated as a result are normative truths: they are truths about how I ought to apply the expression, if I am to apply it in accord with its meaning, not truths about how I will apply it. My meaning something by an expression, it appears, does not guarantee that I will apply it correctly; it guarantees only that there will be a fact of the matter about whether my use of it is correct. (1989, p.509)

The primary lesson of KW’s second objection, on Boghossian’s view, is that one cannot ‘read off’ the correct use of an expression from a speaker’s disposition to use that expression: “to be disposed to use an expression in a certain way implies at most that one will, not that one should” use that expression in that way (1989, p.509). Boghossian continues:
Any theory which, like the crude dispositional theory currently under consideration, simply equates how it will be correct for me to use a certain expression with how I am disposed to use it, would have ruled out, as a matter of definition, the very possibility of error. And as Wittgenstein was fond of remarking, if the idea of correctness is to make sense at all, then it cannot be that whatever seems right to me is (by definition) right. (1989, p.531)

The problem here is further illustrated by the fact that I could after all be disposed to answer incorrectly. For example, I could be disposed to forget to carry, e.g. I might report ‘115’ when asked for the sum of ‘68 + 57 = ?’ by forgetting to carry the ‘1’ from right to left in a standard addition table. Given this, dispositionalism would imply, not that I am disposed to make mistakes, but that I actually mean some non-standard function skaddition obtained by deleting carrying from performing addition computations. Moreover:

... even if there were a suitably selected disposition that captured the extension of an expression accurately, the disposition could still not be identified with that fact of meaning, because it still remains true that the concept of a disposition is descriptive whereas the concept of meaning is not. In other words, according to Kripke, even if there were a dispositional predicate that logically covaried with a meaning predicate, the one fact could still not be identified with the other, for they are facts of distinct sorts. (Boghossian 1989, p.532)

As Boghossian points out, KW assumes that meaning facts would essentially amount to normative facts and so there could be no reduction of meaning facts to dispositional facts—facts that are only descriptive and thus non-normative in kind. One way to bring this out is by noting that, for example, the question “S is disposed to apply ‘bellbird’ to bellbird, but does S really mean bellbird by ‘bellbird’?” would be an open question in the sense that sincerely asking it is perfectly coherent. So even if we stipulate that S is disposed to apply ‘bellbird’ to all and only bellbirds and exhibits a flawless record of doing so, it seems that we can still coherently ask whether he actually means what we think he does. The property of being disposed to apply ‘bellbird’ to an object if and only if that object is a bellbird cannot, as a matter of conceptual necessity, be identical to the property of actually meaning ‘bellbird’ by bellbird. This is the heart of KW’s normativity objection, according to Boghossian, and it is a topic that we will return to at length in Chapter Two once we have a grasp on the open question argument as it appears in metaethics.7

§1.6.7 Sophisticated Dispositionalism

One way that a dispositionalist could address the objections outlined above would be to introduce the idea that under some ceteris paribus or ideal conditions a speaker would be disposed

---

7 In Chapter Two, I claim that, in line with Zalabardo (1997), this argument cannot succeed as an a priori and fully general objection against any dispositionalist account of meaning given that modern proponents of this view deny the claim that meaning facts are reducible to dispositional facts as a matter of conceptual necessity. Modern proponents of dispositionalism instead accept the more modest thesis that meaning facts are reducible, as a matter of empirical fact on a par with the a posteriori reduction of water to H2O. This blocks KW’s normativity objection as it is construed here. I argue in Chapter Two that a revised version of this argumentative strategy can potentially be used against modern dispositionalist theories of meaning by developing an analogue of the Moral Twin Earth argument proposed by Horgan and Timmons (1992).
to apply an expression in a certain pattern, e.g. one that corresponds to ‘plus’ referring to addition and no other function. This view we will call ‘sophisticated dispositionalism’. In the next sections I explain how this response to KW’s argument can apparently meet the preceding objections and then consider whether this response can stave off three subsequent objections: KW’s revived finitude objection, KW’s circularity objection, and Boghossian’s belief-holism objection.

We can elucidate the thesis suggested by the sophisticated dispositionalist by drawing an analogy with the fact that, for example, salt is roughly-speaking ‘disposed’ to dissolve when exposed to water under certain conditions:

... common salt possesses a disposition to dissolve when placed in water, but it will not dissolve in a sample of water which is already saturated; so to say that salt is water soluble is really to say that it is disposed to dissolve in water under certain conditions, one consequence of whose obtaining is that the water in question is not already saturated. (Miller 2007, p. 208)

If dispositionalism about meaning can be sketched in a similar fashion, i.e. by equating what a speaker means by an expression with how he or she is disposed to use it under ideal conditions, then a dispositionalist account of a meaning fact can avoid the two preceding objections against simple dispositionalism. The finitude objection can be met, according to Miller, by specifying the ideal conditions in such a way that rules out the thesis that ‘plus’ refers to skaddition:

The problem which the finiteness of our actual dispositions posed might then be avoided if we could include some specification in these conditions to the effect that the speaker lives long enough to hear out the relevant arithmetical enquiry: it seems plausible to say that if I were to live long enough, even for numerals “n” and “m” which are in fact too large for me to take in in my normal life span, I would, in response to an appropriate query, utter a numeral denoting the sum of the two numbers rather than “5”. This would rule out the interpretation of “+” as standing for the skaddition function. (2007, p. 208)

Likewise, the normativity objection can be met by specifying the ideal conditions with respect to my disposition in such a way that if certain ideal conditions obtain, then I would both be disposed to report the sum for any arithmetical query of the form ‘x + y = ?’ and be justified in doing so:

... if the conditions specified are genuinely ideal, or at least conditions of proper functioning, then doesn’t it follow from the fact that I would respond in a certain way under conditions of that type that I have a reason for responding in that fashion? For example, if someone in ideal conditions for appraising arithmetical claims were to say that a certain answer to an arithmetical problem was appropriate, and moreover, I know that the conditions are in fact ideal, doesn’t it follow that I ought to accept the verdict given on the arithmetical problem? (2007, p.208)

Moreover, this allows the scope for the distinction that KW calls for, between my performance with respect to using an expression and my competence or correctness in using that expression as I do:

Competence would be a matter of acting as one would act under ideal conditions; this could come apart from actual performance, in the cases where the ideal conditions fail to obtain. One could thus be systematically disposed to make a mistake, because the ideal conditions could systematically fail to obtain. (2007, p.208)
All that is left, of course, is to specify the set of ideal conditions in requisite way, i.e. in such a way that just in case these conditions obtain, then it would be true that I would be disposed to report the sum for any query of the form ‘x + y = ?’ (2007, p.208-9). Crucially, the specification of these conditions must abide the dispositionalist’s general project of reducing meaning facts to dispositional facts without referencing facts about meaning, otherwise:

... this would put the cart before the horse, since the story about how speakers are disposed to behave under those conditions is supposed to tell us what constitutes the facts about meaning. So, no use of semantic or intentional materials is to be made in the specification of the ideal conditions (2007, p.209).

A simpler example from Boghossian illustrates how such a specification could go for the dispositionalist.

Intuitively, “magpie” stands for magpies: bullfinches, sparrows, people, and tables do not fall within its extension. Suppose that Jones means magpies by the predicate “magpie”. The dispositionalist will say that this fact about Jones is constituted by the fact that he is disposed, under ideal conditions, to apply the predicate “magpie” to all and only those objects which are magpies. In other words, Jones’s meaning magpie by “magpie” is constituted by the fact that he is disposed, under ideal conditions, to token the belief that x is a magpie, if and only if x is in fact a magpie. Can Sophisticated Dispositionalism spell out a set of ideal conditions such that this is true, in a way which avoids the use of prior semantic and intentional materials? (Miller 2007, p.209)

KW and Boghossian argue that sophisticated dispositionalism cannot spell out an appropriate set of ideal conditions on the basis of three objections: the revived finitude objection, the circularity objection, and the belief-holism objection. The first two are proposed by KW and the third by Boghossian.

§1.6.8 KW’s Revived Finitude Objection

In his first pass at sophisticated dispositionalism, KW imagines how a dispositionalist could try to avoid the finitude objection outlined above by introducing a ceteris paribus clause that would guarantee my answering with the sum for any query of the form ‘x + y = ?’:

[H]ow should we flesh out the ceteris paribus clause? Perhaps as something like: if my brain had been stuffed with sufficient extra matter to grasp large enough numbers, and if it were given enough capacity to perform such a large addition, and if my life (in a healthy state) were prolonged enough, then given an addition problem involving two large numbers, m and n, I would respond with their sum, and not with the result according to some quus-like rule. (1982, p.27)

KW rejects this suggestion on account of an epistemic worry. How could we know what would happen if such a ceteris paribus clause was met? According to KW, this kind of speculation would be best left to “science fiction writers and futurologists”:

But how can we have any confidence of this? How in the world can I tell what would happen if my brain were stuffed with extra brain matter, or if my life were prolonged by some magic elixir? Surely such speculation should be left to science fiction writers and futurologists. We have no idea what the results of such experiments would be. They might lead me to go insane, even to behave according to a quus-like rule. The outcome really is obviously indeterminate, failing further specification of these magic mind-expanding processes; and even with such specifications, it is highly speculative. (1982, p.27)
KW concludes that a dispositionalist reply of this sort cannot meet the sceptical challenge, given that a speaker could not be reasonably viewed as being disposed *ceteris paribus* to apply an expression in a determinate pattern across a potentially infinite number of future cases.

§1.6.9 KW’s Circularity Objection

KW’s second attempt at spelling out a set of *ceteris paribus* conditions improves on the preceding construal by using a more realistic counterfactual:

If I somehow were to be given the means to carry out my intentions with respect to numbers that presently are too long for me to add (or to grasp), and if I were to carry out these intentions, then if queried about \(m + n\) for some big \(m\) and \(n\), I would respond with their sum (and not with their quum). (1982, p.28)

KW argues that though this counterfactual would be true, it would have no more traction against the sceptical challenge in that it effectively presupposes that the speaker intends to use ‘plus’ to refer to the addition function. In other words, it was already settled for the dispositionalist what function I meant by ‘plus’, in virtue of some prior fact. This is problematic, of course, because presupposing that there is some such fact is off limits to the reductive dispositionalist:

. . . if the dispositionalist attempts to define which function I meant as the function determined by the answer I am disposed to give for arbitrarily large arguments, he ignores the fact that my dispositions extend to only finitely many cases. If he tries to appeal to my responses under idealized conditions that overcome this finiteness, he will succeed only if the idealization includes a specification that I will still respond, under these idealized conditions, according to the infinite table of the function I actually meant. But then the circularity of the procedure is evident. The idealized dispositions are determinate only because it is already settled which function I meant. (1982, p.28)

The introduction of a *ceteris paribus* cannot help the dispositionalist, KW argues, even if that clause is designed to “exclude ‘noise’, or by [drawing] a distinction between ‘competence’ and ‘performance’” (1982, p.30):

No doubt a disposition to give the true sum in response to each addition problem is part of my ‘competence’, if by this we mean simply that such an answer accords with the rule I intended, or if we mean that, if all my dispositions to make mistakes were removed, I would give the correct answer. (Again I waive the finiteness of my capacity.) But a disposition to make a mistake is simply a disposition to *give an answer other than the one that accords with the function I meant*. To presuppose this concept in the present discussion is of course viciously circular. If I meant addition, my ‘erroneous’ actual disposition is to be ignored; if I meant skaddition, it should not be. Nothing in the notion of my ‘competence’ as thus defined can possibly tell me which alternative to adopt. (1982, p.30)

We could try to “specify the ‘noise’ to be ignored” without presupposing that my use of ‘plus’ refers to *addition*, KW says, though this effort will not be fruitful either:

Recall that the subject has a systematic disposition to forget to carry in certain circumstances: he tends to give a uniformly erroneous answer when well rested, in a pleasant environment free of clutter, etc. One cannot repair matters by urging that the subject would eventually respond with the right answer after correction by others. First, there are uneducable subjects who will persist in their error even after persistent correction. Second, what is meant by ‘correction by others’? If it means rejection by others of ‘wrong’ answers (answers that do not accord with the rule the speaker means) and
suggestion of the right answer (the answer that does accord), then again the account is
circular. If random intervention is allowed (that is, the 'corrections' may be arbitrary,
whether they are 'right' or 'wrong'), then, although educable subjects may be induced to
correct their wrong answers, suggestible subjects may also be induced to replace their
correct answers with erroneous ones. The amended dispositional statement will, then,
provide no criterion for the function that is really meant. (1982, p.31-32)

Consequently, KW concludes that sophisticated dispositionalism cannot succeed in addressing
the sceptical challenge on pain of circularity. In the next two sections I outline Boghossian’s claim
that KW’s revived finitude objection fails to convince and then explain Boghossian’s claim that
a circularity objection similar to KW’s can be based on a principle about belief holism.

§1.6.10 Boghossian’s Reply to KW’s Revived Finitude Objection

Boghossian (1989) argues that KW’s revived finitude objection against sophisticated
dispositionalism cannot succeed given that KW does not sufficiently argue for why we should
omit idealized generalizations about how a speaker would, for example, be disposed to use an
expression in the distant future. According to Boghossian, KW’s objection boils down to the claim
that “there will always be a serious indeterminacy in what my dispositions are, and thus . . .
dispositional properties [prove to be] an inappropriate reduction base for meaning properties”
(1989, p.528). Boghossian argues that KW’s demands for refuting this thesis are unwarranted.
KW effectively demands that:

. . . if it is indeed the property horse that I am disposed to apply the term ['horse’] to, then
I should be disposed to apply it to all horses, including horses so far away and so far in
the past that it would be nonsense to suppose that I could ever get into causal contact with
them. Otherwise, what is to say that my disposition is not a disposition to apply the term
to the property nearby horse, or some such? But no one can have a disposition to call all
horses ‘horses’, for no one can have a disposition with respect to inaccessible objects [e.g.
horses on Alpha Centauri]. (1989, p.528)

Boghossian claims that this argument fails to convince:

If I were now to go to Alpha Centauri, I probably would not be in any position to call
anything by any name, for I would probably die before I got there. But that does by itself
need not pose an insuperable obstacle to ascribing the disposition to me. All dispositional
properties are such that their exercise—the holding of the relevant counterfactual truth—
is contingent on the absence of interfering conditions or equivalently, on the presence of
ideal conditions. And it certainly seems conceivable that a suitable idealization of my
biological properties will render the counterfactual about my behaviour on Alpha
Centauri true. (1989, p.529)

KW does not provide a robust argument against these idealized considerations, but instead says
that “such speculation should be left to science fiction writers and futurologists” (Kripke 1982,
p.27). According to Boghossian, this is a serious problem for KW given that the burden of proof
is on him to show why idealized generalizations about semantic dispositions are inadmissible:

If the point is supposed to be . . . that one can have no reason for accepting a generalization
defined over ideal conditions unless one knows exactly which counterfactuals would be
true if the ideal conditions obtained, then, as Jerry Fodor has pointed out, it seems
completely unacceptable. For example, no one can claim to know all of what would be
true if molecules and containers actually satisfied the conditions over which the ideal gas
laws are defined; but that does not prevent us from claiming to know that, if there were ideal gases, their volume would vary inversely with the pressure on them. Similarly, no one can claim to know all of what would be true if I were so modified as to survive a trip to Alpha Centauri; but that need not prevent us from claiming to know that, if I were to survive such a trip, I would call the horses there ‘horse’. (1989, p.529)

According to Boghossian, nothing in KW’s argument demonstrates that relevant idealizations of this sort are inadmissible and so this objection of KW’s cannot convincingly block a dispositionalist response.\(^8\) However, Boghossian argues that there are other considerations that are potentially much more problematic for the sophisticated dispositionalist.

§1.6.11 Boghossian’s Belief-Holism Objection

Boghossian argues that the dispositionalist’s efforts to specify a set of ideal conditions that make no mention of any meaning or intentional facts must fail because the fixation of a speaker’s beliefs is “a holistic affair” (Miller 2007, p.209). According to Boghossian, the fixation of a speaker’s belief is “mediated” by the background assumptions that that speaker already has—as Boghossian puts it:

Belief fixation is typically mediated by background theory – the contents a thinker is prepared to judge will depend on what other contents he is prepared to judge . . . just about any stimulus can cause just about any belief, given a suitably mediating set of background assumptions. (1989, p.539)

For example, a speaker may token the belief there is a magpie just in case “there is some other kind of bird present, due to the presence of a belief to the effect that there are no birds apart from magpies in the relevant neighbourhood, and so on” (Miller 2007, p.209). The problem for dispositionalism is that the dispositionalist is committed to specifying in non-semantic and non-intentional terms the conditions under which a speaker is disposed to token the belief there is a magpie only in the presence of a magpie:

The Sophisticated Dispositionalist wishes to identify someone’s meaning such and such with facts of the form: S is disposed to token the belief B under conditions C, where the C are to be specified non-semantically and non-intentionally. Boghossian’s argument focuses on the fact that these conditions will have to include some proviso to the effect that certain other clusters of background beliefs B1, . . ., Bn . . . are absent. Thus, in specifying the optimal conditions for the meaning-constituting disposition concerning “magpie”, we will require non-intentional and non-semantic optimal conditions for the range of background beliefs which figure in the clusters B1, . . ., Bn . . .; otherwise we will not be able to stipulate, in non-intentional terms, the conditions under which each member of that range of beliefs is absent. (Miller 2007, p.210)

According to Boghossian, there exist a “potential infinity of such mediating background clusters of belief” and so what is required is “precisely what a dispositionalist theory was supposed to provide: namely, a set of naturalistic necessary and sufficient conditions for being a belief with a certain content” (1989, p.540). For this reason, Boghossian concludes that the dispositionalist’s

\(^8\) See Boghossian (1989), p.529-530 for a more detailed discussion of this objection.
project of specifying some such conditions is doomed and so dispositionalism\(^9\), as a reductive version of meaning factualism, must fail.\(^10\)

\section*{§1.6.12 Simplicity}

A further suggestion that KW rejects is the suggestion that my meaning addition by ‘plus’ “is to be preferred as the simplest hypothesis” (1982, p.38). According to KW, this suggestion cannot work because “[s]uch an appeal must be based either on a misunderstanding of the sceptical problem, or of the role of simplicity considerations, or both” (1982, p.38). On the first front, KW explains, the sceptical challenge is constitutive and not epistemological: the sceptic questions whether there is anything that would constitute my meaning one function or another, addition or quaddition, by my use of the ‘plus’ sign in the past. Considerations about simplicity might help us sort through the competing hypotheses, “but they obviously can never tell us what the competing hypothesis are” (1982, p.38). KW continues: “If we do not understand what two hypotheses state, what does it mean to say that one is ‘more probable’ because it is ‘simpler’? If the two hypotheses are not genuine hypotheses, not assertions of genuine matters of fact, no ‘simplicity’ considerations will make them so” (1982, p.38).

\section*{§1.6.13 Mental Images and other Mental States}

A further type of mental state reply that KW considers is the idea that meaning addition by ‘plus’ would be constituted by a kind of distinct and irreducible experience known to me by introspection:

Why not argue that “meaning addition by ‘plus’” denotes an irreducible experience, with its own special quale, known directly to each of us by introspection? (Headaches, tickles, nausea are examples of inner states with such qualia.) Perhaps the “decisive move in the conjuring trick” has been made when the sceptic notes that I have performed only finitely many additions and challenges me, in the light of this fact, to adduce some fact that ‘shows’ that I did not mean quus. Maybe I appear to be unable to reply just because the experience of meaning addition by ‘plus’ is as unique and irreducible as that of seeing yellow or feeling a headache, while the sceptic’s challenge invites me to look for another fact or experience to which this can be reduced. (1982, p.41)

KW rejects this response because, as Miller points out:

\begin{quote}
9 So far we have targeted individualist versions of dispositionalism and thus it may be objected that communitarian versions of dispositionalism may fare better, i.e. by appealing to the idea that an individual’s dispositions may fail to match those of his or her community. This move will not save dispositionalism because KW’s sceptical argument and the preceding objections turn on the attribution of a systematic error, something that can manifest at the community level as well. A community could be disposed to make systematic errors. For further discussion on the generality of the above objections against communitarian versions of dispositionalism, see Boghossian (1989), p.534-536 and Kusch (2006), p.123-125.
\end{quote}

\begin{quote}
10 This objection is discussed further by Miller (2007). According to Miller, the need to rule out background clusters of beliefs generates the circularity worry even if the set of background beliefs is finite (2007, p.211). Miller argues that a dispositionalist may be able to avoid this charge by adopting a kind of ‘ultra-sophisticated’ dispositionalist account of meaning, in line with a David Lewis style reductive analysis. This move gets this dispositionalist out of trouble from Boghossian’s circularity objection, Miller says, but also generates a new problem: there is no way for ultra-sophisticated dispositionalism to specify the optimality conditions in such a way that blocks out a potential infinitude of background clusters of beliefs (See Miller 2007, p.212-216).
\end{quote}
[I]t is not a necessary condition for understanding that some particular item come before one’s mind when one hears or uses a given expression . . . even in cases where there does seem to be an empirical regularity between a particular expression and a particular such item we can still perfectly well conceive of someone understanding the expression in the absence of that item (indeed, we can conceive of someone understanding the expression even when no such item is present at all (Miller 2002, p.6).

For example, it seems that I could understand the meaning of the expression ‘cube’ without having a cube before me or a mental image of a cube in my mind (2002, p.6). Moreover, it is not even a sufficient condition for meaning something by an expression or grasping the meaning of an expression that a certain mental image comes to mind, given that that mental image “does not by itself determine the correct use of the associated word, because the [image] thus associated is really just another sign whose meaning too requires to be fixed” (2002, p.6). KW’s sceptic can here challenge the relation between any mental image I cite and the expression associated with that image: “the relation of any picture or image to the associated word can be construed in such a way that any future pattern of use of the expression can count as correct” (2002, p.6). As such, The sceptic concludes: “Whatever comes before the mind can be made to accord with a deviant application of the expression [and thus] mental images set no standards for the correct use of an expression; one cannot “read off” from a mental image what counts as the correct use of an associated expression” (2002, p.6). And what goes for mental images also goes for any mental state conceived to have its own distinctive quale. The occurrence of such a mental state is neither necessary nor sufficient for understanding a word in a particular way.

§1.6.14 Primitive Mental State

One way that the above mental state suggestion could be potentially strengthened in order to deal with the objection above would be to suggest that a “non-quotidian” state of mind, a state that is not introspectable, constitutes my meaning addition by ‘plus’. In other words, my meaning addition by ‘plus’ is constituted by a primitive and sui generis state “not to be assimilated to sensations or headaches or any ‘qualitative’ states, nor to be assimilated to dispositions, but a state of a unique kind of its own” (1982, p.51). This might block the objection developed in the previous section, though at the cost of opening up two further lines of objection.

First, even if the primitive state reply is apparently “irrefutable”, it is irrefutable precisely because it “leaves the nature of this postulated primitive state - the primitive state of ‘meaning addition by ‘plus’’ - completely mysterious” (1982, p.51). In particular, it would be completely mysterious how a primitive state of this kind could inform my future use of an expression if indeed that state is inaccessible to me via introspection. How could I be aware of and yet guided by such a state? Second, this primitive mental state response faces a ‘considerable’ if not ‘impossible’ logical problem: If there is such a primitive state as my meaning addition by the ‘plus’ sign, then that state would have to be a finite object, contained in my finite mind.11 As such, KW continues,

11 See Kripke 1982, p.52.
such a state could “not consist in my explicitly thinking of each case of the addition table, nor even of my encoding each separate case in the brain: we lack the capacity for that” (1982, p.52). Nonetheless, whatever primitive state there is that would constitute my meaning addition by ‘plus’ would have to be present “in a queer way” (1982, p.52, KW cites Wittgenstein’s *Investigations*, §195). KW wonders how this could be the case: “Can we conceive of a finite state which could not be interpreted in a quus-like way? How could that be?” (1982, p.52). Even if we set aside this worry, KW says, and concede that meaning addition by ‘plus’ is a primitive state of a totally distinct sort, this does not get us out of trouble from the sceptic. Presumably, KW imagines, what the advocate of this primitive mental state response has in mind is a sort of entailment thesis like, “If I now mean addition by ‘plus’; then, if I remember this meaning in the future and wish to accord with what I meant, and do not miscalculate, then when asked for ‘68 + 57’, I will respond ‘125’” (1982, p.53). This thesis cannot succeed according to KW because “it remains mysterious exactly how the existence of any finite past state of my mind could entail that, if I wish to accord with it, and remember the state, and do not miscalculate, I must give a determinate answer to an arbitrarily large addition problem” (1982, p.53).

§1.6.15 Platonism

The final response KW considers is the suggestion that my meaning addition by ‘plus’ is constituted by my grasp of an abstract entity on par with the non-mental entities posited by mathematical realists and ‘Platonists’ (1982, p.53). On this account, KW explains, “[t]he addition function is not in any particular mind, nor is it the common property of all minds [instead] it has an independent, ‘objective’, existence” (1982, p.53). Therefore, there is no problem with respect to how the meaning of ‘plus’ could have consequences for a potential infinitude of addition problems, given that this abstract entity could be an infinite object. After all, KW says: “The proof that the addition function contains such a triple as (68, 57, 125) belongs to mathematics and has nothing to do with meaning and intention” (1982, p.54). Frege espoused a view that could account of my meaning addition by ‘plus’ in line with this suggestion:

Frege's analysis of the usage of the plus sign by an individual posits the following four elements: (a) the addition function, an ‘objective’ mathematical entity; (b) the addition sign ‘+’, a linguistic entity; (c) the ‘sense’ of this sign, an ‘objective’ abstract entity like the function; (d) an idea in the individual’s mind associated with the sign. The idea is a ‘subjective’ mental entity, private to each individual and different in different minds. The ‘sense’, in contrast, is the same for all individuals who use ‘+’ in the standard way. Each such individual grasps this sense by virtue of having an appropriate idea in his mind. The ‘sense’ in turn determines the addition function as the referent of the ‘+’ sign. (1982, p.54)

As KW points out, it seems that there is no special problem for this position because the sense of an expression determines its reference, or what that expression refers to (1982, p.54). However, KW argues that this suggestion cannot meet the sceptical challenge because what is in question is precisely how “the existence in my mind of any mental entity or idea can constitute ‘grasping’
any particular sense rather than another” (1982, p.54). Given that that mental entity or idea is a finite object in my mind, it can be interpreted in a quus-like way:

The idea in my mind is a finite object: can it not be interpreted as determining a quus function, rather than a plus function? Of course there may be another idea in my mind, which is supposed to constitute its act of assigning a particular interpretation to the first idea; but then the problem obviously arises again at this new level. (A rule for interpreting a rule again.) And so on. For Wittgenstein, Platonism is largely an unhelpful evasion of the problem of how our finite minds can give rules that are supposed to apply to an infinity of cases. Platonist objects may be self-interpreting, or rather, they may need no interpretation; but ultimately there must be some mental entity involved that raises the sceptical problem. (1982, p.54)

Consequently, KW claims that a Platonist reply of this kind cannot meet the sceptical challenge.

This concludes our brief discussion of candidate facts that might meet the sceptic’s demands. As we’ve seen, KW argues that there are no promising candidates for a meaning fact that would satisfy both the extensionality constraint and the normativity constraint. In the next two sections, I clarify and emphasize the importance of the latter constraint in KW’s argument.

§1.7 What is the Normativity of Meaning?

One crucial feature of KW’s argument that is necessary for its success against any factualist account of meaning is the thesis that meaning is in some sense *normative*. But what does it mean to say that meaning is normative? This point has not been thoroughly clear in reading KW because, as Anandi Hattiangadi (2007) points out, KW often conflates two broad claims with respect to the normativity of meaning: The first is the thesis that meaning is normative in the ‘norm-relative’ sense that the meaning of an expression sets a standard relative to which applications of that expression count as either *correct* or *incorrect*. According to Hattiangadi, this is an ‘anodyne’ or ‘platitudinous’ thesis for any theory of meaning. The second thesis KW employs is the ‘normativist’ or ‘prescriptivist’ thesis that meaning is essentially normative or *prescriptive* in the sense that the meaning of an expression normatively compels or prescribes how a speaker *ought* to apply that expression. Hattiangadi argues that it is this thesis that KW’s argument crucially turns on. More specifically, it is this thesis that enables KW’s argument to succeed against any form of meaning factualism. Hattiangadi addresses the unclarity between this thesis and the preceding one by differentiating the thesis that ‘meaning is normative’ from its cousin above, embodied in the slogan ‘meaning is norm-relative’. I follow Hattiangadi in teasing these two views apart and emphasize how KW requires that meaning is normative on the ‘normativist’ or ‘prescriptivist’ reading in order to maintain his argument with the kind of generality he aspires to.

§1.7.1 Hattiangadi: Normativity and Norm-Relativity

In order to adumbrate the thesis that meaning is normative, we need to tease apart two conceptions of that thesis. These two conceptions conveniently map onto the two constraints set by KW on any adequate account of my meaning *addition* by ‘plus’, the extensionality constraint
and the normativity constraint. The first thesis consists in the view that meaning is normative in the norm-relative sense that the meaning of an expression determines how that expression is correctly applied, i.e. in terms of the uses of that expression that count as correct and those that count as incorrect. Hattiangadi (2007) labels this thesis ‘Norm-Relativity’. This conception of the normativity of meaning maps onto KW’s extensionality constraint in the sense that if I do mean addition by ‘plus’, then that fact about my meaning rules out the deviant hypothesis that my use of that term denotes quaddition or anything else, i.e. given that addition is the function referred to by my term ‘plus’. Given that addition is the function that governs my use of the term ‘plus’, ‘125’ is the only answer that accords with my past meaning—it is the only answer that would count as correct insofar as ‘plus’ denotes addition. The intuitive idea behind the sceptic’s demand on this front is that the meaning of the term ‘plus’ must determine a standard of correctness for the application of that term, the grasp of which would determine whether I use that term correctly or not (in line with the principle embodied by Norm-Relativity). So in light of whatever fact there is about me that secures addition as the function that regulates my use of the term ‘plus’, that fact dictates what behaviour counts as correct and incorrect going forward (in terms of how that term would be correctly and incorrectly applied by me). Of course, this is not the only constraint on meaning for KW and likewise it is not the only construal of the claim that ‘meaning is normative’. KW’s second constraint involves a more robust conception of this thesis, as we saw earlier, in terms of how a speaker’s meaning intention with respect to a given expression dictates or prescribes how a speaker ought to go on in using that expression.

Recall that on KW’s second constraint for any candidate fact as to what I meant by ‘plus’, whatever fact secures addition as the unique function governing my use of that term must also prescribe how I ought to go on in applying that term. Whatever fact is cited with respect to the first constraint must also satisfy the second. An adequate account of my meaning addition by ‘plus’ in terms of some fact must also provide an account of how that fact prescribes how I ought to use that term at present and in the future. This constraint of KW’s is based on a stronger conception of the normativity of meaning, embodied in a thesis that Hattiangadi dubs ‘Normativity’. According to this thesis, the claim that meaning is normative consists in the intuitive idea that the meaning of an expression prescribes how that expression ought to be applied, e.g. that ‘green’ ought to be applied only to green things and not, say, blue things, and that ‘125’ and not ‘5’ is the answer that ought to be given in response to the query ‘68 + 57 = ?’. Normativity, as such, essentially involves the preceding thesis that meaning is norm-relative but expands on that thesis by taking the notion of correctness to immediately imply a prescription. According to Hattiangadi, KW runs this conception of the normativity of meaning together with the former:

On the one hand, Kripke clearly attributes to the semantic realist [i.e. factualist] the assumption that understanding the meaning of a word is analogous to following a rule—that when someone means something by a word, such as ‘plus’, she has a specific rule in mind (either the rule for addition or quaddition) which determines whether any answer she might give to an addition sum is correct. This assumption is evident, for instance,
when he says ‘[t]his is the whole point of the notion that in learning to add I grasp a rule: my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future.’ On the other hand, Kripke attributes to the semantic realist the view that meaning is normative, that is, that what I mean is prescriptive, or action-guiding. This is evident, for instance, when Kripke expresses the ‘eerie feeling’ someone under the sway of the intuitive, semantic realist view, might have. He says: ‘Even now as I write, I feel confident that there is something in my mind—the meaning I attach to the “plus” sign—that instructs me what I ought to do in all future cases. I do not predict what I will do . . . but instruct myself what I ought to do to conform to the meaning.’ That is, Kripke seems to maintain that, at least according to the semantic realist, meaning must be both norm-relative and normative. (2007, p.37-38)

KW requires the stronger thesis that meaning is normative in the prescriptive or action-guiding sense, Hattiangadi says, in order to maintain the a priori success of his sceptical argument against any account of a meaning fact (i.e. to draw the conclusion with the sort of generality that KW intends) as opposed to having to deal with particular versions of meaning factualism on a case-by-case basis.

In brief, Hattiangadi claims that by assuming that meaning is normative, KW is able to “remove the gaps in the sceptical argument” by making use of the arguments originally devised by Hume and Moore against the existence of moral facts, arguments that “purport to rule out all naturalistic reductions of normative moral concepts and properties” (2007, p.37-38). Hattiangadi points out that this argument is structurally similar to A.J. Ayer’s argument against the existence of moral facts (2007, p.39):

Ayer holds that moral realists, in committing themselves to the existence of objective, moral facts, face a dilemma. If the putative moral facts are said to be ‘natural’ facts, the realist commits the ‘naturalistic fallacy’. If the putative moral facts are said to be non-natural, moral facts turn out to be unlike ordinary empirical facts and therefore unknowable. (2007, p.39)

KW’s argument takes analogous route by attempting to show that meaning factualism faces a similar dilemma. As Hattiangadi explains:

. . . if we grant that meaning is normative, Kripke’s sceptic could be taken to show that the semantic realist faces the same dilemma as the moral realist: if she says that semantic facts are natural, as the dispositionalist does, she commits a fallacy, and if she says that semantic facts are non-natural, she makes them inherently mysterious and unknowable. (2007, p.39)

These arguments share two crucial assumptions, according to Hattiangadi. First, it is assumed that the judgments involved are action-guiding, and second, it is assumed that an internalist account of action-guidingness for these judgments is true, i.e. “they presuppose that if someone judges that she ought to do something, then she is ipso facto motivated to do it” (2007, p.39). So, in the meaning case, this amounts to assuming both that semantic judgments are essentially action-guiding and that semantic ‘oughts’ are categorically prescriptive. Most importantly, Hattiangadi explains, “semantic oughts cannot simply be means/end prescriptions . . . concerning what one ought to do in order to achieve the object of a wish or a desire or to achieve an end” because this would not engage the metaethical arguments that KW requires in order to maintain the a priori
success of his argument against any putative account of a meaning fact (2007, p.182). As Hattiangadi puts it, “[t]he crucial point is that semantic prescriptions must be categorical as opposed to merely instrumental if the sceptic is to use the normativity of meaning in his service” (2007, p.183). Hattiangadi argues that we could not reasonably view semantic obligations as categorically prescriptive. According to her, the ‘oughts’ that issue from semantic judgments could only be merely instrumental in the latter sense. So, given that KW’s argument against the existence of meaning facts hinges on the claim that meaning is normative in the sense that it involves categorical ‘oughts’, this argument can be blocked by showing that we do not have semantic obligations in the sense that Normativity implies. As Hattiangadi has mentioned, the assumption that meaning is norm-relative is not sufficient for grounding these arguments against the existence of any meaning facts. So, if meaning is only norm-relative and semantic prescriptions are only instrumental, as Hattiangadi claims, then KW’s argument cannot have the a priori and fully general success that we initially viewed that argument to have.12

§1.8 Conclusion

In this chapter I introduced KW’s sceptical argument according to which there are no meaning facts in virtue of which ascriptions of meaning such as “Gerard means addition by ‘plus’” are true or false. I explained how KW motivates this argument via a sceptical paradox that he imagines a bizarre sceptic poses. This sceptic hypothesizes that my past meaning intention with respect to the ‘plus’ sign was actually to refer not to addition, but to a non-standard function quaddition. KW set two constraints on any adequate account of a meaning fact that would rule out this hypothesis proposed by the sceptic: the extensionality constraint and the normativity constraint. Following KW, I enumerated several candidates for a meaning fact that would meet these two constraints, and explained why KW found all of them wanting—including even the most sophisticated version of a dispositionalist account of meaning. Thus it was concluded that there simply are no facts of the sort in virtue of which a speaker means one thing rather than another by an expression. One of the key premises in this argument was the thesis that meaning is essentially normative. As Hattiangadi suggested, this thesis could amount to two different claims: The weak conception of that thesis states that meaning is normative in the norm-relative sense that the meaning of an expression implies a standard according to which applications of that expression can count as correct or incorrect. The stronger conception of the thesis states that the meaning of an expression normatively compels or prescribes how that expression ought to be applied. KW requires the latter thesis, Hattiangadi says, in order to maintain his conclusion that there are no meaning facts.

12 Very roughly, the thought is that one can be released from semantic ‘obligations’ via a mere change in desires. If I have no desire to tell the truth, for example, in what sense am I obliged to apply a term to objects that fall within its extension? However, the details of Hattiangadi’s argument against Normativity are beyond the scope of this dissertation. For a complete exposition of the argument, see Hattiangadi (2006) or Hattiangadi (2007), ch.7.
Chapter Two continues the discussion of KW’s normativity objection against dispositionalist theories of meaning and links this discussion to ongoing debate in metaethics concerning ‘open question’ style arguments of a similar kind. We’ll see that KW’s normativity objection apparently fails due to the fact that the reduction of meaning facts to facts about speakers’ dispositions can be conceived to be \textit{a posteriori} in the manner of the reduction of water to H$_2$O. However, we’ll go on to argue that despite this a normativity challenge to dispositionalism can be given by constructing an analogue of the ‘Moral Twin Earth’ argument developed by Terry Horgan and Mark Timmons against \textit{a posteriori} versions of ethical naturalism. The eventual upshot will be that meaning is normative in a sense that poses a genuine—and as yet unanswered—challenge to reductive dispositionalist accounts of meaning.
Chapter Two: Twin Earth and the Open Question Argument

§2.1 Introduction

In this chapter I continue the discussion of KW’s ‘normativity’ objection against dispositionalism by viewing that objection as a kind of open question argument. I explain that KW’s version of this argument cannot succeed against reductive dispositionalist accounts of meaning, given that these accounts conceive meaning facts as reducible to facts about speakers’ dispositions by means of an \textit{a posteriori} reduction, i.e. in the manner of the reduction of water to $\text{H}_2\text{O}$. Nonetheless I argue that this objection can be revived by constructing an analogue of the ‘Moral Twin Earth’ argument proposed by Terry Horgan and Mark Timmons against \textit{a posteriori} versions of ethical naturalism. The upshot of this argument is that there is evidence that meaning is normative in a way that poses a genuine threat to reductive dispositionalist accounts of meaning.

In the following I first outline Moore’s open question argument and explain how, following Zalabardo (1997), KW’s analogue of this argument cannot succeed against the \textit{a posteriori} versions of dispositionalism. I then develop a revised open question argument that does challenge these accounts, modelled on the Moral Twin Earth argument developed by Horgan and Timmons (1992a).\textsuperscript{13}

§2.2 Moore: The Classical Open Question Argument

In his \textit{Principia Ethica} (1903), G.E. Moore develops an objection to \textit{definitional} naturalism, “the thesis that moral properties are identical or reducible to natural properties as a matter of definitional or conceptual fact[,\]” by arguing that this view commits what he calls the ‘naturalistic fallacy’ (Miller 2013, p.11). According to Moore, this fallacy is committed by any attempt to define ‘good’ in terms of a property or set of properties, natural or otherwise. On Moore’s view, ‘good’ is not definable at all and so even if good was a natural property, Moore says, “that would not alter the nature of the fallacy nor diminish its importance one whit. All that I have said about it would remain quite equally true: only the name which I have called it would not be so appropriate as I think it is” (Moore 1903, p.65). As such, Moore argues, the fallacy is “thus committed by anyone who tries to give any definition of ‘good’ or analysis of the concept which it expresses” (Miller 2013, p.12).

\textsuperscript{13} Horgan and Timmons develop this argument in a number of different formats. The ‘revised’ open question argument outlined here is only one version of the argument. Horgan and Timmons see themselves as developing a formula for constructing arguments as opposed to developing a single specific argument. See Horgan and Timmons 1991, 1992a, 1992b, 1996a, 1996b, 2000a, 2009, 2015. Horgan and Timmons discuss the broad strategy of their Moral Twin Earth argument more generally in a 2013 entry to the \textit{International Encyclopedia of Ethics}, “Moral Twin Earth”.
Moore motivates his case against any attempt to define ‘good’ by means of some conceptual or *a priori* reduction to a certain property or set of properties by constructing an argument that we will call the ‘classical open question argument’. Before we outline that argument, we’ll need to define what it is for a question to be ‘open’ or ‘closed’ on Moore’s technical use of those terms. First, a question is closed just in case sincerely asking that question implies that a speaker does not grasp the meanings of the terms or concepts involved in that question’s formulation (Miller 2007, p.12). If a speaker were to ask for example ‘Smith is an unmarried man, but is he a bachelor?’, this would suggest that that speaker is prey to some linguistic or conceptual misunderstanding of the expressions ‘unmarried man’ or ‘bachelor’. Anyone who truly grasps the meaning of these expressions (and thus the concepts they stand for) knows that they are equivalent. Any competent speaker would be able to answer the question ‘Smith is an unmarried man, but is he a bachelor?’ simply on the basis of their linguistic competence. By contrast, a question is open just in case it is not closed. With this crucial distinction in place, we can reconstruct Moore’s argument:

1. Suppose that the predicate ‘good’ is synonymous with, or analytically equivalent to, the naturalistic predicate ‘N’.

In that case, it would also be true that:

2. It is part of the meaning of the claim that ‘x is N’ that x is good.

And consequently:

3. Someone who seriously asked ‘Is an x which is N also good?’ would betray some conceptual confusion. The question would be closed.

But quite the contrary seems to be true, according to Moore:

4. For, given any natural property N, it is always an open question whether an x which is N is good. That is to say, asking the question ‘Is an x which is N also good?’ betrays no conceptual confusion. For example, it would be coherent to ask “Is a pleasurable action good” or “Is something which we desire to desire good?” (13).

Thus, Moore concludes:

5. It cannot be the case that ‘good’ is synonymous with, or analytically equivalent to, ‘N’.

And consequently:

6. The property of being good cannot as a matter of conceptual necessity be identical to the property of being N. (Miller 2013, p.12-13)

Though this argument faces some formidable objections in the moral case, we will bracket these and focus on whether a similar argument can be run in the meaning case. We will see that even

---

if the standard objections can be met, an open question argument against dispositionalism has its own troubles.

§2.3 Zalabardo: KW’s Normativity Argument

Jose Zalabardo argues in his 1997 that if KW’s normativity objection is construed as an open question argument in the style of Moore, this argument cannot succeed against modern reductive dispositionalist accounts of meaning. Zalabardo credits this insight to Fodor (1990), who interprets KW as claiming that “we cannot draw evaluative conclusions from descriptive premises concerning speakers’ dispositions because an open question argument would invalidate these inferences” (Zalabardo 1997, p.282). KW may have been drawn to this argument, Zalabardo explains, on account of the following line of reasoning:

How a speaker would apply a predicate under ideal conditions cannot be the fact that determines how he should apply it because we can meaningfully ask whether the applications that he would endorse under ideal conditions are the ones that he should endorse. (1997, p.282)

On this construal, KW would be arguing that for any speaker disposed to apply an expression in a certain pattern, one could sincerely and coherently ask whether that speaker ought to apply that expression in that way.

In Chapter One, we briefly discussed how an ‘open question’ interpretation of the normativity objection could be grounded in KW’s demand that meaning facts must essentially amount to normative facts. Boghossian (1989) gestures at this interpretation of the objection by emphasizing KW’s point that “[t]he relation of meaning and intention . . . is normative, not descriptive” (1982, p.37). Dispositionalism is unequipped to explain this relation, on KW’s view, because facts about speakers’ dispositions are essentially descriptive, whereas what is required is a normative fact. As Boghossian explains:

. . . even if there were a suitably selected disposition that captured the extension of an expression accurately, the disposition could still not be identified with that fact of meaning, because it still remains true that the concept of a disposition is descriptive whereas the concept of meaning is not. In other words, according to Kripke, even if there were a dispositional predicate that logically covaried with a meaning predicate, the one fact could still not be identified with the other, for they are facts of distinct sorts. (Boghossian 1989, p.532)

In light of this, KW rejects dispositionalism as an adequate response to the sceptical challenge. We can formalize this construal of the objection by tracking Moore’s argument as formulated above. It can be formulated thus:

(1) Suppose that ‘means addition by the ‘plus’ sign’ means the same as ‘disposed to answer with the sum’.

In that case, it would also be true that:

15 Zalabardo also considers what he calls the ‘justification argument’, but as the cogency of this argument is irrelevant to the main strategy of my thesis, I don't discuss it here.
(2) It is part of the meaning of the claim that ‘S is disposed to answer with the sum’ that S means addition by the ‘plus’ sign. And consequently:

(3) Someone who seriously asked ‘S is disposed to answer with the sum, but does S mean addition by the ‘plus’ sign?’ would betray some conceptual confusion. The question would be closed, just as ‘Smith is an unmarried male, but is he a bachelor?’ would be closed.

But quite the contrary seems to be true:

(4) Given any S that is disposed to answer with the sum, it is an open question as to whether S means addition by the ‘plus’ sign. The question ‘Smith is disposed to give the sum, but does he mean addition by the ‘plus’ sign?’ would be open in that it betrays no conceptual confusion in asking that question. (It is open because ‘Smith is disposed to give the sum, but ought he to give the sum?’ is open.)

Therefore:

(5) It cannot be the case that ‘means addition by the ‘plus’ sign’ means the same as ‘disposed to answer with the sum’.

And subsequently, KW concludes:

(6) The property meaning addition by the ‘plus’ sign cannot as a matter of conceptual necessity be identical with the property being disposed to answer with the sum.

This argument might in principle succeed against versions of dispositionalism grounded in an allegedly analytic connection between ‘means addition’ and ‘disposed to give the sum’, though it cannot succeed against the reductive versions of dispositionalism popular today. As Fodor puts it, if KW’s normativity objection amounts to this kind of open question argument, then it is not a new issue and moreover it is unclear how this objection could have any traction in the meaning case:

In short, I’m not clear how - or whether - “open question” arguments can get a grip in the present case. I am darkly suspicious that the Kripkensteinian worry about the normative force of meaning is either a non-issue or just the reduction issue over again; anyhow that it’s not a new issue. (Fodor 1990, p.136)

Fodor does not explain why this open question argument ends up being a “non-issue” in his brief comments on the topic. Zalabardo tries to flesh this out on Fodor’s behalf.

According to Zalabardo, the heart of Fodor’s objection is that dispositionalism can avoid an open question argument of the sort sketched above by rejecting the idea that meaning facts reduce to dispositional facts as a matter of conceptual necessity, i.e. by means of an a priori reduction grounded in a claimed analytic equivalence or synonymy relation. Instead, the dispositionalist can say that meaning facts reduce to dispositional facts as a matter of metaphysical necessity, i.e. by means of an a posteriori reduction, similar in kind to the reduction of water to H₂O. So the fact that we can coherently ask whether a speaker’s disposition to use an expression conforms to the way he ought to use that expression does not rule out the theoretic reduction of
meaning facts to facts about idealized dispositions—instead it rules out a particular *a priori* conception of that reduction. As Zalabardo explains, this objection cannot succeed against forms of dispositionalism that view meaning facts as *a posteriori* reducible to facts about dispositions:

> [P]roponents of naturalistic reductions of semantic notions see their task as on a par with other theoretic reductions, such as the identification of water with H₂O or of heat with kinetic energy. The proponents of [modern reductive dispositionalist accounts of meaning] aim at revealing the nature of [meaning], and of the attending normative facts, precisely in the sense in which the nature of water is revealed by its identification with a certain molecular structure . . . No doubt we can meaningfully ask whether water is H₂O, but the meaningfulness of this question does not undermine the identification of water with H₂O[.] (Zalabardo 1997, p.282-283)

Or as Miller (2006) puts it:

> Just as one cannot undermine the theoretical identification of water with H₂O by showing that judgments about water are not analytically equivalent to judgments about H₂O, the versions of reductive dispositionalism on offer cannot be undermined via an argument to the effect that semantic judgments are not analytically equivalent to judgments about optimal dispositions. (Miller 2006, p.110)

Consequently, it looks as though KW’s normativity objection, conceived as an open question argument, cannot be applied generally to rule out any dispositionalist account of meaning. Some versions of reductive dispositionalism can avoid this objection and thus block KW’s thesis that there are no meaning facts by suggesting that meaning facts are constituted by facts about dispositions, by means of an *a posteriori* reduction.¹⁶

In the next sections I consider whether the open question argument can be conceived in a way that does apply against reductive dispositionalist accounts of meaning. I argue that it can and take the Moral Twin Earth argument posed by Horgan and Timmons (1992a) as a model for devising this objection.

### §2.4 Horgan and Timmons: The Open Question Argument Revised

In the past Moore’s open question argument was thought to deliver a death blow to naturalistic moral realism. As Horgan and Timmons (hereafter ‘H&T’) explain in their 1992:

> . . . the argument’s persuasive appeal held a good many philosophers in its sway. Its appeal was felt by W.D. Ross (1930: 7-11, 92-3) and A.C. Ewing (1948: 41-2) who, like, Moore were led to espouse ethical nonnaturalism. Its appeal was also felt by A.J. Ayer (1952: 104-5) and R.M. Hare (1952: ch.5) who, because they found the metaphysical and epistemological commitments of nonnaturalism unpalatable, were led to noncognitivist accounts of moral discourse. (1992a, p.153)

Moore’s argument, as it turned out, was only persuasive insofar as a certain semantic view was popular among ethical naturalists. In time, this thesis was abandoned in favour of a more plausible view pioneered by Kripke (1972) and Putnam (1975). Consequently, “a novel strain of ethical naturalism - a kind allegedly immune to Moorean open question arguments - was bound to sprout

¹⁶ Nothing changes if these dispositions are idealized in the way we discussed in Chapter One. Boghossian (1989) claims that all dispositions are idealized, it’s just that some are *tacitly* so.
H&T argue that though recent forms of ethical naturalism side-step Moore’s argument, that argument can be reconfigured such that it *does apply* against these views: “the currently popular version of naturalism, despite its immunity to Moore’s version of the open question argument, succumbs to a newly fashioned open question argument” (1992a, p.153). According to them, a ‘Moral Twin Earth’ scenario supports this contention.

§2.4.1 New Wave Moral Semantics

H&T claim that contemporary ethical naturalism is developed around a general project that they call ‘new wave moral semantics’. The main aim of this project is to extend recent developments in the philosophy of language to moral semantics, where three developments are of primary importance: First, the *synonymy criterion of property identity*—the idea that two properties are identical only if the terms used to refer to those properties are synonymous—is widely rejected. This is so in light of numerous scientific counterexamples. For example:

The (sortal) property *being water* is identified with the property *being composed of H₂O molecules*; *being a cloud* is identified with *being a mass of water droplets*, temperature is identified with mean kinetic energy and so on. But no one supposes that ‘being water’ is synonymous with ‘being composed of H₂O molecules’, or that ‘temperature’ is synonymous with ‘mean kinetic energy’, and so forth for many other scientific identities. Quite simply it doesn’t seem that, in general, property identity requires synonymy. (1992a, p.156)

Second, the work of Kripke and Putnam has brought “widespread acceptance of the idea that names and natural kind terms are *rigid designators*—rigid in the sense that such expressions designate the same entity with respect to every possible world in which that entity exists” (1992a, p.157). Identity statements that involve rigid designators are necessarily true but not analytically true, and some such statements are ‘synthetic definitions’, e.g. ‘Water = H₂O’ (1992a, p.157). These statements “give the real nature or essence of the entity, property, or kind designated by a particular term” (1992a, p.157). So, for example, the identity statement ‘Water = H₂O’ is, if true, an expression of the “real, underlying essence of water and provides us with a (synthetically true) definition of ‘water’” (1992a, p.157). Third, there has been widespread acceptance of so-called ‘causal’ theories of reference. On these theories

. . . the semantical property of reference is to be understood as essentially involving appropriate causal connections between speakers’ use of a term and the thing to which the term refers. (1992a, p.157)

These theories are designed to explain two linguistic phenomena, according to H&T:

---

17 Horgan and Timmons specifically target Richard Boyd’s version of ethical naturalism, though the argument can be applied more generally. As they explain in a 2013 entry to the International Encyclopedia of Ethics: “the argument, understood generically, represents a recipe . . . that can be used as a basis for challenging any version of naturalist moral semantics” (2013).
Such theories propose to explain (i) how the reference of a term is originally determined (e.g. there being some sort of baptism or dubbing ceremony through which speakers in causal contact with an item acquire the ability to refer to that item through the use of some expression used in the ceremony), and (ii) how the capacity to refer is spread throughout a linguistic community (again, by speakers’ causally interacting with one another and with the item). (1992a, p.157-158)

The basic idea behind causal theories of reference is that “reference is ‘grounded’ by relevant causal hook-ups between speakers and the world” (1992a, p.158).

Richard Boyd (1988) defends a version of naturalistic moral realism characterized by these recent developments. First, Boyd claims that moral terms have synthetic definitions and as such, we can reject the synonymy criterion of property identity for moral terms: “the property goodness is identical with such and such natural property even though the term ‘good’ is not synonymous with any naturalistic term or phrase designating the relevant natural property” (1992a, p.158). Second, Boyd claims that moral terms rigidly designate the natural properties that they are used to refer to. Some such statements involving moral terms constitute definitions and these statements are necessarily true without being analytic. Third, Boyd extends a causal theory of reference to moral terms: “for moral terms, just as for names and natural kind terms, reference is a matter of there being certain causal connections between the use of moral terms and the relevant natural properties” (1992a, p.158). On Boyd’s version of a causal theory of reference, “reference is essentially an epistemic notion and so the relevant causal connections constituting reference are just those causal connections involved in the knowledge gathering activities” (1992a, p.158):

Roughly, and for nondegenerate cases, a term \( t \) refers to a kind (property, relation, etc.) \( k \) just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term \( t \) will be approximately true of \( k \) (excuse the blurring of the use-mention distinction). Such mechanisms will typically include the existence of procedures which are approximately accurate for recognizing members or instances of \( k \) (at least for easy cases) and which relevantly govern the use of \( t \), the social transmission of certain relevantly approximately true beliefs regarding \( k \), formulated as claims about \( t \) (again excuse the slight to the use-mention distinction), a pattern of deference to experts on \( k \) with respect to the use of \( t \), etc. . . . When relations of this sort obtain, we may think of the properties of \( k \) as regulating the use of \( t \) (via such causal relations) . . . (Boyd 1988, p.195)

Boyd extends this version of a causal theory of reference to moral terms and this commits him, according to Horgan and Timmons, to what they call the ‘causal regulation thesis’ (1992a, p.159):

CRT  Causal regulation thesis: For each moral term \( t \) (e.g., ‘good’), there is a natural property \( N \) such that \( N \) alone, and no other property, causally regulates the use of \( t \) by humans.

It is because moral terms regulate their use by humans in this way that moral terms “behave semantically like natural kind terms: they rigidly refer to certain natural properties and hence possess synthetic definitions” (1992a, p.159). Thereby, new wave moral semantics is expressed by this principle:

CSN  Causal semantic naturalism: Each moral term \( t \) rigidly designates the natural
property N that uniquely causally regulates the use of t by humans.

If we combine this thesis with metaphysical ethical naturalism:

**MN** *Metaphysical naturalism*: There are moral properties (and facts); and these are identical with natural properties (and facts).

Then the result is *causal ethical naturalism*. According to H&T, CSN does not fall to the classical Moorean argument, given that the standard Moorean argument targets ASN:

**ASN** *Analytic semantic naturalism*: Fundamental moral terms like ‘good’ have analytically true naturalistic definitions.\(^{18}\) (1992a, p.154)

A naturalist who accepts CSN can maintain MN without fear of Moore’s argument—that argument is only effective against *analytic ethical naturalism* (the combination of ASN and MN) and not *causal semantic naturalism*. The fact that moral terms and natural terms do not stand in a relation of synonymy is not a problem for MN when combined with CSN. For CSN, the link between a moral term t and a natural property N is, as we have seen, synthetic and not analytic. Consequently, it would be perfectly coherent to ask whether т is N, but the fact that this is an open question *does not* show that т is not N.\(^{19}\)

§2.4.2 General Strategy

H&T assume for the sake of argument that CRT is true and then go on to show, via a Moral Twin Earth scenario, that “nevertheless moral terms do *not* rigidly refer to the natural properties that causally regulate their use by humans” (1992a, p.160). According to them:

Although causal regulation may well coincide with - or even constitute - reference for certain terms (e.g. names and physical natural-kind terms) . . . for moral terms anyway, causal regulation does not coincide with reference. (1992a, p.160)

If H&T are right and CRT cannot accurately capture the meanings of moral terms, then CSN and thereby MN are in danger.

In order to approach H&T’s argument, we’ll first need to discuss how a causal theory of reference could capture the meanings of natural kind terms. This we will do via Putnam’s Twin Earth thought experiment. On Putnam’s view, our intuitions about this scenario suggest that we should adopt a causal theory of reference for natural kind terms such as ‘water’. CSN implies that ‘good’ is like ‘water’ and can therefore be explained via a causal theory of reference. If this were true, then a moral analogue of Putnam’s scenario should yield analogous results. According to H&T, it does not and they deploy the Moral Twin Earth scenario to show this.

---

\(^{18}\) Cf. the synonymy relation mentioned in step (1) of Moore’s open question outlined in §2.1 above.

\(^{19}\) Please excuse the blurring of the use-mention distinction here, this is carried over from Boyd, as quoted from his 1988 above.
§2.4.3 Putnam’s Twin Earth

In his 1975 essay “The meaning of ‘meaning’,” Putnam envisions a distant planet that he calls ‘Twin Earth’. This planet, Putnam explains is identical to Earth in every way save for the fact that the watery substance on this planet—the liquid that fills its oceans, rivers, and streams—is constituted not by H₂O, but by a different molecular compound, XYZ. So despite the fact that the superficial appearance of this watery substance is the same as that of the watery substance on Earth—they are both clear, odourless liquids, for example—on Earth this substance consists of H₂O molecules, whereas on Twin Earth this substance consists of XYZ molecules. So the substance that Earthers apply their ‘water’ term to consists of H₂O, while the substance that Twin Earthers apply their ‘water’ term to consists of XYZ. On reflection, Putnam says, Earthers and Twin Earthers do not mean the same thing by ‘water’.

Suppose that a Twin Earther visits Earth and after some time converses with an Earther about ‘water’. The Twin Earther, let’s suppose, is visiting Dunedin, New Zealand. The Earther and Twin Earther visit a local pub, the Albar. After some time of discussing the striking similarities between Earth and Twin Earth, the two come to a dispute about what one of the other patrons at the Albar is drinking. The Twin Earther says “That man is drinking water”, using his Twin English term ‘water’ to communicate this to the Earther. The Earther responds, “No, that man is not drinking water” and uses the English term ‘water’ to communicate this reply. According to Putnam, there is an important sense in which the Earther and Twin Earther are simply speaking at cross purposes: the two could not reasonably be viewed as disagreeing about what is in the glass held by the other pub patron. In other words, any ensuing dispute would be merely verbal. If the Earther and Twin Earther were to become cognizant of this difference, say by consulting some scientific texts from each planet, then whatever dispute they had would likely cease. Putnam takes these intuitions as evidence that the Earther and the Twin Earther refer to different substances by ‘water’, and that ‘water’ has a different meaning for each of them.

Putnam argues that our linguistic intuitions about Twin Earth can be explained by the view that natural kind terms such as ‘water’ possess a meaning that is crucially tied to the physical essence of the stuff that that term refers to, H₂O. So, on Twin Earth, one would say something that is incorrect if one were to apply the English term ‘water’ to refer to liquid filling the lakes and rivers of Twin Earth, the liquid composed of XYZ. Likewise, the Twin Earther would say something incorrect in applying the Twin English term ‘water’ to the outwardly similar liquid on Earth, composed of H₂O. This principle about the meaning of natural kind terms explains why we think that an Earther and Twin Earther do not genuinely disagree about whether the pub patron’s...

---

20 See also Putnam “Is Semantics Possible?” (1970) for a broad discussion of semantics and natural kind terms.

21 Granted, there is a sense in which they still might disagree, in that they could disagree about whether the patron is drinking ‘water’ in either the English or Twin English sense or gin (or indeed some other clear liquid). The example is not perfect, but the key point is that there is a way in which these two speakers fail to genuinely disagree about whether a liquid is ‘water’ or not (in their respective languages).
glass is filled with ‘water’ or not. H&T claim that we can view Putnam’s argument for this conclusion as a kind of semantic competence argument and close relative of Moore’s open question argument.

§2.4.4 Putnam and the Open Question Argument

According to H&T, Moore’s open question argument against definitional naturalism is based on a semantic competence test: If a competent speaker could readily answer a question solely in virtue of his or her grasp of the meanings of the terms involved in that question, then that question is closed. For example, ‘Sam is an unmarried man, but is he a bachelor?’ would be closed. A question is open, by contrast, just in case it is not closed. According to Moore, certain moral questions such as ‘Action x is pleasurable, but is x good?’ are open and this, he says, is evidence against definitional naturalism. Putnam’s argument can be viewed as using a similar strategy in support of treating natural kind terms as rigid designators by showing that the relevant analogous questions are closed.

Suppose that someone asks whether water is $H_2O$. This would amount to asking, ‘Liquid L is $H_2O$, but is it water?’ or ‘Liquid L is water, but is it $H_2O$?’ (1992a, p.161). The questions ‘Liquid L is $H_2O$, but is it water?’ and ‘Liquid L is water, but is it $H_2O$?’ are open, but this does not undermine CSN. Given that it does not hold that ‘water’ and ‘$H_2O$’ are synonymous, it is not committed to the idea that those questions are closed. However, the questions ‘Given that the use of ‘water’ by humans is causally regulated by the natural kind $H_2O$, is Liquid L, which is $H_2O$, water?’ and ‘Given that the use of ‘water’ by humans is causally regulated by the natural kind $H_2O$, is Liquid L, which is water, $H_2O$?’ would be closed given Putnam’s account of the semantics of ‘water’. So the fact—allegedly born out of the Twin Earth thought experiment—that they are closed is evidence for the truth of Putnam’s semantic view. They encapsulate the story about reference which holds that ‘water’ refers to $H_2O$, which is held by Putnam to explain our semantic intuitions concerning Earth (and Twin Earth). H&T aim to show, via their Moral Twin Earth argument, that the analogous questions in the case of the term ‘good’ would be open. If this is the case, then despite the fact that it survives Moore’s version of the open question argument, CSN would be challenged by a ‘revived’ version of the open question argument.

§2.5 Moral Twin Earth

In setting up their Moral Twin Earth argument, H&T first stipulate a few things about the ‘Moral Earth’ of their scenario. First they assume the truth of CRT and CSN, in line with Boyd’s thesis ‘that ‘good’ and ‘right’ are regulated by certain functional properties; and that, as a matter of empirical fact, these are properties whose functional essence is captured by some specific consequentialist normative theory’ (1992a, p.163). This theory we will refer to as ‘T’. Second, they assume that “there is some reliable method of moral inquiry which, if properly and thoroughly employed, would lead us to discover this fact about our uses of moral terms” (1992a, p.163). With these two assumptions in place, H&T construct an analogue of Putnam’s Twin Earth
designed to support their contention that the analogous questions from Putnam’s scenario sketched above, concerning ‘good’, would turn out to be open.

H&T envision a distant part of the actual world, a planet they call ‘Moral Twin Earth’. This Moral Twin Earth is identical to Earth in every way, including the fact that Twin Earthers have a certain terminology in Twin English that they use to evaluate the praiseworthiness of agents and their actions. They use the Twin English expressions ‘morally right’ and ‘wrong’ to express these sentiments and in general they are disposed to comply with their judgments about the actions that they call ‘morally right’ and ‘wrong’. They take seriously how these judgments bear on the well-being of Twin Earthers generally, and so on. All in all, this practice should be very familiar to the inhabitants of Earth as the kind of thing that we do when we speak about our moral practice, for which we reserve the terms ‘morally right’ and ‘wrong’ in English—as H&T put it, “the uses of [twin-moral] terms on Moral Twin Earth bear all of the ‘formal’ marks that we take to characterize moral vocabulary and moral practice” (1992a, p.164).

Despite these similarities, there is one important difference between Earth and Twin Earth: For Twin Earthers, the use of their term ‘morally right’ in Twin English is causally regulated by natural (functional) properties, whose essence is captured by a deontological normative theory, Td. H&T explain that this difference in the causal regulation of their moral terminology is “due at least in part to certain species-wide difference in psychological temperament that distinguish Twin Earthlings from Earthlings. (For instance, perhaps Twin Earthlings tend to experience the sentiment of guilt more readily and less intensively, than do Earthlings)” (1992a, p.165). Moreover, they explain that:

. . . if Twin Earthers were to employ in a proper and thorough manner the same reliable method of inquiry which (as we are already supposing) would lead Earthlings to discover that Earthling uses of moral terms are causally regulated by functional properties whose essence is captured by the consequentialist theory Tc, then this method would lead the Twin Earthers to discover that their own uses of moral terms are causally regulated by functional properties whose essence is captured by the deontological theory Td. (1992a, p.165)

H&T claim that there are two options for assessing the status of an exchange between the two invoking these terms:

On the one hand, we could say that the differences are analogous to those between Earth and Twin Earth in Putnam’s original example, to wit: the moral terms used by Earthlings rigidly designate the natural properties that causally regulate their use on Earth, whereas the twin-moral terms used by Twin Earthers rigidly designate the distinct natural properties that causally regulate their use on Twin Earth; hence, moral and twin-moral terms differ in meaning, and are not intertranslatable. On the other hand, we could say instead that moral and twin-moral terms do not differ in meaning or reference, and hence that any apparent moral disagreements that might arise between Earthlings and Twin Earthers would be genuine disagreements - i.e., disagreements in moral belief and in normative moral theory, rather than disagreements in meaning. (1992a, p.165)
H&T contend that the latter option is the most plausible assessment of the scenario, given that our intuition is that the Earther and the Twin Earther genuinely disagree. This causes trouble for CSN in that:

. . . if CSN were true, and the moral terms in question rigidly designated those natural properties that causally regulate their use, then reflection on this scenario ought to generate intuitions analogous to those generated in Putnam’s original Twin Earth scenario. I.e., it should seem intuitively natural to say that here we have a difference in meaning, and that [T]win English ‘moral’ terms are not translatable by English moral terms. But when it comes to characterizing the differences between Earthers and [T]win Earthlings on this matter, the natural-seeming thing to say is that the differences involve belief and theory, not meaning. (1992a, p.165-6)

H&T reinforce this assessment of their scenario by explaining how it would strike Earthers and Twin Earthers if they discovered this difference in the causal regulation of their moral terminology:

If CSN were true, then recognition of these differences ought to result in its seeming rather silly, to members of each group, to engage in inter-group debate about goodness - about whether it conforms to normative theory T^e or to T^d. (If, in Putnam’s original scenario, the two groups learn that their respective uses of ‘water’ are causally regulated by different physical kind-properties, it would be silly for them to think they have differing views about the real nature of water.) But such inter-group debate would surely strike both groups not as silly but as quite appropriate, because they would regard one another as differing in moral beliefs and moral theory, not in meaning. (1992a, p.166)

H&T argue that our intuitions about Moral Twin Earth suggest that CSN is probably false on account of the fact that the best explanation of Moral Twin Earth should respect the above intuitions rather than repudiate them. They use the phrase ‘probably false’ because “the inference to CSN’s falsity [is] inductive, an inference to the best explanation” and after all, our intuitions concerning Twin Earth scenarios “are empirical evidence about matters of semantics (just as their syntactic intuitions about grammaticality are empirical evidence about matters about syntax” (1992a, p.163). Nonetheless these intuitions “especially when robustly present among most all competent speakers, are quite powerful evidence; ceteris paribus, a semantic hypothesis that respects the intuitions is preferable to, and is more likely to be correct than, a semantic hypothesis that repudiates them” (1992a, p.163).

§2.5.1 The Open Question Argument Revised

H&T claim that a revised open question argument can be mounted from the Moral Twin Earth scenario:

. . . since it is a highly non-trivial issue whether the basic good-making natural property is the one (if there is just one) that causally regulates the use of ‘good’ by humans, or instead is the one that causally regulates this term’s use by twin-humans, or instead is some natural property distinct from either of these, the outcome of the Moral Twin Earth thought experiment also undergirds [a] ‘revised open question thesis’[.] (1992a, p.166)

According to them, our intuitions about Moral Twin Earth tell against CSN by showing that the analogs of the Putnam open questions outlined above are open when it comes to the term ‘good’. In particular, the question ‘Given that the use of ‘good’ by humans is causally regulated by natural
property N, is entity e, which has N, good?’ would be open, and likewise ‘Given that the use of ‘good’ by humans is causally regulated by natural property N, does entity e, which is good, have N?’ would also be open. CSN implies that these two questions would be closed, but as the Moral Twin Earth scenario attempts to demonstrate, it appears that they are open. Therefore, H&T argue that we should reject CSN.

§2.5.2 Generalizing the Revised Open Question Argument

H&T generalize this revised open question argument against any version of metaphysical naturalism (MN) of the same kind. As they note, CSN is only one version of a more general semantic thesis (1992a, p.167):

SSN Synthetic semantic naturalism: Fundamental moral terms like ‘good’ have synthetic naturalistic definitions. (1992, p.167)

H&T concede that there may yet be some form of SSN that would be able to meet the open question argument mounted from Moral Twin Earth, but according to them, Moral Twin Earth is:

. . . more than a specific thought experiment directed at the specific semantic thesis CSN. It is, in addition, a recipe for thought experiments. For any potential version of SSN that might be proposed, according to which (i) moral terms bear some relation R to certain natural properties that collectively satisfy some specific normative moral theory T, and (ii) moral terms supposedly refer to the natural properties to which they bear this relation R, it should be possible to construct a Moral Twin Earth scenario suitably analogous to the one constructed above - i.e., a scenario in which twin-moral terms bear the same relation R to certain natural properties that collectively satisfy some specific normative theory T’, incompatible with T. The above reasoning against CSN should apply, mutatis mutandis, against the envisioned alternative version of SSN. (1992a, p.167)

Therefore, H&T claim, Moral Twin Earth can be used to mount a generalized argument against SSN: “Questions analogous in form to [those sketched for CSN], for ‘good’ and for other moral terms, are open questions” (1992, p.167).

§2.5.3 How to Explain Moral Twin Earth

In closing, H&T briefly consider how we might explain our linguistic intuitions about Moral Twin Earth, once we have abandoned CSN or SSN more generally. What could be responsible for our intuition that an Earther and Twin Earther mean the same thing and genuinely disagree in spite of the stipulations about reference? H&T gesture at the kind of theory that might suffice to answer this question:

Although this is not the place to elaborate such a theory, the sort of rationale any such theory might be expected to provide for our meaning intuitions can be briefly outlined as follows. (1) One of the defining characteristics of a moral code is that it performs an action-guiding role for members of the community in which it is in force. (2) This normative aspect thus amounts to a semantic constraint for interpreting the practices of a community as moral practices, and so is plausibly taken to be built in to the meaning of moral terms like ‘good’ and ‘right’. (3) This action-guiding, normative feature of the meanings of moral terms helps explain why our intuitions engaged by the Moral Twin Earth thought experiment go the way they do: essential to the meaning of moral terms like ‘good’ and ‘right’ is their action-guiding function, which both Earthen usage of
moral terms and twin Earthian usage of moral terms share. Hence, despite the fact that the use of moral terms by the two groups is regulated by different natural properties, the (orthographically identical) moral terms of the two groups mean the same, contrary to Boydian semantics and consistent with the results of reflection on [the questions posed from Moral Twin Earth, assuming CSN’s truth above]. (1992a, p.170)

In the next sections and in the remainder of this dissertation, I extend this normativist explanation from H&T to a Meaning Twin Earth. I argue that a Meaning Twin Earth scenario generates intuitions at odds with dispositionalist accounts of meaning, including the recently popular a posteriori versions of dispositionalism. I claim that a revised open question argument, analogous to the one constructed by H&T, supports this challenge to dispositionalism, and that the best explanation of the relevant intuitions in the Meaning Twin Earth case proceeds via the claim that meaning is normative.22

§2.6 Meaning Twin Earth

Imagine that there exists a distant planet, which we will call ‘Meaning Twin Earth’. This Meaning Twin Earth is identical to Earth in most every way, including the fact that the inhabitants of this planet, Twin Earthers, engage in a linguistic practice of ascribing linguistic understanding to speakers and meaning to their utterances. They use expressions like ‘means bicycle’ in “Greg means bicycle by ‘bicycle’” to express judgments of this sort. This practice is very much like the one that we are familiar with on Earth with respect to the English language. Meaning Twin Earth is so similar in fact that if a group of Earthers were to visit Twin Earth, they would be disposed to translate the Twin English ‘means’ to be identical in meaning to the orthographically identical English term ‘means’ on account of these terms having the same role in the practices of each community.

In spite of the outward appearance that meaning ascription works exactly the same on Twin Earth as it does on Earth, there is one key difference that sets the two planets apart: The use of the English term ‘means’ as it pertains to Earther meaning ascriptions is causally regulated by natural (functional) properties that are best captured by a communitarian theory of meaning, $T^o$, whereas the use of the Twin English term ‘means’ as it pertains to Twin Earther meaning ascriptions is causally regulated by natural (functional) properties that are best captured by an individualist theory of meaning, $T^{in}$.23 This difference in the regulation of meaning-ascribing predicates between the two populations is generated by certain discrepancies in the psychological makeup of Earthers and Twin Earthers. As a result of these basic differences, the two populations each have robust justifications for why the relevant properties govern their meaning ascriptions,

---

22 For an attempt to use Twin Earth style considerations to provide evidence that the concept of belief is normative, see Miller and Kalantari (2016).
23 Roughly, the communitarian theory may be along the lines of those developed by Kripke (1982) or Wright (1980), while the individualist theory may be along the lines of that defended in Blackburn (1984) or Wee (2016).
i.e. their uses of the term ‘means’ in English and Twin English. Earthers claim that what a speaker means by a certain expression is primarily determined by the dictates of that speaker’s linguistic community, whereas Twin Earthers claim that what a speaker means by a certain expression is primarily determined by the dictates of that speaker’s idiolect.

The same degree of importance is placed in both theories on what counts as a speaker’s meaning something by his expressions, though the two theories depart from one another with respect to the features that are central and peripheral in determining meaning. The Earthers take it that conformity to the standards set by a community are more central, whereas the Twin Earthers take it that conformity to the standards set by a speaker him or herself are more central. These two theories come apart in some challenging examples. In some cases, Earthers may be disposed to judge that a lifelong solitary speaker cannot mean something by his or her words (e.g. on account of the absence of a linguistic community), whereas Twin Earthers may be disposed to judge that such a speaker can indeed mean something by his or her words (e.g. on account of their having some system for applying expressions devised by the speaker him or herself). They would, as such, use the English ‘means’ and Twin English ‘means’ in light of these considerations. For example, an Earther may express his view ‘Smith does not mean addition by the ‘plus’ sign’ and the Twin Earther might say ‘Smith does mean addition by the ‘plus’ sign’.

§2.6.1 Robinson Crusoe

Suppose that on Earth there is a man, Robinson Crusoe, who is the sole inhabitant of an isolated island. Crusoe has never had contact with any other human beings and moreover has inhabited the island since very soon after his birth. Against all odds, the young Crusoe survived a shipwreck and washed ashore unharmed. Suppose that Crusoe has a system for identifying blackberries, a staple of his diet, whereby he exclaims his Crusoean term ‘blackberry’ whenever he encounters such a berry. So whenever Crusoe encounters a berry with a certain outward appearance, he reports ‘Blackberry!’.

Suppose that a Twin Earther visits Earth and learns of Crusoe’s existence. This Twin Earther and one of the Earthers are able to actively observe Crusoe—perhaps by using a telescope and a directional microphone aboard a vessel near the island. The two observe him together and come to exchange their views on the puzzling behaviour of Crusoe. The Twin Earther engages the Earther on the topic of whether Crusoe means anything by his words. He says to the Earther, “Crusoe means blackberry by ‘blackberry’” and the Earther responds to him thus, “Crusoe does not mean blackberry by ‘blackberry’”. Naturally, the Earther uses his term for meaning ascription,

24 The precise source of these psychological differences is unimportant, but if necessary we could take them to be whatever differences incline actual communitarians and individualists to their respective views.
25 The literature on rule-following and private language is rife with Crusoe cases such as this. Wittgenstein is credited with coining the first case of this kind in the Investigations. Kripke briefly mentions such a case in his discussion of the sceptical solution and refers to Ayer and Rhees’ extended discussion of the topic. See Kripke (1982), p.110, Ayer and Rhees (1954). See also Baker and Hacker (1985), p.160. A more recent discussion can be found in Hacker (2010).
the English term ‘means’, as it relates to his communitarian theory of meaning to convey his sentiments about Crusoe, while the Twin Earther uses his term for meaning ascription, the Twin English ‘means’, as it relates to his individualist theory of meaning to convey his sentiments about Crusoe.

§2.6.2 How to Explain Meaning Twin Earth

Two options are available for explaining Meaning Twin Earth and the difference between the Earther and Twin Earthers’ use of their respective terms for meaning ascription: Either the two parties simply speak at cross purposes on account of the fact that their terms for meaning ascription, ‘means’ in English and ‘means’ in Twin English, refer to distinct functional properties and so mean different things; or on the other hand, the two parties genuinely disagree and as such mean the same thing by their terms for meaning ascription in English and Twin English. I argue that the latter assessment best captures our linguistic intuitions: our intuition in the above case is that the Earther and Twin Earther do indeed genuinely disagree.

The Earther and Twin Earther of the Meaning Twin Earth are most plausibly viewed as disagreeing on account of the incompatible contents they express in their judgments about Crusoe’s meaning. The Earther claims that Crusoe cannot mean anything by his words, whereas the Twin Earther claims that Crusoe can mean something by his words. The two parties disagree about whether Crusoe means something or nothing at all on account of the different theories that regulate ascriptions of meaning. So, despite the fact that Earthers and Twin Earthers use their terms for meaning ascription to refer to different theoretical properties, this fact does not apparently compromise the intuition that the two disagree.26 We can flesh the scenario a little more as follows.

The Earther and Twin Earther fundamentally disagree about whether there is any room for a distinction between what seems right and what is right vis-a-vis Crusoe’s use of the term ‘blackberry’. The Earther contends that Crusoe does not mean anything by ‘blackberry’ (or indeed any of his other expressions) on account of the fact that, according to the Earther, meaning is determined by the dictates of a speaking community—something that Crusoe does not have access to. Consequently, the Earther argues that there can be no distinction between what seems right and what is right for Crusoe in his use of the expression ‘blackberry’. Crusoe cannot mean anything by ‘blackberry’ because there is no room to draw this distinction and thus there is no sense in which we can reasonably speak of his uses of the expression ‘blackberry’ as being either

26 It may be objected that the Earther could concede in some peripheral cases of meaning ascription, like Crusoe’s, that an individual’s idiolect is sufficient for meaning (even if in general the more central cases are up to communal considerations). This is unlikely given that communitarian views of meaning are motivated by examples like Crusoe’s. The Earther may admit that speaker intentions are pertinent in some more ordinary examples. Nonetheless, according to him, communal checks are the crucial ones in determining meaning. In Crusoe’s case, where no communal check on meaning is present due to the absence of any speaking community whatsoever, the Earther is apt to judge that we cannot plausibly view Crusoe as saying anything with a determinate meaning.
correct or incorrect. The Twin Earther argues that Crusoe does mean something by his words and claims that there is room for the requisite distinction between what seems right and what is right for Crusoe. The Twin Earther perhaps emphasizes Crusoe’s system for identifying blackberries that have a particular colour and shape, something he has written down for himself and refers to every time he believes he has encountered one—and thus exclaims ‘Blackberry!’ . According to the Twin Earther, this system is sufficient for the distinction between what seems right and what is right for Crusoe in that Crusoe could, on occasion, make a mistake and correct himself by using his system for reporting berries he can eat. There is room, on the Twin Earthers view, to judge that Crusoe has applied the expression ‘blackberry’ correctly or incorrectly in light of this system.

§2.6.3 Meaning and Normativity

I suggest that the best characterization of Meaning Twin Earth centres on the difference in normative positions that the Earther and Twin Earther accept. As such, Meaning Twin Earth may provide an interesting new argument for the normativity of meaning if this thesis constitutes the best explanation, as in the case of Moral Twin Earth, of why our intuitions fail to match Putnam’s case. The normativist may have a ready explanation for this in that, according to him, ‘means’ plays the same evaluative role in English and Twin English. In this respect, meaning ascription plays an analogous role to the kind of moral evaluation at the centre of the Moral Twin Earth dispute. So the suggestion might be like its cousins in moral discourse, ‘means’ is essentially normative. Thus, it is reasonable to expect that if Earthers and Twin Earthers were to discover the difference in theory that underlies their use of meaning ascription, the two would persist in disagreeing.

§2.7 Conclusion

In this chapter I considered whether a classical or revised version of Moore’s open question argument could apply in a convincing way against modern a posteriori versions of semantic dispositionalism. I found that though the classical version of the argument fails, a revised version the argument mounted from an analogue of H&T’s Moral Twin Earth scenario does potentially pose a genuine challenge to these dispositionalist theories. Although this argument may not a priori rule out dispositionalism as a theory of meaning, it may yet provide a serious challenge to dispositionalism if a normativist explanation best captures our intuitions about a Meaning Twin Earth scenario. Fleshing out this line of thought will be the task of the remainder of this dissertation. In Chapters Three, Four, and Five I counter some prominent objections that might be raised against the use of Moral Twin Earth style arguments before returning to the question, in Chapter Six, of the shape of a normativist explanation of the relevant linguistic intuitions.
Chapter Three: Plunkett and Sundell on ‘Metalinguistic Negotiation’

§3.1 Introduction

This chapter evaluates a potential objection to the Meaning Twin Earth argument based on an argument devised by Plunkett and Sundell (2013). According to Plunkett and Sundell, some recent arguments from disagreement, including the Moral Twin Earth argument devised by Horgan and Timmons (H&T), rest on the allegedly dubious premise that all genuine disagreements are canonical disputes—disputes in which speakers literally express conflicting contents (beliefs, propositions, or attitudes). Plunkett and Sundell argue that some disagreements are expressed non-canonically; sometimes the immediate topic of a dispute is expressed non-literally, e.g. by implicature. In the most dramatic examples of non-canonical dispute, the topic of disagreement is what the terms used in communicating that dispute ought to mean, i.e. about whether they ought to denote one set of properties or another. Plunkett and Sundell call these cases ‘metalinguistic negotiations’ and count Moral Twin Earth as showcasing this kind of dispute. I argue that even if Moral Twin Earth can be plausibly viewed as a metalinguistic negotiation (over what properties ‘good’ ought to denote), Meaning Twin Earth cannot be viewed in this way.

In the next sections I outline Plunkett and Sundell’s argument against the shared meaning thesis cited above and detail how the Moral Twin Earth argument can be blocked by viewing the focal dispute of that scenario as a metalinguistic negotiation. I attempt to apply this same objection to Meaning Twin Earth but ultimately find that this scenario constitutes a limit case for Plunkett and Sundell’s method—a Plunkett and Sundell style objection cannot be convincingly applied against Meaning Twin Earth.

§3.2 Plunkett and Sundell: Shared Meaning and Disagreement

In their 2013, David Plunkett and Tim Sundell criticize a popular argumentative strategy in contemporary metaethics, which they call the ‘shared meaning strategy’ or ‘SMS’. This strategy is based on the principle that speakers who genuinely disagree must mean the same thing by key terms used in conveying that disagreement. According to Plunkett and Sundell, arguments by Hare (1952), Smith (1994), and Horgan and Timmons (1992a) follow this pattern:

In each of these cases, theorists take there to be a tight link between the fact that two speakers genuinely disagree with each other and the facts about what the respective speakers mean by their words. In cases where speakers are imagined to mean different things by their words (whether in virtue of ambiguity or difference in external environment or contextual variation), the worry is that the two speakers could not genuinely disagree with each other. To account for the purported datum that the two
speakers do genuinely disagree with each other, each theorist argues that we should take the speakers to mean the same things by the words they use in that exchange. (2013, p.6)

This link between disagreement and meaning is motivated by the collective intuition that certain verbal disputes appear to evince genuine disagreement.

According to Plunkett and Sundell, the standard way to understand two speakers as genuinely disagreeing begins by supposing that those “speakers disagree over the literally expressed content of what they are saying. Insofar as one views content in terms of propositions, this will amount to thinking that the speakers disagree about the truth of the propositions that they literally express” (2013, p.6). Such a dispute, in which speakers disagree about the truth or correctness of the content they literally express, can be called a canonical dispute. If a dispute is canonical, one can reasonably conclude that speakers engaged in that dispute mean the same thing by their words:

If it is true that a dispute is canonical, this lends powerful support to the conclusion that the speakers mean the same things by the words they use in that exchange. Why? Because if speakers meant different things by their words, then they would be very unlikely to express inconsistent contents. Hence, from the assumption that a dispute is canonical, one is in a reasonably good position to conclude that, insofar as speakers in that exchange use the same words, those speakers mean the same things by those words. (2013, p.7)

The general pattern of this argumentative strategy, the shared meaning strategy (SMS), takes at least four steps to show that a target semantic theory is implausible:

1. Introduce an actual or imagined dispute between two speakers, A and B, that yields the general intuition that A and B genuinely disagree. In the course of that dispute, A and B use some key term ‘t’ to convey their dispute (where ‘t’ is used, not mentioned).
2. Suppose that a target semantic theory is true.
3. If the semantic theory from (2) is true, then A and B attach different meaning to ‘t’.
4. If A and B attach different meaning to ‘t’, then (contrary to the intuition in (1)) A and B do not genuinely disagree.
5. Therefore, the target semantic theory from (2) must be false.27

The purpose of this strategy is to reject a metaethical theory by showing its underlying semantic theory to be false. The strategy does this by showing how a target semantic theory cannot vindicate the intuitive data in (1) for certain paradigm cases of apparently genuine moral disagreement. If the general aim of a metaethical theory is to explain the basic features of moral discourse and if moral disagreement is one such feature, then we can reasonably reject theories

27 Plunkett and Sundell sum up this shared meaning strategy in the following way: “Recall that the argument type we critique begins with the premise that a given dispute expresses genuine disagreement. From there, one argues (perhaps on inference to best explanation, perhaps on other grounds) that parties to the dispute literally express incompatible contents. In our terminology, one goes on to infer that the dispute is canonical. Finally, from the intermediate premise that the dispute is canonical, one can infer that parties to the dispute mean the same things by the relevant terms. (And from that conclusion, it follows that theories positing the relevant sort of variation in meaning—contextualism (variation in content only), ambiguity (variation in content because of variation in character), etc.—are false” (2013, p.12).
which cannot explain basic cases of apparent moral disagreement. This is how things generally go in the arguments outlined in the next three sections.

§3.3 Hare: The Missionary and the Cannibals

In his 1952, R.M. Hare argues that analytic ethical naturalism, which depends on a descriptivist moral semantics, is untenable. To show this, Hare introduces a now seminal thought experiment that, according to him, lends support to his so-called prescriptivist metaethical theory.28

Hare envisions a missionary who visits an island of cannibals. The missionary, in attempting to give an enlightened moral education to the island’s inhabitants, quickly discovers that the cannibals use the word ‘good’ in ways that differ from the applications made by the missionaries. In particular, the cannibals apply ‘good’ to people who collect the most human scalps, where the missionary applies ‘good’ to people who are meek and mild. As such, the two use the word ‘good’ to express apparently contradictory moral judgments. The cannibals are in fact equally surprised by the missionary’s deviant use of the word ‘good’:

[They] know that when he uses the word he is commending the person or object he applies it to. The only thing they find odd is that he applies it to such unexpected people, people who are meek and gentle and do not collect large quantities of scalps; whereas they themselves are accustomed to commend people who are bold and burly and collect more scalps than average. (1952, p.148)

According to Hare, descriptivism is forced to draw the absurd conclusion that the missionary and cannibals do not mean the same by ‘good’. As such, we cannot understand the missionary and cannibals as having a genuine disagreement:

If this were so, then when the missionary said that people who collected no scalps were good (English), and the cannibals said that people who collected a lot of scalps were good (cannibal), they would not be disagreeing, because in English (at any rate missionary English), “good” would mean among other things “doing no murder”, whereas in the cannibals’ language “good” would mean something quite different, among other things “productive of maximum scalps”. (1952, p.148-9)

This theory must be false according to Hare, given that we believe two parties genuinely disagree when they apply the word ‘good’ in such radically different ways. Hare concludes that the missionary and the cannibals must mean the same thing by ‘good’ and moreover, that his prescriptivist theory can explain this result. If the meaning of ‘good’ is understood to express an imperative that can be universalized, then we can vindicate our intuitions about disagreement in the case Hare describes by viewing the missionary and cannibal as expressing incompatible universal prescriptions or imperatives (this involves treating ‘good’ as a normative term in the sense that judgments involving ‘good’ express prescriptions).

28 Hare introduces this theory in The Language of Morals (1952) and subsequently develops it in Freedom and Reason (1965), and Moral thinking: Its levels, method, and point (1981).
Suppose for example that the missionary and the cannibal come to speak about the late Mother Teresa, the 1979 Nobel Peace Prize winning Roman-Catholic missionary whose hospices sheltered and cared for victims of HIV, AIDS, tuberculosis, etc. The missionary informs the cannibals about Mother Teresa and her deeds. In due course, the missionary and the cannibals come to exchange their views about Mother Teresa:

(1) (a) Mother Teresa is good.
(b) Mother Teresa is not good.

On a descriptivist view, the missionary and the cannibals simply speak past one another: The missionary expresses the judgment that Mother Teresa is meek and mild and the cannibals express the judgment that Mother Teresa does not collect as many scalps as possible. In this way, they express different beliefs but there is nothing incompatible about those beliefs on the descriptivist theory of meaning. Consequently, there is no sense in which the two genuinely disagree.

On Hare’s view and for most people, the missionary and the cannibals genuinely disagree when they exchange A and B. Given their respective schemas for applying ‘good’ to people, the missionary and cannibals express incompatible judgments about Mother Teresa. The missionary, in making the judgment that Mother Teresa is good, commands us to be meek and mild, i.e. he conveys the universal prescription that being meek and mild (like Mother Teresa) is what everyone ought to do. The cannibals, in making the judgment that Mother Teresa is not good, commands us to collect as many scalps as possible, rather than be meek and mild, i.e. he conveys the universal prescription that collecting as many scalps, and not being meek and mild, is what everyone ought to do.\(^{29}\) Those universal prescriptions are fundamentally incompatible and cannot be co-executed. This explanation vindicates our intuition about disagreement between A and B, something that the descriptivist theory of meaning cannot capture.

§3.4 Smith: Objectivity and the Pitfall of Moral Relativism

The general pattern of argument targeted by Plunkett and Sundell (2013) also appears in Michael Smith’s argument for his own brand of moral realism in *The Moral Problem* (1994). Smith’s target in this argument is an *a posteriori* version of ethical naturalism which he calls ‘non-definitional naturalism’.\(^{30}\) According to Smith, this naturalist view cannot explain moral disagreement as a feature of our moral phenomenology. Smith follows Hare’s argument in motivating this argument against non-definitional naturalism.

Hare’s prescriptivism is designed to avoid the same fate of descriptivism, which as the missionary scenario purports to reveal, collapses into *moral relativism*—a thesis that is deeply at

\(^{29}\) The secondary purpose of the missionary’s words is to express his attitude of approbation toward Mother Teresa, who is doing what she should, being meek and mild (and not, among other things, collecting as many scalps as possible). Likewise, for the cannibals, the secondary purpose of their words is to express their attitude of disapprobation toward Mother Teresa, given that she is not doing what she should, i.e. collecting as many scalps as possible as she should be.

\(^{30}\) Smith targets a moral theory similar to what we called ‘causal semantic naturalism’ in Chapter Two.
odds with moral phenomenology. On Smith’s view, the virtue of Hare’s theory is that it captures the objectivity of moral discourse.\textsuperscript{31} Hare is able to explain how our moral judgments are not merely relativistic and as such, he can explain our intuition that people genuinely disagree about moral matters. Descriptivism cannot capture these same features:

If descriptive accounts of the meaning of ‘good’ and ‘right’ make our moral judgments inescapably relativistic, and non-descriptive accounts don’t, then we surely have good reason to reject the idea that our moral judgments are descriptive. For such a relativism flouts core platiitudes about moral disagreement: platiitudes about the objectivity of morality (see also Brink, 1989, p.24, 29-35; though contrast Harman, 1985).\textsuperscript{32} (1994, p.35)

Any adequate moral semantics cannot allow disagreement to collapse into relativism, i.e. such that speakers in moral dispute merely speak past one another. Metaethical theorists…

\ldots must make sure that moral claims do not turn out to have different contents in different contexts. And yet this seems inevitable if they simply say that, for example, the word ‘right’ is used to refer to the feature of acts that is causally responsible for our uses of the term ‘right’. For if the cause of A’s and B’s uses of the word ‘right’ are not the same, then contrary to the platitude that if A says ‘x is right’ and B says ‘x is not right’ then A and B disagree, A and B are not disagreeing. A’s judgment that x is right has a different content from B’s judgment that x is right. (1994, p.35)

Here Smith mounts the argument from disagreement against metaphysical but non-definitional naturalism. According to Smith, this version of ethical naturalism fares no better than its analytic descriptivist ancestor given that its underlying semantic theory forces the absurd conclusion that A and B do not genuinely disagree.

§3.5 Horgan and Timmons: Moral Twin Earth

As we saw in Chapter Two, Horgan and Timmons (1992a) claim that a revised open question argument is effective against the modern \textit{a posteriori} versions of synthetic ethical naturalism popular today. According to them, an analogue of Putnam’s Twin Earth scenario, Moral Twin Earth, supports their contention: our intuitions about this scenario are incompatible with those about Putnam’s original scenario, contrary to the tenets of a modern naturalist semantic theory for moral terms—as what we called ‘causal semantic naturalism’ or ‘CSN’ in Chapter Two. On account of this mismatch, H&T conclude that moral terms like ‘good’ cannot be plausibly

\textsuperscript{31} Smith outlines a complete list of constraints for any adequate metaethical theory. See Smith 1994, p.127. The objectivity of moral judgment is the first of five constraints he lists here.

\textsuperscript{32} Smith’s reasoning here matches Horgan and Timmons’ comments on moral phenomenology and relativism in “Morality without Moral Facts” (2006). According to Horgan and Timmons (2006), any tenable metaethical theory must account for the basic features of moral phenomenology. They embody this in the \textit{Moral Phenomenology Criterion} (MPC), which states “A metaethical theory ought, if possible, to account for and vindicate as many of the deeply embedded aspects of moral thought and moral discourse—the phenomena of morality—as possible” (2006, p.223). One of the key features of moral thought and discourse is moral disagreement, which Horgan and Timmons capture in a platitude they call MP3: “There can be genuine, deep moral disagreements—including disagreements between people over fundamental moral assumptions and principles” (2006, p.223). According to Horgan and Timmons, moral relativism is untenable because it cannot accommodate the disagreement as a feature of (broadly-speaking) moral phenomenology, i.e. MP3: “The trouble with relativism, we maintain, is that it cannot make good sense of moral disagreements—disagreements that cannot be explained as disagreements about the non-moral facts of the case” (2006, p.228).
viewed as possessing a meaning in line with a Putnam-style naturalist analysis. Moral terms do not behave like their natural kind counterparts. This argument follows the format embodied in SMS above. This is revealed by the fact that they think two explanations of Moral Twin Earth are possible:

Given all these assumptions and stipulations about Earth and Moral Twin Earth, what is the appropriate way to describe the differences between moral and twin-moral uses of ‘good’ and ‘right’? Two hermeneutic options are available. On the one hand, we could say that the differences are analogous to those between Earth and Twin Earth in Putnam’s original example, to wit: the moral terms used by Earthlings rigidly designate the natural properties that causally regulate their use on Earth, whereas the twin-moral terms used by Twin Earthlings rigidly designate the distinct natural properties that causally regulate their use on Twin Earth; hence, moral and twin-moral terms differ in meaning, and are not intertranslatable. On the other hand, we could say instead that moral and twin-moral terms do not differ in meaning or reference, and hence that any apparent moral disagreements that might arise between Earthlings and Twin Earthlings would be genuine disagreements—i.e., disagreements in moral belief and in normative moral theory, rather than disagreements in meaning. (1992a, p.165)

The second option captures our linguistic intuitions about the scenario—Earthers and Twin Earthers are best viewed as genuinely disagreeing and as such, the two must mean the same thing by ‘good’:

We submit that by far the more natural mode of description, when one considers the Moral Twin Earth scenario, is the second. Reflection on the scenario just does not generate hermeneutical pressure to construe Moral Twin Earthling uses of ‘good’ and ‘right’ as not translatable by our orthographically identical terms. (1992a, p.165)

Our intuitions are simply different in the two scenarios.

H&T go on to explain why ‘good’ and ‘right’ must be translatable between English and Twin English33, but the key assumption about the link between disagreement and shared meaning, SMS, is in place. If the Earther and Twin Earther mean different things by ‘good’ then they could not genuinely disagree. We do think that they genuinely disagree, so they must mean the same thing by ‘good’. 34

33 Chapter Two outlines how Horgan and Timmons take the best explanation of Moral Twin Earth to involve the normativity of moral terms. See §2.5.3.
34 Like Hare, Horgan and Timmons ultimately advocate a version of ethical expressivism, which they call ‘cognitivist expressivism’. They rely on the notion of incompatible ought-commitments in order to vindicate the intuition that Earthers and Twin Earthers genuinely disagree. In their case, the literally expressed content of a moral judgment is an ought-commitment, something than is incompatible with the implementation of some other ought-commitments. So in a move that is very similar to Hare, they say that some ought-commitments cannot be co-executed. For instance, if being ‘good’ amounts to an ought-commitment to “be meek, mild, and not murder people,” then one could not also “collect as many scalps as possible.” If ‘good’ amounts to an ought-commitment to “collect as many scalps as possible,” then one could not also “be meek, mild, and not murder people.” For Hare, one could not universally prescribe both “collect as many scalps as possible” and ‘be meek, mild, and not murder people,” so ‘good’ (on the corresponding moral semantic theory) cannot be consistently applied to both someone who collects as many scalps as possible and to someone else who is meek, mild, and does not murder people.
§3.6 Plunkett and Sundell: Disagreement and Literally Expressed Content

Plunkett and Sundell claim that the shared meaning strategy (SMS) depends on the false assumption that all genuine disagreements are canonical disputes—disputes in which speakers conflict over the literally expressed content of what they are saying. Some genuine disagreements are in fact non-canonical disputes—disputes in which speakers do not literally express conflicting contents:

We argue that speakers can, and often do, genuinely disagree with each other even while in the disputes reflecting those disagreements, those speakers do not mean the same things by their words. (2013, p.3)

According to Plunkett and Sundell, an alternative strategy can better explain disagreement in some cases of this sort:

Taking a dispute to be canonical is not always the best—let alone only—non-debunking explanation of the intuition of genuine disagreement. There are many instances of non-canonical disputes—disputes that do not center on literally expressed content—that nevertheless express genuine disagreement. Non-canonical disputes that express genuine disagreements are, in fact, pervasive. And one type of non-canonical dispute in particular—what we call a metalinguistic negotiation—is not only capable of expressing genuine disagreement, but is, we argue, particularly plausible as an analysis of many normative and evaluative disputes. (2013, p.7)

This strategy does not make the false assumption that SMS stands on and, as such, can explain the non-canonical examples of genuine disagreement. This strategy focuses on the functions that speakers take their words to denote. According to Plunkett and Sundell, some genuine disagreements are non-canonical precisely because speakers disagree at the metalinguistic level, about the function that a word ought to denote. These disputes are best explained when viewed as metalinguistic negotiations—disputes in which speakers use words the way they do in part to communicate how they think those words ought to be used.

§3.6.1 Some Disagreements are Non-Canonical Disputes

Plunkett and Sundell claim that some disputes are cases in which speakers genuinely disagree even though they do not literally express incompatible contents. Indeed, according to them, some rather ordinary examples of non-canonical dispute are cases in which speakers literally express compatible contents. Take for example the following dispute:

(2) (a) There is one proton in the nucleus of a helium atom.

(b) No, there are two protons in the nucleus of a helium atom.

Taken literally, (2a) and (2b) do not express incompatible contents. There are exactly two protons in the nucleus of a helium atom—so strictly-speaking (2a) and (2b) are both true.

35 Plunkett and Sundell use the word ‘concept’, but this term is imprecise: ‘concept’ could refer to either items at the level of sense or items at the level of reference. I choose to use ‘function’, given that Plunkett and Sundell mean to use this term to refer to items at the level of reference.
Intuitively though, the two disagree. So how could these two speakers be viewed as disagreeing if they literally express compatible contents? According to Plunkett and Sundell, the two disagree on account of the incompatible contents (beliefs) they express *non-literally*, i.e. by implicature. The two genuinely disagree—and presumably they understand one another perfectly well to be disagreeing—but the root of their disagreement is implied, not literally expressed, in the exchange. They disagree in virtue of the *non-literally expressed content* expressed by (2a) and (2b):

The parties to the dispute in (2) disagree in virtue of the fact that the speaker of (2a) believes that there is *exactly* one proton in the nucleus of a helium atom, while the speaker of (2b) believes that there are *exactly* two protons in the nucleus of a helium atom. They not only believe those propositions, but they also communicate them, and therefore correctly perceive themselves to disagree with each other. (2013, p.12)

So, despite the fact that (2a) and (2b) literally express compatible contents, the speakers nonetheless communicate incompatible contents, albeit *non-literally*. In this respect, (2) constitutes a genuine disagreement.

According to Plunkett and Sundell, we can best capture what it is for two speakers to genuinely disagree by the lights of a general principle, *Disagreement Requires Conflict in Content* (DRCC), which states:

> If two subjects A and B disagree with each other, then there are some objects p and q (propositions, plans, etc.) such that A accepts p and B accepts q, and p is such that the demands placed on a subject in virtue of accepting it are rationally incompatible with the demands placed on a subject in virtue of accepting q. (Perhaps, though not necessarily in virtue of q entailing not-p.) (2013, p.11)

The primary virtue of this principle is that it is agnostic about the content of moral judgment. So though Hare, Smith, and H&T each have their own views about the content of moral judgment, they can all accept a principle like DRCC.

The special interest of the dispute in (2) above is that the two speakers engaged in that dispute do not *literally express* the incompatible contents that legitimate their dispute as a genuine disagreement. The two speakers express this disagreement non-literally, via implicature: The non-literally expressed content of (2a) is the belief there is *exactly* one proton in the nucleus of a helium atom; the non-literally expressed content of (2b) is the belief there are *exactly* two protons in the nucleus of a helium atom. Hence the two really do genuinely disagree, but not at the level of literally expressed content (as SMS theorists suppose).

The dispute in (2) delivers two important lessons about disagreement and meaning, according to Plunkett and Sundell. The first lesson is that we should be wary of the quick connection we commonly draw between disagreement and the canonical (or non-canonical) nature of a dispute. This connection is a dangerous one to make in that some non-canonical disputes might, as (2) purports to show, evince genuine disagreement:

> ... [2] demonstrates the danger of assuming, based on the existence of disagreement, that the relevant dispute involves the *literal* (semantic) expression of incompatible contents. In other words, it demonstrates that non-canonical disputes can reflect genuine disagreements. And the types of non-canonical disputes that can do so—disputes centered
on information conveyed via implicature (or presupposition, or connotation, or a host of other communicative mechanisms)—are hardly exotic to philosophers of language. This by itself is enough to block especially quick or simplistic instances of inferences of shared meaning from facts of disagreement. (2013, p.13)

So one take-away message from (2) is that our intuitions about genuine disagreement do not necessarily track cases of canonical dispute. Indeed, many cases of non-canonical dispute similar to (2) are cases in which speakers disagree even though they literally express compatible contents.

The second lesson to learn from (2) is that two questions are often conflated by theorists:

1. Is dispute x a genuine disagreement?
2. How is the content of that disagreement expressed in dispute x, i.e. are the contents literally expressed in dispute x incompatible?

SMS proponents are committed to running these two together. First, the intuition that dispute x is genuine is vindicated by the fact that the speakers engaged in x—say A and B—express different contents. Second, the content of that disagreement is literally expressed. If it was not literally expressed, we could hardly understand A and B as genuinely disagreeing, according to SMS theorists. According to Plunkett and Sundell, this is a serious mistake on the part of these theorists. Any disagreement principle based around incompatible content... will entail that for a dispute to express a genuine disagreement, there must be some stable subject matter over which the parties disagree. But intuitively, questions about whether there is a stable subject matter have nothing to do with the linguistic question of whether competing claims about that subject happen to be communicated semantically or pragmatically. This intuition is precisely correct, and (2) offers a clear case of how the two issues come apart. (2013, p.13)

In this way, (2) already reveals the danger in drawing quick semantic conclusions from our intuitions about certain disagreements.

§3.6.2 Metalinguistic Disputes

Plunkett and Sundell explain that some disputes are non-canonical because speakers mean different things by their words. The most obvious examples of this are cases in which speakers use words that mean different things in different contexts, i.e. words with context-sensitive meanings. Take for instance the word ‘tall’. Whether the word ‘tall’ is correctly applied to someone is partly determined by whether, for instance, one is speaking about philosophy academics or professional basketball players. I say ‘partly’ because:

36 Our intuitions about genuine disagreement might predominantly track cases of canonical dispute, but the point here is that there is nothing analytic about this connection. The link between our intuitions about disagreement and the canonical nature of certain disputes is a synthetic connection, i.e. one settled by explanatory considerations. Theorists who deploy SMS assume the connection is a priori and necessary—if a dispute is genuine, then it must be canonical.

37 It may be important to underscore at this juncture that the case described by Plunkett and Sundell in (2) is non-canonical because the conflicting content is conveyed by implicature and not literally expressed. There is no further reason to suspect that the two mean different things by their words. This itself suggests that SMS stands on shaky ground.
For the contextualist about this type of expression, the “meaning” of the word does not vary in one sense—it has the same character across contexts. But in another important sense, the meaning of the word is variable—it picks out different properties (and thus has different contents) in different contexts. (2013, p.13)

So whether a particular person can correctly be called ‘tall’ in a particular context is determined by what counts as being tall in light of that context. Someone who can be called ‘tall’ by philosophy academic standards may not be ‘tall’ by professional basketball player standards. In speaking under the two circumstances, whether one is correct or incorrect in applying the word ‘tall’ is up to the parameter for tallness in a given context. I might count as ‘tall’ by philosophy academic standards, but not by professional basketball player standards. If we have a grasp of how this parameter is set, according to Plunkett and Sundell, then we can know how things stand in that context given how the word ‘tall’ is used:

... if, for example, we know the threshold for “tallness”—then sentences involving expressions like ‘tall’ can provide us with useful information about the heights of people and objects around us. There is no reason at all, however, that things cannot work in precisely the reverse direction. If we can hold the relevant height facts constant, then expressions involving gradable adjectives like ‘tall’ can provide us with useful information about the context. The latter kind of usage is described in Barker (2002), who calls it a sharpening or metalinguistic usage of a term. (2013, p.13)

Barker (2002) offers an example of this kind of metalinguistic usage:

Normally, [3] will be used in order to add to the common ground new information

concerning Feynman’s height:

[3] Feynman is tall.

But [3] has another mode of use. Imagine that we are at a party. Perhaps Feynman stands before us a short distance away, drinking punch and thinking about dancing; in any case, the exact degree to which Feynman is tall is common knowledge. You ask me what counts as tall in my country. “Well,” I say, “around here, ...” and I continue by uttering [3]. This is not a descriptive use in the usual sense. I have not provided any new information about the world, or at least no new information about Feynman’s height. In fact, assuming that [‘tall’] means roughly ‘having a maximal degree of height greater than a certain contextually supplied standard’, I haven’t even provided you with any new information about the [satisfaction conditions] of the word [‘tall’]. All I have done is given you guidance concerning what the prevailing relevant standard for tallness happens to be in our community; in particular, that standard must be no greater than Feynman’s maximal degree of height. (2013, p.14)

---

38 I understand Barker here to mean the common ground of information that is accessible to all speakers engaged in the conversation. The speakers engaged in the discussion are deciding about whether Feynman can be called ‘tall’ in the present conversational context. Whether he can be called ‘tall’ or not is determined by the information collated from all speakers engaged in that conversation. Given that information, they can decide on a relevant standard for height (in that context) and then say whether he meets that standard or not, i.e. whether he can be called ‘tall’ or not. So, what Barker means here by ‘common ground’ is likely just shared conversational context.

39 I have tried to clear up any confusion between use and mention here by using tall and ‘tall’ as appropriate. I also replaced ‘truth conditions’ with ‘satisfaction conditions’, given that the latter is more accurate to what Barker means to say, i.e. he means to say something about the relevant conditions under which ‘tall’ can be correctly applied.
Plunkett and Sundell suggest that we extend Barker’s example to a case where speakers disagree about the contextually appropriate usage of the word ‘tall’:

In his (2002), Barker does not consider cases where speakers disagree about the information communicated by this type of usage. But it is easy to see how his example could be extended in that way. After all, another party to the conversation might simply object and say “no, Feynman is not tall”. Just as the original utterance conveyed information not about Feynman’s height but rather the appropriate usage of ‘tall’, so too would the ensuing dispute be a matter not of factual disagreement over Feynman’s height, but rather opposing views about the contextually appropriate usage of ‘tall’. Barker uses ‘metalinguistic’ to refer to the type of sharpening use at play here. Accordingly, we call the corresponding disputes over the correctness or appropriateness of those types of usages metalinguistic disputes. (2013, p.14)

For this extension of Barker’s case, there are widely-settled facts about whether it would be appropriate to call Feynman ‘tall’ in various contexts. The disagreement in this case is about the contextually appropriate usage of ‘tall’ and whether indeed Feynman’s height counts as ‘tall’ in light of this standard, i.e. what standard for tallness is appropriate under that context). The relevant facts about Feynman—his height, the average height of people in various regions, and so on—are fixed. So the dispute over which characterization is best is a disagreement that would be settled relatively straightforwardly, e.g. by the average height in the room. In other cases, things are not so easy. Some metalinguistic disputes are cases in which speakers (at least in part) disagree about the parameters that ought to govern the propositions expressed in a dispute. Plunkett and Sundell call such cases ‘metalinguistic negotiations’ and offer a few examples that evince this kind of dispute.

§3.6.3 Metalinguistic Negotiation

The first example of metalinguistic negotiation that Plunkett and Sundell introduce imagines two speakers—Oscar and Callie—who, in cooking a meal together, come to exchange the following words about a chili they are preparing for a dinner party:

(4) (a) That chili is spicy!
    (b) No, it’s not spicy at all.

According to Plunkett and Sundell, the relevant parameter for spicyness is not antecedently settled. The most natural way to understand Oscar and Callie’s dispute is not in terms of “advancing competing factual claims about some independently determined threshold,” but rather as negotiating about what that threshold should be under the circumstances—they want to settle on an appropriate level of spice, one that is amenable to the collective palate of the dinner guests (2013, p.15). So what makes Oscar and Callie’s dispute a genuine disagreement is that they disagree about the parameters for spicyness that ought to govern the use of the term ‘spicy’. Much hinges on this disagreement in that whether the chili counts as ‘spicy’ or not has practical implications for whether they ought to add more spice or not:

For Oscar and Callie, as for many of us, an agreement amongst all the cooks in the kitchen that the chili can be described as “spicy” plays an important role in collective decision-
making. In particular, it plays an important role in decision-making about whether to add more spice. This may have nothing at all to do with what is analytic about ‘spicy’. Rather, it derives from sociological facts about how people in kitchens act when their creations earn that label. Why should Callie have to refrain from further seasoning when the chili cannot even be described as “spicy”? (2013, p.15)

Intuitively, Oscar and Callie genuinely disagree and the best explanation of this intuition, Plunkett and Sundell say, is captured by viewing them as engaging in metalinguistic negotiation. The alternative view, which centres on incompatibility at the level of literally expressed content, cannot capture this intuition about Oscar and Callie given that the two do not express incompatible contents. Rather, the two disagree about the appropriate use of that term, given their competing understandings of what parameters ought to govern the use of the word ‘spicy’:

Why are such exchanges perceived as disputes, when the speakers fail to assert inconsistent propositions—in fact via their assertion of those propositions—they also pragmatically advocate for the parameter settings by virtue of which those propositions are asserted. The claim that one “spicyness” threshold is preferable to some competing “spicyness” threshold is very much the kind of thing over which two speakers can disagree. (2013, p.15)

To show this, Plunkett and Sundell suggest we imagine Oscar and Callie as having a straightforward canonical dispute about the same topic, in terms of two competing plans for going forward:

(4) (Plan A) Use the word ‘spicy’ such that it does apply to their chili, or…

(Plan B) Use the word ‘spicy’ such that it does not apply to their chili.\(^{40}\)

Oscar and Callie must decide, practically-speaking, which function ought to play the role of mapping the correct uses of the term ‘spicy’, and as such whether their chili is in the extension of that term.\(^{41}\) If the dispute is characterized this way, Oscar and Callie clearly have a genuine disagreement by the lights of the standard way of understanding disagreement, embodied in DRC (2013, p.16). Moreover, Plunkett and Sundell explain:

. . . given the right context—for example, a context where we must coordinate our chili seasoning, or thermostat adjusting, or our basketball picks, or our progressive taxation brackets—such disagreements can be very much worth having. (2013, p.16)

§3.6.4 Content and Character, Sense and Reference

A key feature that the ‘spicy’ case shares with those glossed previously is that it centres on a term that is context-sensitive. In every case, what is up for debate is whether a word ought to denote one function or another, given two competing contextually-set parameters for the use

\(^{40}\) I use the notion of a plan in a broad and non-technical sense here. I do not commit to, for instance, Gibbard’s account of normative judgment as “plan-laden”, i.e. as in Gibbard (2012).

\(^{41}\) This species of dispute is common, according to Plunkett and Sundell: “We think that metalinguistic disputes of [this type] are common. Indeed, we think such usages extend well beyond the kitchen, to disagreements about what should count as “tall” during our basketball draft, or “cold” in our shared office, or “rich” for our tax base. In any such case, speakers each assert true propositions, but they express those true propositions by virtue of the fact that they set the relevant contextual parameters in different ways.” (2013, p.15)
of that word. As Plunkett and Sundell note, the preceding examples of metalinguistic negotiations are “concerned with ‘meaning’ in the sense of content, but not character” (2013, p.16). Plunkett and Sundell follow Kaplan (1989) in making this distinction. According to him, an expression has a meaning in a given context in two different senses. First, an expression has a meaning in virtue of its content, “what it picks out, relative to the context,” i.e. what Frege calls the reference of an expression (2013, p.8). Second, an expression has a meaning in virtue of its character, i.e. “its linguistically encoded, contextually invariant meaning” (2013, p.8). So far, we have looked at cases that involve terms that vary at the first level, at the level of content or reference. For example, recall the case about whether Feynman is ‘tall’. Suppose that two speakers use the term ‘tall’ to denote different heights, respective to the different average heights in their communities:

. . . consider a contextualist view of ‘tall’ according to which speakers use the term to pick out different height-properties relative to a context. If such a view is right, then speakers in different contexts “mean” different things by the word in one sense [i.e. content or reference] but not another. In particular, for those speakers the word ‘tall’ has different contents, but nevertheless it has the same character: something like having a maximal degree of height greater than the contextually supplied standard. (2013, p.8-9)

The idea of metalinguistic negotiation can explain these cases. We can readily view two speakers as genuinely disagreeing even if they use terms with different meanings at the level of character or sense.43

§3.6.5 Negotiation Over Character, not Content

Plunkett and Sundell argue that the idea of metalinguistic negotiation can capture some disputes in which speakers use terms to mean different things, given the different character they associate with their words. Plunkett and Sundell bring this out with an example described by Peter Ludlow in a 2008 paper in which two sports radio hosts engage in a heated debate over whether the racing horse Secretariat is an ‘athlete’ or not.44 The two express their sentiments at in the following simple exchange of words:

(5)    (a) Secretariat is an athlete.
       (b) Secretariat is not an athlete.

In this case, it is unlikely for the dispute between the two hosts to be resolved by some fact about Secretariat. For example, we could not reasonably expect the dispute to be resolved by pointing out how fast Secretariat can run, or how many podium finishes he has achieved. In fact, Ludlow

42 I use the Fregean term ‘function’, though Plunkett and Sundell put this differently, in terms of “how to fix parameter settings for bits of context-sensitive terminology.” (2013, p.16)
43 Plunkett and Sundell choose to follow Kaplan (1989), but qualify that they do not commit to his picture: “Our argument does not require us to endorse the specifics of Kaplan’s picture as against competing ones, such as those offered by Lewis or Stalnaker. Rather we use his terminology to mark the relevant distinctions where necessary, however those distinctions are ultimately to be understood.” (2013, p.8)
44 See Ludlow (2008), p.118.
claims that we could suppose that the two hosts know everything there is to know about Secretariat and yet this knowledge would not settle their dispute. We could suppose that they…

. . . mutually know all the facts about Secretariat’s speed, strength, etc., and what races, awards, medals he won, etc., just as Oscar and Callie mutually know all the facts about the chemical hotness of the chili (2013, p.16).

In spite of knowing all this, it would be reasonable to expect the two to persist in disagreeing. The basic question, according to Plunkett and Sundell, is how we can account for this intuition about the dispute: How could the two genuinely disagree even though they mutually know all the facts?

Plunkett and Sundell claim that we can preserve the intuition that the two radio hosts genuinely disagree about whether Secretariat is an athlete (despite knowing all the relevant facts) by viewing the two as disagreeing about the relevant character of the term ‘athlete’—something that they communicate in their dispute metalinguistically, rather than literally. So one way to preserve the intuition that (5) is a genuine disagreement is to start with how the two speakers are disposed to (or at least typically do) apply the term ‘athlete’. If we have this information, then we are likely to be in a good position to understand why the two conflict despite knowing all the facts about Secretariat. If the two categorically apply the word ‘athlete’ differently, this would explain their disagreement:

Suppose then that one speaker, the speaker of (5a), systematically applies the term ‘athlete’ in such a way as to include non-human animals. The other speaker, the speaker of (5b), systematically applies the term ‘athlete’ in such a way as to never include non-human animals. This holds true even when all of the relevant factual information is on hand, including, as noted, the facts about Secretariat’s speed, strength, etc. (2013, p.16)

If we know as much about the two, then we know the two are apt to mean different things by ‘athlete’. Ludlow concludes just as much. According to him, the two radio hosts mean different things by ‘athlete’ because they disagree about the function that ‘athlete’ ought to denote, say athlete¹ or athlete², in that conversation: “What is at issue is how the term ‘athlete’ should be used in this context. In other words, the dispute is about the character of the expression ‘athlete’” (2013, p.17). According to the speaker of (5a), ‘athlete’ ought to denote function athlete¹ that maps Secretariat to TRUE. According to the speaker of (5b), ‘athlete’ ought to denote function athlete² that maps Secretariat to FALSE. The centre of their dispute is best captured as a kind of negotiation over two plans for going forward. Either accord with the interpretation of ‘athlete’ advanced by (5a), or accord with the interpretation of ‘athlete’ advanced by (5b):

(5) (Plan A) Make ‘athlete’ denote athlete¹.
   (Plan B) Make ‘athlete’ denote athlete².

In this respect, (5) is a sort of practical dispute about what to do, to make ‘athlete’ mean one thing or the other, and thus to apply ‘athlete’ to some things and not others, like Secretariat.⁴⁵

---

⁴⁵ Plunkett and Sundell classify similar disputes about concept choice under the heading ‘conceptual ethics’. According to them, a dispute falls under this heading just in case it centres on matters about how a term ought to be used, i.e. about what function (concept) a word ought to denote. In Ludlow’s case, the dispute centres on whether ‘athlete’ ought to denote one function or the other, but many philosophical debates have
Notice that insofar as the speakers of (5a) and (5b) use ‘athlete’ with different meaning (in terms of character), they both express true propositions. The important feature—the one that vindicates the intuition about disagreement—is the fact that they disagree about the appropriate characterization of ‘athlete’ under the circumstances, i.e. the one that ought to determine whether it would be correct or incorrect to apply the term ‘athlete’ to Secretariat (at least in the shared conversational context). On Ludlow’s view, the two “pragmatically advocate for the concept that they are using and in virtue of which they assert those propositions” (2013, p.17). In sum, the dispute between the two radio hosts counts as a genuine disagreement (in spite of their exhaustive shared knowledge about Secretariat) because the two disagree about whether ‘athlete’ ought to denote one function \((\text{athlete}^1)\) or another \((\text{athlete}^2)\) in that conversation.

According to Plunkett and Sundell, Ludlow’s case is revealing for at least two reasons. First, the case is remarkably familiar:

Many of us are familiar with disputes about whether Missouri is in the “midwest”, or whether Pluto is a “planet”, or whether the American federal anti-drug effort constitutes a “war”. In each case, the relevant facts—the location of Missouri, the size and orbit of Pluto, the contents of the relevant anti-drug policies—are mutually known among the parties to the dispute. And yet it seems that the disputes are, or at least have the potential to be, genuine disagreements in any plausible sense of the term ‘genuine disagreement’. It may not matter very much which states we choose to include in the midwest. But it can matter a great deal whether a policy is meant to address a social ill or advance our cause in a war. As in the case of Oscar and Callie’s debate about the “spicyness” of their chili, metalinguistic negotiations influence and advance more general processes of collective decision-making and action. (2013, p.17)

Second, Ludlow’s case is revealing because it highlights the normative character of many similar disputes:

It is likely that the reason why the two speakers bother to go in for this argument in conceptual ethics (an argument about how to use the term ‘athlete’) is because they ultimately have different normative views about how to live and what to do.46 In this case, perhaps the speakers have different normative views about what sorts of creatures are deserving of which sorts of recognition and rewards. One might therefore be tempted to ask: is this normative issue (rather than “merely linguistic” issue about how to use the word ‘athlete’) not really what their disagreement is about? (2013, p.17)

Plunkett and Sundell grant that disputes are motivated by the background normative views that speakers have, and indeed, this background is often not expressed pragmatically or semantically by speakers when they converse with one another. Imagine for example that two people, Martha and George, are shopping for a car and come to argue over whether Subarus are good cars (2013, p.17). In this case, the exchange is motivated by a practical decision that Martha and George need to make. They need to decide what sort of car they ought to buy together. Plunkett and Sundell

---

46 Where ‘normative’ is used broadly to cover both moral and non-moral cases, i.e. ‘normative’ here could be used to speak about a purely practical issue as well.
say that though this information is certainly relevant in that it explains how the two have arrived at the present dispute, it is not strictly-speaking the immediate topic of that dispute. The immediate topic is what makes a Subaru a good car in terms of some functional definition of ‘good car’, e.g. most reliable, most safe, or most horsepower, or whatever. In any case, what function they pick for ‘good car’ (i.e. what function ‘good car’ denotes) is up to them. Likewise, in the example Ludlow imagines, the background motivation behind the dispute about Secretariat is not the immediate topic of that dispute:

In the case of Martha and George, we wouldn’t be required to build into the content that they express (pragmatically or semantically) the background conditions that explain why they are having this linguistic exchange. It would be just as mistaken to do so in the case of metalinguistic negotiations such as the Secretariat case. (2013, p.18)

According to Plunkett and Sundell, the cases they describe share three key features:

1. They are non-canonical: the speakers involved literally express mutually consistent contents.
2. They are non-canonical in virtue of variation in meaning: the speakers express mutually consistent contents because they do not mean (in the relevant sense) the same things by their words.
3. They nevertheless serve as expressions of genuine disagreement: the speakers involved do accept (and communicate) incompatible contents, and thus satisfy DRCC (2013, p.18).

As such, the preceding examples, as examples of metalinguistic disputes:

... demonstrate not only that speakers who genuinely disagree with each other need not literally express incompatible contents, but that they need not even mean the same things by their words. They might employ context-sensitive terms with the same character but with different contents, as with ‘spicy’. Or they might employ ordinary expressions with entirely distinct characters, as with ‘athlete’. (Of course, in the typical case, when the characters are distinct, the contents will be distinct as well.) Either way, the connection between genuine disagreement and sameness of meaning is broken. From the single premise that some linguistic exchange reflects a genuine disagreement between the speakers involved, nothing follows with respect to the semantics of the expressions employed in the exchange. (2013, p.18)

§3.6.6 Normative Disputes

Plunkett and Sundell claim that some normative or evaluative disputes, i.e. disputes that involve ought statements or categorical imperatives, can be best explained when viewed as metalinguistic negotiations. According to them, the case imagined by Ludlow about Secretariat glossed already is one such example. In this case, the dispute between the two radio hosts about whether the race horse Secretariat is an ‘athlete’ could plausibly be viewed as a kind of normative dispute. This is because ‘athlete’ functions as a normative or evaluative term (i.e. a “thick” normative concept) in that whether Secretariat is an athlete or not implies whether he deserves praise or not.
Ludlow’s case can be used as a model for detecting disputes that are metalinguistic negotiations. Any suitably similar case, according to them, should have two features in common: “(a) they would be taken to involve metalinguistic usage (a distinctive mechanism) and (b) they would be analysed as centring on a question of conceptual ethics (a distinctive normative topic [i.e. about what function a word ought to denote])” (2013, p.19).47 According to Plunkett and Sundell, at least two cases support their claim that the metalinguistic view can be extended to explain the normative or evaluative cases that metanormative theorists (Hare et. al.) are concerned with.

Remember that, according to Plunkett and Sundell, the Moral Twin Earth argument advanced by H&T and other similar arguments depend on the assumption that all genuine disagreements are canonical disputes in which speakers use terms with the same meaning to literally express conflicting contents. On account of the counterexamples above, this assumption is on shaky ground: some disagreements are expressed non-literally, in non-canonical disputes. As such, it looks as though the move from disagreement to shared meaning is too quick. Nonetheless, there is one way that proponents of SMS might curb this worry. So far, we have only discussed non-normative disputes. The cases that SMS theorists are interested in are normative disputes—disputes that involve normative expressions like ‘morally right’ and ‘wrong’. Is there something special about these normative or evaluative cases that vindicates the assumption about shared meaning?

§3.6.7 What counts as ‘torture’?

Let’s suppose that two speakers, Ben and Tracy, come to exchange their views about what practices they believe constitute torture. When they arrive at the topic of waterboarding, the two express their views in the following simple dispute:

(6) (a) Waterboarding is torture.
(b) Waterboarding is not torture.

Ben and Tracy diverge in their functional definitions of torture. Ben, the speaker of (6a) “follows the United Nations in defining torture as any act inflicting severe suffering, physical or mental,

47 The reader may wonder if it needn’t be exhausted by (a) and (b). Couldn’t there be some underlying normative question that explains why the normative issue about concepts arises? Plunkett and Sundell’s example about whether Subarus are good cars is designed to address this worry. According to them, that dispute is motivated by a background normative issue, but this issue is not the immediate topic of that dispute. Plunkett clarifies this point using a slightly different example in a 2015 paper:
First, there are the reasons why two speakers enter into a dispute, or, similarly, what reasons they have for pursuing it. Second, there is the immediate topic of disagreement. For example: suppose Anil and Seth are at the bike shop trying to figure out which bike Seth should buy. They are arguing about whether or not a bike at the store looks cool. They are arguing about this because they each have a view about which bike Seth should buy and so their argument is a way of arguing about that broader topic. But the immediate topic of disagreement is more limited: it is about whether the bike looks cool or not. With this in mind, return to my proposal about the Secretariat case. What Sundell and I propose is that the immediate topic of disagreement in the Secretariat case is the following normative issue: which concept should be used in this context, by being paired with the term ‘athlete’. (Plunket 2015, p.844-845)
in order to obtain information or to punish,” whereas Tracy, the speaker of (6b) “follows former U.S. Justice Department practice in defining torture as any such act inflicting pain rising to the level of death, organ failure, or the permanent impairment of a significant body function” (2013, p.19). Given their respective views on torture, Ben is disposed to use the word ‘torture’ as it applies to waterboarding among other practices, whereas Tracy is disposed to use the word ‘torture’ as it applies to many practices, though not waterboarding. According to Plunkett and Sundell, even if we assume that Ben and Tracy mean quite different things by ‘torture’, there is still room to understand the two as genuinely disagreeing:

Even if we suppose that the speakers mean different things by the word ‘torture’, it is clear that we have not exhausted the normative and evaluative work to be done here. After all, in the context of discussions about the moral or legal issues surrounding the treatment of prisoners, there is a substantive question about which definition is better. By employing the word ‘torture’ in a way that excludes waterboarding, the speaker of (6b) communicates (though not via literal expression) the view that such a usage is appropriate to those moral or legal discussions. In other words, she communicates the proposition that waterboarding itself is, in the relevant sense, unproblematic—a proposition that is, we submit, well worth arguing about. (2013, p.19)

This case is suitably analogous to Ludlow’s case, according to Plunkett and Sundell, because it has two features in common. First, it would be reasonable to understand the two speakers, Ben and Tracy, as using ‘torture’ metalinguistically to express their conflicting beliefs about whether waterboarding is torture. Second, the two could be understood as engaging in conceptual ethics, given the conflict expressed in their metalinguistic use of the word ‘torture’. The topic that their dispute centres on is the question of whether ‘torture’ ought to denote Ben’s concept of torture, the function torture\(^1\), or Tracy’s concept of torture, torture\(^2\). With Ludlow’s case in mind, a detailed account of how this goes should be quite familiar.

According to the speaker of (6a), ‘torture’ ought to denote function torture\(^1\) that maps waterboarding to TRUE. According to the speaker of (6b), ‘torture’ ought to denote function torture\(^2\) that maps waterboarding to FALSE. Given the different functions they use ‘torture’ to denote, the dispute between the two can be best understood as a negotiation over two plans for going forward—either accord with the interpretation of ‘torture’ advanced by (6a), or accord with the interpretation of ‘torture’ advanced by (6b):

(6) (Plan A) Make ‘torture’ denote torture\(^1\).
   (Plan B) Make ‘torture’ denote torture\(^2\).

Like Ludlow’s case, (6) is a practical dispute about what to do, to make ‘torture’ mean one thing or the other, and thus to apply ‘torture’ to waterboarding or not. According to Plunkett and Sundell, the Moral Twin Earth scenario can be explained in a similar manner, as a negotiation over the function that ‘good’ ought to denote.

Plunkett and Sundell use ‘metalinguistic use’ as a technical term to describe cases in which a speaker uses a word in a certain way to non-literally express a belief about what function that word ought to denote. In this case about torture, each speaker uses the word ‘torture’ to express their belief that ‘torture’ ought to denote a function that either maps waterboarding to TRUE or to FALSE. I detail these functions below.
§3.7 Back on Twin Earth

Recall the Moral Twin Earth scenario described by H&T. In the scenario, we are to imagine a distant Moral Twin Earth, identical in every which way to Earth, save for the fact that Moral Twin Earthers use the word ‘good’ in line with a set of natural (functional) properties, T\textsuperscript{d}, that play into a deontological ethical theory, whereas Earthers use the word ‘good’ in line with a different set of natural (functional) properties, T\textsuperscript{c}, that play into a consequentialist ethical theory.\textsuperscript{49} Plunkett and Sundell put this more simply:

Suppose that by ‘morally right’ Earthlings meant something akin to “maximizing overall aggregate utility”. That is: suppose that some form of analytic utilitarianism is true of the term ‘morally right’ as used by Earthlings. In contrast, suppose that some form of analytic Kantianism is true of the term ‘morally right’ for Twin Earthlings, such that, for Twin Earthlings, what their term ‘morally right’ means is something akin to “being in accord with the Categorical Imperative”. (2013, p. 20)

If the Twin Earther were to visit Earth and speak with an Earther about whether various things are ‘morally right’, we would intuitively expect the two parties to genuinely disagree. According to H&T, we could only understand the Earther and Twin Earther as genuinely disagreeing if we suppose they mean the same thing by ‘morally right’—otherwise the two would only speak at cross purposes. This is where H&T (and others) misstep, according to Plunkett and Sundell. We can understand the Earther and Twin Earther as genuinely disagreeing without supposing they mean the same thing by ‘morally right’.

Plunkett and Sundell argue that we can understand the Earther and Twin Earther as genuinely disagreeing even if they mean different things by ‘morally right’. This is because, according to them, “at least some of the disputes involving moral terms between Earthlings and Twin Earthlings may be metalinguistic” (2013, p.20). These cases would be better explained by Plunkett and Sundell’s view, as metalinguistic negotiations. So if Plunkett and Sundell are right, we do not need to assume anything about shared meaning to make sense of disagreement. We can capture our intuitions about disagreement in some cases without resorting to the claim that speakers mean the same thing by their words. Plunkett and Sundell imagine a dispute between the Earther and Twin Earther that, according to them, supports this contention.

Suppose that the Twin Earther—let’s say Chris—visits Earth and comes to exchange moral views with his new Earther friend, Bob. At one point, they have the following dispute:

(7) (a) Lying with the aim of promoting human happiness is sometimes morally right. In fact it often is!
(b) No, you are wrong. It is never morally right to lie in order to promote human happiness.

\textsuperscript{49} Plunkett and Sundell choose to use the word ‘morally right’ whereas Horgan and Timmons choose to use the word ‘good’ here. From what I understand, nothing turns on using one or the other. In this section, I follow Plunkett and Sundell and use ‘morally right’ to avoid confusion.
Imagine that Bob, given his consequentialist leanings, is the speaker of (7a), and that Chris is the speaker of (7b), given his deontological leanings. According to Plunkett and Sundell, we can understand (7) as a genuine disagreement without supposing that ‘morally right’ means the same thing. It seems that we can suppose that ‘morally right’ means different things, respective to Chris and Bob’s different moral views, without endangering the intuition that they genuinely disagree.

According to Plunkett and Sundell, the Moral Twin Earth dispute they propose is a suitable analogue to Ludlow’s case and thus can be explained via their metalinguistic strategy. Remember that, on their view, two features mark a suitable analogue. First, that dispute must involve speakers who use terms metalinguistically, i.e. they use a term (e.g. ‘athlete’) in a way that conveys (non-literally) the function they believe that word ought to denote. Second, the topic of that dispute centres on whether in fact that word ought to denote one function or another. The dispute they envision on Moral Twin Earth exhibits both features. First, Bob the Earther and Chris the Twin Earther can be understood as using the term ‘morally right’ metalinguistically, and second, they can be understood as doing so precisely because they have a tacit disagreement that clearly falls under the umbrella of conceptual ethics, i.e. they disagree about the function that ‘morally right’ ought to denote:

Just as in [the Secretariat case], we might see Bob the Earthling and Chris the Twin Earthling as engaged in a (perhaps tacit) disagreement about which concept is the right one to use in this context. In this context, the issue is arguably about what concept to use in figuring out how to live and in guiding one’s plans about what to do. (2013, p.20)

Plunkett and Sundell continue:

More specifically, it has something to do with which concept should play a functional role that concerns matters of how we navigate our decisions about how to treat others, what to hold each other responsible for doing, and how to live more generally. The disagreement might take place for much the same reason people care about which concept is expressed by ‘torture’: given a certain socio-historical setting—a setting in which certain words (largely independent of which specific concepts they express) fill specific and important functional roles in our practices—participants might care a great deal (and genuinely substantive results could hang on) which concept/word pairings we employ in a given context. The debate between Bob the Earthling and Chris the Twin Earthling would ultimately turn on substantive normative matters, namely, the substantive normative issue of (roughly) which is the better concept to employ in figuring out what to do. Nevertheless, it would not turn on the substantive normative issue that Horgan and Timmons (and perhaps Bob and Chris) seem to think is at issue, namely, what is morally right, in accordance with a meaning of ‘morally right’ shared between the speakers. (2013, p.20)

Following suit with the preceding examples of metalinguistic negotiation, we can spell out the details of this dispute as involving a kind of practical choice between two plans for going forward, respective to the different functions that Chris and Bob believe ‘morally right’ ought to denote.50

50 At this point it is important to reiterate that the aim of extending Plunkett and Sundell’s analysis to cases like Moral Twin Earth is not to offer a competing explanation for any disagreement we can imagine, but rather to show that some philosophers, like Horgan and Timmons, make a mistake when they draw a necessary link between disagreement and shared meaning. Our intuitions about genuine disagreement in
Given the dispute described by Plunkett and Sundell, it is likely that ‘morally right’
denotes two distinct functions for Chris and Bob. For Chris, the Twin Earther, ‘morally right’
denotes the function \( \text{morally right}^1 \) which maps lying with the aim of promoting human happiness
to FALSE. For Bob, the Earther, ‘morally right’ denotes the function \( \text{morally right}^2 \) which maps
lying with the aim of promoting human happiness to TRUE. Given the different functions they
have in mind, there are at least two plans for going forward—they could either align the meaning
of ‘morally right’ with Bob’s function, \( \text{morally right}^1 \), or with Chris’s function, \( \text{morally right}^2 \):

(7) (Plan A) Make ‘morally right’ denote \( \text{morally right}^1 \).

(Plan B) Make ‘morally right’ denote \( \text{morally right}^2 \).

Here, Chris and Bob have a practical choice to make between two plans going forward. As
Plunkett and Sundell say, much hinges on what they choose to do. What we should do with our
lives and how we should treat others is a non-trivial matter. What function ‘morally right’ denotes,
whether that be \( \text{morally right}^1 \) or \( \text{morally right}^2 \) (or something else), has implications for moral
talk and practice. This, for Plunkett and Sundell, highlights the viability and importance,
moreover, of bringing metalinguistic negotiation into metanormative discourse. As Plunkett and
Sundell have shown, some genuine normative disagreements may be expressed in non-canonical
disputes, disputes that, like those in the cases outlined previously, would be better explained by
viewing them as metalinguistic negotiations.\(^{51}\)

In the remainder of this chapter, I extend Plunkett and Sundell’s view about
metalinguistic negotiation to try to explain my Meaning Twin Earth scenario. If Plunkett and
Sundell’s view can be extended to capture this case, then one could potentially object to the
Meaning Twin Earth argument by claiming that our intuitions about the Meaning Twin Earth
scenario can be explained without invoking the thesis that meaning is normative.

§3.8 Meaning Twin Earth

Remember that the scenario I introduce in Chapter Two imagines a distant Meaning Twin
Earth, identical to Earth in every which way, save for the fact that its inhabitants use the word
‘means’ as it plays into a theory of meaning distinct from the one that Earthers ascribe to. These
Twin Earthers apply the word ‘means’ in accordance with an individualist theory of meaning,

\(^{51}\) I say ‘better explained’ because the shared meaning strategy (SMS) cannot even get started in these cases.
If speakers mean different things by the words they use to communicate a dispute, then we cannot make
sense of them as genuinely disagreeing—at least on that view. But, as Plunkett and Sundell demonstrate, it
does seem that some of these non-canonical disputes are cases of people who genuinely disagree and this
evidence counts against SMS.
regulated by some natural (functional) properties, \(M^i\), whereas Earthers apply the word ‘means’ in accordance with a communitarian theory of meaning, regulated by some natural (functional) properties, \(M^c\). Simply put, Twin Earthers think that what someone means is primarily determined by that speaker’s idiolectic understanding of what his or her words mean, whereas Earthers think that what someone means is primarily determined by the linguistic community that that speaker is part of. Given this difference, the two parties would be apt to ascribe meaning differently in various cases. I described a rather difficult case of this sort in Chapter Two.

Recall that in the Meaning Twin Earth scenario, we imagined a lifelong solitary individual named ‘Robinson Crusoe’ on Earth who is the sole inhabitant of an isolated island. Miraculously, the young Crusoe was washed ashore here shortly after his birth. In order to survive on the island, Crusoe has developed a primitive system for identifying one of the staples of his diet, blackberries. Whenever he encounters such a berry, he exclaims “Blackberry!” . Now whether Crusoe can be said to mean something by his words may be the subject of debate, and in particular it may be the subject of debate between a Twin Earther and Earther, if they were to encounter one another.

Imagine that a Twin Earther, let’s say Greg, were to visit Earth and speak with an Earther, Alice, about whether Crusoe speaks meaningfully when he says “Blackberry!” (or anything for that matter). In their conversation, suppose that Alice and Greg come to exchange their views on the matter in a simple dispute:

(8) (a) Crusoe does not mean blackberry by “Blackberry!”
(b) Crusoe means blackberry by “Blackberry!”

In Chapter Two I suggested that the most plausible assessment of a dispute of this sort between the Earther and Twin Earther, the speakers of (8a) and (8b), would be to say that they genuinely disagree and thus mean the same thing by their respective terms for meaning ascription (i.e. the English ‘means’ and the Twin English ‘means’), otherwise they would speak at cross purposes. The basic lesson we have learned from Plunkett and Sundell is that this move from disagreement to shared meaning could be dubious. As we saw in the above discussion, we can apparently explain the intuition that an Earther and Twin Earther disagree about what is ‘morally right’ without resorting to the claim that they mean the same thing by that expression. Can this same strategy be applied in the meaning case?

Let’s suppose that, in accordance with Plunkett and Sundell’s strategy from metalinguistic negotiation, Alice the Earther (8a) and Greg the Twin Earther (8b) mean different things by ‘means’ given the conflicting semantic theories they accept as governing meaning ascription. In particular, they use ‘means’ to denote two distinct functions, respective to each theory: According to Alice the Earther, the speaker of (8a), ‘means’ denotes the function \(\text{means}^1\) which maps Crusoe’s meaning blackberry by “Blackberry!” to FALSE. According to the Greg the Twin Earther, the speaker of (8b), ‘means’ denotes the function \(\text{means}^2\) which maps Crusoe’s meaning blackberry by “Blackberry!” to TRUE. Between these two interpretations, Alice and
Greg have a practical choice to make. They could either align the meaning of ‘means’ with Alice’s function, $\text{means}_1$, or with Greg’s function, $\text{means}_2$:

$$\text{(8)} \quad \begin{align*}
\text{(Plan A)} & \text{ Make ‘means’ denote } \text{means}_1. \\
\text{(Plan B)} & \text{ Make ‘means’ denote } \text{means}_2.
\end{align*}$$

Despite the fact that they mean different things by ‘means’, there is sufficient room to claim that the two genuinely disagree. There is a substantial question about what function ‘means’ ought to denote and much hinges on the function that Alice and Greg select in this case. In particular, the function they select determines whether Crusoe speaks meaningfully or not when he utters things such as “Blackberry!”, and as such, whether we can understand him or any lifelong solitary individual as having a language. So contrary to my assessment of the scenario in Chapter Two, it looks as though there is a way to understand an Earther and Twin Earther as having a genuine disagreement without supposing they mean the same thing ‘means’. In the next section I argue that though this explanation via metalinguistic negotiation is attractive, it cannot work in the case of meaning.

§3.9 Why Metalinguistic Negotiation Can’t Explain Meaning Twin Earth

I argue that metalinguistic negotiation cannot explain Meaning Twin Earth because in this special case, the two speakers must possess a common understanding of the meaning of ‘meaning’ in order to genuinely disagree. There is a limit to metalinguistic negotiation that is embodied by the Meaning Twin Earth scenario: we could not understand two speakers as having a genuine disagreement about what someone means by their words when they issue meaning ascriptions in which ‘means’ is used, without supposing that the two share a common understanding of what it is to mean something at all, for it is this understanding that enables those speakers to communicate that disagreement in the first place. The case about meaning that I envision is rather special in that meaning itself is the subject of debate. If we apply metalinguistic negotiation in this case, as I described earlier, this dispute becomes nonsensical. To show this, I return to the case about ‘spicy’ and then to my case about ‘means’ vis-à-vis Meaning Twin Earth.

In the spicy case, Oscar and Callie disagree about whether ‘spicy’ ought to denote one function or another. Oscar claims that ‘spicy’ ought to denote function $\text{spicy}^1$ that maps their chili ‘C’ to TRUE. Callie claims that ‘spicy’ ought to denote function $\text{spicy}^2$ that maps their chili ‘C’ to FALSE. Going forward, they are apt to negotiate over two sorts of plans, one that accords with Oscar’s preferences and the other to Callie’s preferences:

$$\text{(4)} \quad \begin{align*}
\text{(Plan A)} & \text{ Make ‘spicy’ denote } \text{spicy}^1. \\
\text{(Plan B)} & \text{ Make ‘spicy’ denote } \text{spicy}^2.
\end{align*}$$

In order to negotiate between Plan A and Plan B, the two parties must have a common understanding of what they are negotiating about. In particular, they need to share a common conception of what Plan A and Plan B amount to, i.e. what it would be for Plan A and Plan B to be satisfied—for ‘spicy’ to mean one thing rather than another. A key part of that shared
understanding is a shared understanding of what function the meaning predicate ‘means’ denotes. So assuming that Oscar and Callie share a common understanding of ‘means’ in the background, the two can get on perfectly well in negotiating, even though they do not mean the same thing by ‘spicy’. This much we can grant for the present scenario but for Meaning Twin Earth, things get more difficult.

In the Meaning Twin Earth scenario, the Earther and Twin Earther—Alice and Greg—are at odds over whether ‘means’ ought to denote one function or another. According to Alice, ‘means’ ought to denote \( \text{means}^1 \), which maps Crusoe’s meaning blackberry by “Blackberry!” to TRUE. According to Greg, ‘means’ should denote \( \text{means}^2 \), which maps Crusoe’s meaning blackberry by “Blackberry!” to FALSE. The two can choose between two plans, one that accords with the Alice’s theory of meaning and the other with the Greg’s alternative theory of meaning:

\[(8)\]
(Plan A) Make ‘means that…’ denote \( \text{means}^1 \).
(Plan B) Make ‘means that…’ denote \( \text{means}^2 \).

Remember that in the ‘spicy’ case, the two parties must have a common understanding of what they are negotiating about in order to negotiate between Plan A and Plan B. In particular, they need to know what it would be for Plan A and Plan B to be satisfied, for ‘spicy’ to mean one thing rather than another. In the meaning case, these same requirements hold true. If Alice the Earther and Greg the Twin Earther are to negotiate between Plan A and Plan B, they must assume a common grasp of the meaning of ‘meaning’. Here lies the problem: Given that ‘meaning’ itself is the subject of dispute and that speakers don’t mean the same thing by ‘meaning’ in the first place, the metalinguistic strategy cripples itself in this special case. It is as if, as Wittgenstein once remarked, we have tried to saw off the branch on which we sit.\(^{52}\) It is impossible for two parties to have a debate about the meaning of ‘meaning’ without the proverbial branch that enables that dispute to take place in the first place. So though, at least for the sake of argument, I agree with Plunkett and Sundell that metalinguistic negotiation may better explain some disputes, like Moral Twin Earth, it cannot capture the focal dispute on Meaning Twin Earth. Two objections might be raised against this argument.

First, it may be objected that given the practical dilemma I pose in (8) above, we could suppose that Alice the Earther and Greg the Twin Earther mean the same thing by ‘denote’ and yet do not mean the same thing by ‘means’. In other words, the two could be viewed as sharing in the notion of reference but not in the notion of meaning. If possible, then the negotiation sketched above, between (8a) and (8b) would be a plausible way to view Alice and Greg as genuinely disagreeing without meaning the same thing by ‘means’—thus avoiding my criticism that the two must share a notion of ‘means’ in order to disagree at all. Though I believe that metalinguistic negotiation is a helpful way to view many disputes, I fail to see how the preceding rebuttal can save that explanatory strategy in the meaning case. It is hard to imagine how two

---

\(^{52}\) See *Investigations* §55.
people could have a common grasp of how denotation is fixed without a common grasp of ‘means’. Supposing that sense determines reference, if one shares a theory of how reference is determined with another, then one can reasonably expect to likewise share a notion of meaning. This, I take it, captures the standard position. I do not see how the two could come apart in a way that would damage my argument, but I am open to the possibility. If there is a way they could, then this could pose a problem for my proposed understanding of Meaning Twin Earth, but the onus is squarely on my opponent at this point to come up with the relevant detailed argument.

Second, it might be objected that since Plunkett and Sundell claim that in the non-canonical case the disagreement concerns a content pragmatically communicated, semantics doesn't come into it, so that my objection fails. However, this seems to miss the essential point I was making, namely, that in order for metalinguistic negotiation to be possible, those involved in the negotiation need to be assumed to have a common conception of what they are negotiating about, and that this is precisely what is lacking in the meaning case we are concerned with.

§3.10 Conclusion

In this chapter I evaluated whether Plunkett and Sundell’s objection to Moral Twin Earth could be convincingly applied against the Meaning Twin Earth argument advanced in Chapter Two. I explained that though Plunkett and Sundell’s objection may succeed against H&T’s argument in the moral case, it cannot succeed against my analogue of that argument. Metalinguistic negotiation may helpfully explain many disputes, including perhaps the one imagined on Moral Twin Earth, but it cannot explain Meaning Twin Earth. In the next two chapters I evaluate two further objections to H&T’s Moral Twin Earth argument and consider whether these objections can be applied against the Meaning Twin Earth argument.
Chapter Four: Copp ‘Milk, Honey, and The Good Life on Moral Twin Earth’

§4.1 Introduction

In this chapter, I outline David Copp’s two objections to Horgan and Timmons’ Moral Twin Earth argument from his “Milk, Honey, and The Good Life on Moral Twin Earth” (2000) and argue, in turn, that each of these objections does not pose a threat to the Meaning Twin Earth argument presented in Chapter Two.

§4.2 Milk, Honey, and The Good Life on Moral Twin Earth

Copp (2000) claims that the Moral Twin Earth argument advanced by Horgan and Timmons (1992a) fails to undermine synthetic ethical naturalism. Copp proposes two replies against Horgan and Timmons (H&T): The first accepts the intuition that the two gather from their scenario and claims that this intuition can be explained without abandoning naturalism. The second revises H&T’s assessment of the scenario using a broader interpretation of Putnam’s work in order to escape H&T’s conclusion. Each of Copp’s replies functions as an independent objection, though Copp admits the latter is the more formidable of the two: A Putnam-style moral semantics can explain Moral Twin Earth and so “naturalists can begin again to issue visas for travel to Moral Twin Earth” (Copp 2000, p.134).

§4.3 Brief Review

Recall that on reflection, the Moral Twin Earth scenario elicits linguistic intuitions that are significantly different from the intuitions we have about the classical Twin Earth scenario introduced by Putnam (1975). In the scenario that H&T describe, our linguistic intuition is that the Earther and the Twin Earther genuinely disagree when they use the term ‘morally right’ to exchange their moral views, where the use of ‘morally right’ is regulated by different functional properties in each case, respective to their different moral theories. For the Earther, the use of ‘morally right’ is causally regulated by some natural (functional) properties, in line with his normative ethical theory \( T_e \); for the Twin Earther, ‘morally right’ is causally regulated by some alternative natural (functional) properties, in line with his normative ethical theory \( T_d \). Given this difference, it is easy to come up with examples of disputes that may occur between the two—disputes in which the Earther and Twin Earther express different moral judgments, and use the
word ‘good’ to communicate these views to each other. Plunkett and Sundell (2013) suggest the following dispute:

(7)  
(a) Lying with the aim of promoting human happiness is sometimes morally right. In fact it often is!
(b) No, you are wrong. It is never morally right to lie in order to promote human happiness.

Intuitively, (7) expresses a genuine disagreement. H&T argue that our linguistic intuitions about disagreements such as this, which take place in their Moral Twin Earth scenario, count as evidence against Boyd-style synthetic ethical naturalism, or what they call ‘causal semantic naturalism’ or ‘CSN’. According to them, these naturalist theories imply (absurdly) that the Earther and Twin Earther simply speak at cross purposes on account of their meaning different things by their moral terminology (i.e. which refer to different properties, respective to Tc and Td). Copp argues that there are at least two ways that Boyd and the other synthetic naturalists can avoid H&T’s conclusion.

§4.4 Copp’s First Reply: Accommodating Moral Twin Earth

In his first reply to H&T, Copp concedes the general intuition that the two extract from their Moral Twin Earth scenario. The challenge for the synthetic ethical naturalist is to offer a story that captures this intuition: “the intuition that the Earthlings and Twin Earthlings have substantive moral disagreements, and not merely verbal disagreements, is robust and widely shared. For this reason, naturalism must try to accommodate the intuition in some way” (2000, p.119). One way that the naturalist can do this, according to Copp, is by showing how there is a sense in which the Earther and Twin Earther can be viewed as meaning the same thing and as genuinely disagreeing despite the fact that the two mean different things in the “philosophically preferred” sense (on account of their terms referring to different properties). This is made possible by deploying a weaker notion of shared meaning that Copp argues can still capture our intuitions about the scenario.

Copp claims that we can understand the Earther and Twin Earther as meaning the same thing and as genuinely disagreeing even if they use ‘morally right’ to refer to different natural (functional) properties: “even if the Earthling term “wrong” expresses a property that is distinct from the property expressed by the Twin-Earthling term “wrong” – which, of course, is what H&T think that any form of synthetic moral naturalism would be committed to saying in a relevant Moral Twin Earth thought experiment – it still might be the case that the Earthlings and Twin Earthlings disagree morally in the situation we have imagined” (2000, p.120). According to Copp, this is possible because, given the construction of the scenario, we can reasonably expect moral terms to be roughly inter-translatable between English and Twin English, even if they mean

53 You could use ‘morally right’ or ‘morally wrong’ or indeed some other moral predicate, but for our purposes nothing in particular hinges on this.
different things in the “philosophically preferred” sense of referring to different properties. Copp says that we can gather this much from the scenario because H&T describe in detail the various ways in which Earthen and Twin Earthen moral discourse is uniformly similar, e.g. that moral and twin-moral terms are used in similar ways to apply to similar sorts of actions, that Twin Earthers are disposed to perform the sort of actions that they apply ‘morally right’ to, etc. Given these details, Copp says it is highly likely that the best translation of the Twin Earther’s term ‘morally right’ is indeed the Earther’s term ‘morally right’ (and likewise for ‘wrong’, etc.). H&T mention this in their original discussion; they explain that a group of explorers from Twin Earth would be likely to translate twin-moral terms in just this way, and these explorers would likely be correct in doing so (2000, p.121, Horgan and Timmons 1992a, p.164). According to Copp, we only need inter-translatability at this general level in order to treat the Earther and Twin Earther as meaning the same thing (on this weaker notion of ‘same meaning’) and thereby as genuinely disagreeing. Copp introduces a useful analogy to show how his preferred notion of shared meaning is the best we can expect to use successfully in many cases, including the one described by H&T:

To understand this point, it needs to be understood that translation is more like trying to find someone who looks enough like you to pass muster in a police lineup than it is like trying to find your identical twin. Many people lack identical twins, of course, and many words lack precise synonyms in certain other languages. So it is possible that the English term “wrong” is the best translation into English of the twin-English term “wrong” even though, by stipulation, the terms express different properties. If this is so, then even though corresponding moral and twin-moral terms express different properties and therefore have different “meanings” in the philosophically preferred sense of the term, there is also a sense in which they might have the same “meaning”. (Copp 2000, p.121-122)

The main point for Copp is that if we understand ‘same meaning’ in this other sense—in terms of rough inter-translatability—we can view the Earther and Twin Earther as meaning the same thing by the English ‘morally right’ and the Twin English ‘morally right’ and therefore as disagreeing, despite the stipulated difference in reference between those terms.

According to Copp, we could not reasonably expect there to be a perfect translation between twin-moral and moral terms in the philosophically preferred sense that H&T propose. The best we can do, Copp claims, is to look for the closest translation of the English term ‘morally right’ in Twin English. Given the description of the scenario, the closest thing is the Twin English term ‘morally right’, which plays a very similar though not identical role (twin-moral terms denote different properties than moral terms do). This much, Copp says, should highlight that H&T are too quick to “take our intuitions in the Moral Twin Earth scenario to be strong evidence that

54 Notice that the best available translation here could still, in absolute terms, be a bad one. For example, the best discus thrower in the philosophy department may not be much of an athlete. In this case, Copp must mean something stronger by ‘best’ than ‘best available’, otherwise even though the best translation of ‘F’ might be ‘G’, there might be no plausible sense in which they can be taken to mean the same thing. What Copp needs to clarify for us is when a candidate translation “passes muster” to use his own term.

55 My emphasis.
corresponding English and Twin English moral terms [must] express the same property” if the Earther and Twin Earther are to be understood as meaning the same thing by those terms (2000, p.124). As Copp has demonstrated, there is room to argue that the contrary is true: “there is an interpretation of our intuitions on which they are compatible with the thesis that corresponding terms in their languages express distinct properties. Our intuitions can therefore be accommodated by semantic theories of the kind that are in question” (Copp 2000, p.124). If Copp is right, then there is an avenue of escape for the naturalist. If the less robust notion of shared meaning that Copp introduces is sufficient for capturing the disagreement between the Earther and the Twin Earther, then the naturalist can head off the H&T challenge: There is a sense in which the two parties mean the same thing and therefore genuinely disagree, even if they strictly speaking do not refer to the same properties in using their words (and thus do not mean the same thing in the philosophically preferred sense assumed by H&T).  

I argue that this first reply from Copp is unconvincing on the grounds of two objections.

§4.4.1 First Response: Can We Dispose of the Principle that Meaning Determines Reference?

Copp’s first reply amounts to the suggestion that if we use a less robust notion of meaning, we can view the Earther and Twin Earther of the Moral Twin Earth scenario as meaning the same thing and as genuinely disagreeing, even though strictly speaking moral and twin-moral terms refer to different properties. The cost of this reply is that we must dispose of something like the general principle that meaning determines reference, e.g. Frege’s principle that ‘sense determines semantic value’. This can be shown by subtly reframing Copp’s first reply. Basically, Copp argues that if we dispose of the idea that sense determines reference, then we can view the Earther and Twin Earther as meaning the same thing by ‘morally right’ even though by stipulation the English ‘morally right’ and the Twin English ‘morally right’ refer to different properties. But, in supposing that (1) the two speakers mean the same thing by ‘morally right’ (in each of their languages) and that (2) the two expressions do not share in reference, it must be the case that (3)

56 It is worth pointing out that there is another potential line of objection against Copp that takes the form of an ad hominem: Notice that Copp’s preferred notion of shared meaning could potentially undermine Putnam’s original Twin Earth argument. In Putnam’s classical scenario, the Earther and Twin Earther could not plausibly be viewed as genuinely disagreeing, given that the two use the term ‘water’ in their respective languages to denote different properties, i.e. H$_2$O and XYZ. As such, the two could only speak past one another, intuitively speaking. If we deploy Copp’s preferred notion of ‘shared meaning’ here from the first reply, this causes problems: On this notion, we can view the Earther and Twin Earther as meaning the same thing and as genuinely disagreeing, contrary to Putnam’s assessment. This is plausible because, on Copp’s view, the English term ‘water’ would be the best translation of the Twin English term ‘water’. As in Horgan and Timmons’ case, this much can be gathered from the construction of the scenario. So given that the relevant terms are roughly inter-translatable, there is a sense in which the Earther and Twin Earther can be viewed as meaning the same thing and genuinely disagreeing. It is very unlikely, as will be highlighted in the second reply, that this is part of Copp’s agenda. If Copp wishes to retain a Putnamian account of meaning, he cannot use the notion of ‘shared meaning’ that his first reply depends on. Likewise, if the reader is persuaded by Putnam’s classical Twin Earth argument, then the reply should also be abandoned.
the meaning of ‘morally right’ does not determine which property that expression refers to. 

Ceteris paribus this is a bad thing, even if it allows us to meet the Moral Twin Earth argument.

I argue that we should reject Copp’s first reply because we should, in general, aim to preserve the principle that meaning determines reference, not because we must necessarily preserve the Fregean notion of sense, but because giving it up would mean giving up any kind of view that, like Frege’s, bases the compositionality of meaning on the compositionality of reference. Doing this would seriously limit our options for saying anything systematic about the intuitive notion of meaning and in particular, about the compositionality of meaning. As such, we could only accept Copp’s first reply at the cost of losing the ability to deal with more fundamental issues that have to do with the compositionality of meaning and reference. So crucially, the objection here stands even if we do not accept every detail in Frege’s account of sense. This is accentuated by the fact that Russellian accounts of meaning that attempt to effectively dispense with the Fregean notion of sense (and thus the principle that meaning determines reference) do not do well in accounting for the compositionality of meaning.

§4.4.2 Second Response: Disagreeing for ‘Most Practical Purposes’?

Even if Copp can address the problems that arise from disposing of the idea that meaning determines reference, the notion of genuine disagreement that Copp ends up with is insufficient for capturing the intuition that the Earther and Twin Earther genuinely disagree. Essentially, what Copp’s first reply amounts to is the thought that although two speakers do not genuinely disagree, for most practical purposes, we can treat them as if they mean the same by ‘morally right’ and disagree. In this respect, Copp is effectively saying that nothing bad will happen if we pretend that the two speakers genuinely mean the same thing. However, this can be conceded by H&T without damaging their argument. That is, they can concede that the Earther and Twin Earther can, for most practical purposes, be treated as if they genuinely disagree. Unfortunately for Copp, this does not enable the synthetic naturalist to escape H&T’s conclusion. The real challenge for the naturalist is to capture the intuition that the Earther and Twin Earther are actually disagreeing.

57 In order to show the importance of the principle that meaning determines reference or that sense determines semantic value, in a brief appendix at the end of the dissertation, I provide a reminder of Frege’s reasons for introducing the notion of sense. To repeat, the objection to Copp is not that he would be committed to giving up a Fregean notion of sense, but that in giving up the principle that meaning determines reference, he would be giving up the possibility of any theory which, like Frege’s, builds an account of the compositionality of meaning/sense on the basis of the compositionality of reference/semantic value. See Appendix: Frege on Meaning, Reference, and Compositionality, p.162-168.

58 It may be objected that this reply begs the question against accounts of meaning that are built around a term’s functional role. I have been careful to avoid doing this. I have not argued that functionalist accounts of meaning must fail if they dispose of the principle that meaning determines reference. Instead I have simply argued that there are good reasons, detailed in the Appendix, for retaining that principle. It is certainly open for an account of meaning to dispose of the principle anyway. What I have attempted to emphasise is that this may cause some problems down the line, especially with regards to capturing the compositionality of meaning.

59 See Miller 2007 p.72-79 for a detailed exposition of these issues for Russellian theories of meaning.
given the stipulated details about the Moral Twin Earth scenario. Copp simply fails to answer this challenge.\footnote{It is no response to argue that Copp might be working with a ‘mundane’ notion of meaning entirely dissociated from systematic semantic theorising: for one thing, a systematic semantic theory which dissociated itself from the intuitive notion of meaning in this way would be in danger of depriving itself of a subject matter, given that one central point of such a theory is to clarify our ordinary notion.}

$\S 4.5$ Copp’s Second Reply: ‘A Theory of Error’

In his second reply to H&T, Copp argues that synthetic ethical naturalism can accommodate our intuitions about Moral Twin Earth by drawing on some strategies used by Putnam. The aim of this reply, according to Copp, is not strictly-speaking to defend synthetic ethical naturalism against all possible objections, but simply to defend the view that a Putnamian semantics can be used in explicating synthetic ethical naturalism.\footnote{As Copp puts it:}

I should stress that my goal is simply to defend synthetic moral naturalism against the Moral Twin Earth argument. I do not aim to defend it or Putnamian semantics against all worries that could be raised. Accordingly, I aim merely to show that a plausible articulation of Putnamian semantics could be used in explicating synthetic moral naturalism. (2000, p.125)

The first reply to Horgan and Timmons’s argument will seem unsatisfying to many philosophers, including, presumably, Horgan and Timmons. For as far as the first reply goes, even if the best English translation of what the Twin Earthlings say in the lying example is that lying would be “wrong”, it is nevertheless true in the example that, in saying that lying would be twin-wrong, the Twin Earthlings attribute to lying a property that the Earthlings do not deny it to have when they say that lying would not be wrong. As far as the first reply goes, the English term “wrong” has a different meaning in the philosophically preferred sense from the corresponding Twin English term. It may seem to many philosophers that any theory with this implication is unacceptable. For it may seem intuitively clear that the Earthlings disagree with the Twin Earthlings precisely about the truth of the proposition about lying expressed by the Twin Earthling sentence, “Lying is wrong”. It may seem clear that the Earthlings are denying exactly what the Twin Earthlings affirm when they say lying would be “wrong”. (2000, p.124)

On Copp’s view, a Putnamian moral semantics can explain this intuition.

Copp argues that we can capture the intuition that (1) the Earther and Twin earther refer to the same property by their moral and twin-moral terms (the English ‘morally right’ and the Twin English ‘morally right’), and that (2) the two genuinely disagree insofar as they use that same term to express conflicting propositions about whether, for example, lying is morally right, i.e. (7) above. We can achieve this by paying careful attention to some of the strategies used by Putnam in certain non-moral examples. Copp expands on this thought in the following passage:

[A Putnamian moral semantics] can accommodate the [above] intuition by deploying strategies some of which Putnam himself uses to explain the possibility of errors of various kinds in non-moral examples, together with a distinction that Putnam draws between speakers’ referential intentions and their interests. The resulting semantics implies that corresponding Earthling and Twin Earthling moral and twin moral terms in Twin Earth scenarios express the same property if they express any property at all. It
therefore would imply that in the lying example either the Earthlings or the Twin Earthlings are mistaken. It would also imply that there is no coherent Moral Twin Earth scenario of the sort described by Horgan and Timmons. That is, in any coherent scenario of the kind described by Horgan and Timmons, the semantics would imply that corresponding Earthling and Twin Earthling moral and twin moral terms express the same property. This is the second reply to the Moral Twin Earth argument. In discussing it, we obviously need to leave behind the special stipulation of the Master scenario that Earthling and Twin Earthling moral and twin moral terms express different properties. (Copp 2000, p.124-125)

The price of this Putnamian approach, according to Copp, is that we must admit that one of the two parties is mistaken, i.e. either the Earther or Twin Earther makes a mistake in their judgements about what’s ‘morally right’ in the Moral Twin Earth scenario.62

§4.5.1 Fallback Conditions and Referential Intentions

Putnam understood that his semantic theory had limits. He understood that the meaning of a word is not always fixed by some hidden structure at, say, the level of chemical composition. In some cases there is more than one such structure, in others there is no hidden structure at all. Putnam recognized that there are many cases where this occurs, for example:

Some diseases, have turned out to have no hidden structure (the only thing the paradigm cases have in common is a cluster of symptoms), while others have turned out to have a common hidden structure in the sense of an etiology (e.g. tuberculosis). Sometimes we still don’t know; there is a controversy still raging about the case of multiple sclerosis. (Putnam 1975, p.160)

The way that Putnam dealt with these cases was to point out how certain ‘fallback conditions’ come into play whenever no hidden structure can account for the meaning of a word. So in these cases, some alternative conditions determine the meaning of that term, e.g. some superficial or functional properties. Putnam describes a few such cases to bring this out.

Suppose that I point to a glass containing a clear liquid and state that “This liquid is ‘water’”. According to Putnam, this ostensive definition of ‘water’ (like many others) comes with certain empirical presuppositions. For one, my definition presupposes that the liquid in the glass is the same type of liquid that speakers in my linguistic community, including myself, apply the

---

62 Copp ultimately argues that the ethical naturalist, as a kind of realist, should concede that one of the parties in the Moral Twin Earth scenario is mistaken. The alternative is to admit that (roughly) both parties say something that is true when they say that lying is ‘morally right’ and that lying is not ‘morally right’. This would be unpalatable for any realist for it would imply that moral relativism is true—something the realist has a deep antipathy towards and will avoid at all costs. This heads off the potential charge that altering the scenario in the way Copp suggests would be ad hoc. According to him, there is sufficient room to understand Moral Twin Earth in this additional way, though Horgan and Timmons apparently fail to realise this. Copp takes this to be the main error in Horgan and Timmons’ assessment of the scenario. Horgan and Timmons (2000) sum up this ‘error’ on their own terms:

In his second reply to the Moral Twin Earth argument, Copp emphasizes that Putnamian semantics has the resources for explaining certain types of error in belief about the subject matter associated with the relevant terms under scrutiny. He thinks that our Moral Twin Earth argument works by not allowing that a moral community can have false beliefs about the referents of their moral terms and by not allowing that they make mistakes in what they take to be instances of actions having certain moral properties. (2000, p.147)
word ‘water’ to. If this presupposition is false and the glass contains, say, gin rather than water, then certain fallback conditions come into play and “I do not intend my definition to be accepted”:

In general Putnam says, ostensive definition has certain empirical presuppositions and when these presuppositions are false, “a series of, so to speak, ‘fallback conditions’ becomes activated” . . . for natural kind terms, “if there is a hidden structure, then generally it determines what it is to be a member of the natural kind, not only in the actual world, but in all possible worlds.” He adds that “the local water, or whatever, may have two or more hidden structures – or so many that ‘hidden structure’ becomes irrelevant, and superficial characteristics become the decisive ones”. (Copp 2000, p.125, Putnam 1975, p.160)

Similar conditions determine meaning for words like ‘milk’ and ‘jade’. In these cases, there simply is no hidden structure shared by all samples of milk or jade and so, according to Putnam, the superficial characteristics become decisive and determine meaning for the words ‘milk’ and ‘jade’. In the case of ‘milk’, the word ‘milk’ refers to any liquid that shares the common functional and genetic property of being “a liquid produced by the mammary glands of a female mammal that has recently given birth, in order to nourish its young” (2000, p.126). In the case of ‘jade’, the word ‘jade’ refers to any stone that possesses the shared disjunctive property of being a stone with the chemical composition of either jadeite or nephrite. For either word, there is no shared hidden structure that grounds meaning:

Both jadeite and nephrite fall in the extension of the term “jade”. There is no hidden structure shared by all samples of jade. All samples do have the disjunctive property of being either jadeite or nephrite, but this of course does not mean that there is a common structure in any relevant sense. Now, just as there are different kinds of jade, there are various kinds of milk, each with a somewhat different chemical composition. Cow milk is chemically different from goat milk. But, even though there is not a chemical composition shared by all milk, there is a “hidden structure” shared by all milk, for all kinds of milk have the functional and genetic property of being, roughly, a liquid that is produced by the mammary glands of a female mammal that has recently given birth, in order to nourish its young. (2000, pp.125-126)

Ultimately, according to Putnam, the meaning of ‘milk’ is linked to the referential intention we use that word with; it is linked to the intention to refer to whatever liquid is suitably similar to the local samples, to “refer to whatever has the relevantly same nature as the local samples, given that milk has a shared nature at the level of functional and genetic properties” and no hidden structure at the level of chemical properties (2000, p.126).

The reason that the functional and genetic properties are semantically relevant (i.e. meaning determining) in the case of ‘milk’ as opposed to some alternative properties is because it is these properties that secure the extension of ‘milk’ that we want to capture. Because milk does not have a shared nature at the chemical level, our use of the term ‘milk’ must track some feature shared at an alternative level. There must be some other similarity between our uses of that term that determine its meaning. Now what similarity (or similarities) is relevant is determined by what we want ‘milk’ to mean, i.e. our referential interests in using that word. If we believe that ‘milk’ refers to the kind of liquid produced by female mammals shortly after birth, then we should intend to refer to whatever shares that functional and genetic similarity. These
properties become the decisive ones that determine the meaning of ‘milk’. If on the other hand we believe that ‘milk’ refers to more things, like say coconut milk or soy milk, then we should intend to refer to whatever kind of liquid shares, for instance, certain superficial properties with the local samples of what we call ‘milk’, more broadly construed (e.g. or maybe the slightly off-white liquid that goes well with Cheerios or Weetbix). So, given different interests, different kinds of properties become the relevant ones for meaning.63

§4.5.2 Back to Twin Earth

Recall that, according to Putnam, the Twin Earth scenario he describes yields the general linguistic intuition that whatever Twin Earthers call ‘water’ on Twin Earth (the liquid composed of the molecular compound XYZ) is not the same thing as what Earthers call ‘water’ on Earth (the liquid composed of the molecular compound H2O), i.e. the referent of the Twin English word ‘water’ (XYZ) is different from the English word ‘water’ (H2O). The basic reason that we have this intuition about the scenario, according to Copp, is that we use the term ‘water’ with the shared referential intention to refer to the liquid with the same hidden structure as the local samples around us (2000, p.127).64 Like the preceding examples, we use the word ‘water’ with a certain referential intention. We use the word ‘water’ with the general intention to refer to whatever has the same hidden structure as the watery stuff around us, i.e. the stuff that we use the word ‘water’ to refer to.

The key difference when we talk about ‘water’—as opposed to any of the other preceding examples—is that water does in fact have a hidden chemical structure that is common to all samples, the chemical compound H2O. The difficulty for the other cases mentioned is that not so much is true for things like jade and milk. Milk and jade do not have a common hidden structure and so the meaning of the words ‘milk’ and ‘jade’ cannot mean what they do in virtue of some story about their hidden structure at the level of, say, chemical composition. The stuff that we call ‘milk’ and ‘jade’ do not share anything uniformly at the chemical level, and so ‘milk’ and ‘jade’ must have a meaning that is determined by some alternative story. Putnam’s point is that we can give a semantics for these terms by falling back to the referential intentions with which we use those words. As Putnam has argued and Copp has underlined, ‘milk’ and ‘jade’ mean what they do because we use those words with the intention to refer to whatever stuff is suitably similar at the level of some superficial or functional properties. As we know, there is no hidden structure

63 Notice that all that has been said about what properties are semantically relevant is compatible with there being multiple similarities or shared properties for any given term. So even if there are many properties that are shared, the ones that are relevant is a matter that is determined by our interests. In this respect, we have a kind of practical choice to make between whether to make one or another set of properties the decisive ones that determine the meaning of a word. In any case, Copp’s point is that our referential intentions play a basic role in determining meaning, no matter what the content of those intentions is, i.e. what properties we intend to refer to. So for ‘milk’, it could easily be some alternative set of properties that determine meaning, depending on what our interests are.

64 As Copp puts it, “the intention with which we actually use the term “water” on Earth is to refer to whatever liquid has the same basic nature as the stuff around us” (2000, pp.126-7).
shared by all samples of milk or jade, so some other properties must determine the meaning of the terms ‘milk’ and ‘jade’. Given that no such hidden structure exists, we must “fall back” to some alternative similarity, in terms of some other shared properties between what we call ‘milk’ and ‘jade’ respectively. On Putnam’s view, we can figure out what properties are semantically relevant by appealing to the referential intentions that we use those words with. Whatever those properties are is determined by what we want to do with those words, with what we want them to pick out and with our broader interests in doing so. In any case, according to Copp, these examples reveal a kind of constraint that should be placed on any plausible semantic theory. It is that any such theory must “incorporate a theory of error”:

In Putnam’s view, for example, the referent of a term, such as “water”, is determined by the actual nature of the stuff that is in the local samples, not by our beliefs about the nature of the stuff. If our beliefs about the nature of the stuff are mistaken, we can make mistakes in using the term. A plausible theory obviously must allow for such mistakes. (2000, p.127)

Copp argues that the same constraint holds for any suitable Putnamian moral semantics. That is, any suitable semantics of that sort must have the resources to account for the possibility of mistakes:

It is the moral facts together with our referential intentions that determine the extension of “wrong”. If our moral views are mistaken, then we might make mistakes in calling actions “right” or “wrong”. Or if there are certain false presuppositions of our usage, then fallback considerations would come into play. The theory needs to allow for such mistakes as well. (2000, p.127)

Copp argues that these sorts of considerations open up space for a reply to H&T.

§4.5.3 Making Room for Error

Copp’s main charge against H&T is that their argument neglects the fact that, loosely speaking, moral terms could be more like ‘milk’ and ‘jade’ than they are like ‘water’.65 If Copp is right, then ‘morally right’ could, under some circumstances, trigger the sort of fallback conditions that determine the meaning of ‘milk’ and ‘jade’ (i.e. some alternative properties relevant to our referential intentions). As mentioned above, the circumstances that Copp has in mind are cases in which we apply the term ‘morally right’ to convey a moral view that is mistaken. H&T fail to make room for the possibility that one of the two parties, the Earther or Twin Earther, in their scenario could make a mistake in applying ‘morally right’ as he or she does. In addition, they do not realize that a Putnam-inspired moral semantics can readily account for this possibility. This is the key problem for H&T:

The most important error in [Horgan and Timmons’] reasoning is that it does not take proper account of the strategies that are available to Putnamian semantics for dealing with

---

65 I say ‘loosely speaking’ because, as Copp grants and most agree, moral terms are normative whereas terms like ‘milk’ and ‘jade’ are not. Copp’s point is that the way that meaning is determined for moral terms might be more like the way it is determined for terms like ‘milk’ or ‘jade’ rather than ‘water’. The use of moral terms may not meet Putnam’s ‘gold standard’ of being causally regulated by some hidden structure at the molecular level.
mistakes. Speaking of “wrong actions” might be like speaking of “unicorn horns”. Calling a lying action “wrong” might be like calling some gin “water”. Or denying that a lying action is “wrong” might be like denying that female dogs give “milk”. (2000, p.128)

The second problem for H&T is that they do not respect the fact that, on Putnam’s view, what similarities are ‘semantically relevant’ is a matter settled by our referential intentions and interests:

The second error is that the reasoning does not take into account the fact that, for Putnam, some similarities among the items that are included in the extension of a term by speakers might be irrelevant to determining its extension. Putnam says that “important” similarities are what matter and, he says, the notion of importance is “interest relative”. (2000, p.128) For example, when it comes to the meaning of ‘milk’, the similarities that are semantically relevant are the functional and genetic properties that are shared by the local samples of milk. The reason that these are the relevant similarities is that, when we use the word ‘milk’, we have a certain interest—to refer to whatever has the relevantly same nature as the local samples that we call ‘milk’. Consequently, Copp explains, the meaning of ‘milk’ has this referential intention built-in. The lesson from this example (and others), according to Copp, is that “in in applying the semantics, we need to fix on respects of similarity that are semantically relevant, given speakers’ intentions and interests” (2000, p.128). In light of this lesson, we should respect the fact that moral terms are also used with certain referential intentions and interests.

Copp claims that we need to settle two “key matters” about moral terms, as they are used in H&T’s scenario, before a Putnamian semantics can hope to capture the meaning of those terms:

First is the content of speakers’ referential intentions in using corresponding English and Twin English moral and twin moral terms. Second is the kinds of similarities that would be semantically relevant, given the interests with which speakers use these terms. (2000, p.129)

According to Copp, both matters are settled by the details that H&T build into the Moral Twin Earth scenario “together with our semantic intuitions and Putnam’s theory” (2000, p.129). First, H&T assume that the relevant terms in English and Twin English are distinctively moral terms—terms that are distinct in kind from, say, the classificatory terms of the social sciences or the terms of etiquette (2000, p.129). Second, H&T assume that Earthers and Twin Earthers have corresponding referential intentions when they use moral terms like ‘morally right’ to convey their respective moral views:

. . . by construction of the Twin Earth example, Earthlings and Twin Earthlings have corresponding intentions and interests in using the corresponding moral terms in their

66 Again, what counts as ‘relevantly same nature’ is determined by our general semantic interest to refer to some things and not others when we use words. When we use the word ‘milk’ for instance, we use that word with the referential intention to pick out whatever it is that speakers in our community use that word to refer to, i.e. anything that is suitably similar to the local samples of milk. If we wanted to pick out some other things, like almond milk or coconut milk, then we would pick some other properties as being the “relevantly same” or suitably similar ones that determine the meaning of ‘milk’. The way the relevantly similar properties are chosen is therefore a kind of two-step process: First, decide what interests you have in using a word; second, pick the properties of similarity that are relevant, given those interests.
languages, and the relevant intentions and interests of the Earthlings in the example are the same as our intentions and interests in the actual world in situations in which we use the term “wrong” or its equivalent. (2000, p.129)

Given that H&T build both features above into their description of the Moral Twin Earth scenario, Copp says we can apply Putnam’s semantics and rely on our intuitions about the scenario in light of applying that semantics. Copp’s focus, in line with the cases he mentions above, is on the “corresponding intentions and interests” that the Moral Earther and Moral Twin Earther share. The details of what these corresponding intentions and interests are can be brought out by looking, as Putnam did for ‘water’, at the presuppositions that we use those terms with, e.g. in ostensively defining the sort of actions that are ‘morally right’ and ‘wrong’.

§4.5.4 The Putnamian Approach to Moral Twin Earth

Copp suggests that we start by asking “What is presupposed in “ostensive definitions” of moral terms?” (2000, p.129). Remember that, according to Putnam, ostensive definition depends on certain empirical presuppositions and that when these are false, certain fallback conditions come into play (2000, p.129). When I point to a glass filled with a clear liquid and utter the word ‘water’, I presuppose that the clear liquid in the glass is in fact the stuff that my linguistic community and I intend to use the word ‘water’ to refer to. If my presupposition was false and the glass contains gin rather than water, then as Putnam put it, “I do not intend my definition to be accepted.” Copp argues that similar presuppositions ground the use of moral terms. He brings this out in the following passage:

Suppose, then, that an English speaker on Earth is teaching a child about right and wrong beginning with the lesson that lying is wrong. Suppose he says, “Lying is wrong”. This teaching does not presuppose that the speaker believes lying to be of the same kind or even the same moral kind as most of the acts that the speaker and most speakers in his linguistic community have been calling ‘wrong’. For he might be a non-conformist who disagrees with most speakers in his community about which kinds of acts are wrong. A non-conformist does not have different linguistic or semantic intentions from moral conformists. His moral terms do not have different meanings. And his belief that most people are mostly mistaken about which acts are wrong would not undermine his view that lying is wrong. So he would not have the intention to withdraw his teaching if it turned out that lying is not of the same (moral) kind as most of the acts that most speakers in his linguistic community have been calling ‘wrong’. If he is sincere, then of course he would intend to withdraw his teaching if he came to believe that lying is not wrong. And if he is sincere, he would be committed to withdrawing his teaching if he came to believe that lying is not of the kind, or does not have the property, that he and most speakers in his linguistic community intend to refer to in using “wrong”. This is the kind we call “wrong action”, and the property is the one we call “wrongness”. As Mackie might say, it is the property of “not-to-be-doneness”. (2000, pp.129-130)

According to Copp, the example implies that ostensive definition for moral terms presupposes that there are moral properties or kinds, such as ‘not-to-be-doneness’, that speakers in my community intend to refer to in using their moral terms such as ‘morally right’ and ‘wrong’. If that presupposition is false, e.g. there is no ‘not-to-be-doneness’ property, then Copp says, “we
would not engage in moral discourse or moral teaching” (2000, p.130). Here the problem for H&T comes to the forefront.

The example that Copp describes suggests that our use of moral terms, like our use of any other sort of terms, involves certain presuppositions about our referential intentions and interests. When we use the expression ‘morally right’, we intend to refer to things that have the property that most speakers in our community intend to refer to in using that expression, i.e. the property being morally right or being to-be-done. H&T seem to grant this in the similarities they describe between Earth and Twin Earth. According to Copp, this is the key similarity that enables the proponent of a Putnam-style moral semantics to escape H&T’s conclusion. Given the scenario, it is open for this objector to claim that, the Earther and Twin Earther use ‘morally right’ in their respective languages with the same referential intention, to refer to whatever has the property that speakers in their community intend to refer to in using the expression ‘morally right’. It is because they share this referential intention that their moral terms are moral terms at all. The two use their orthographically identical terms with the intention to refer to the same property, though disagree about what things actually have this property. So, contrary to H&T’s argument, there is room to understand the Earther and Twin Earther as genuinely disagreeing:

. . . in using “wrong”, an English speaker intends to refer to actions that are of the kind, wrong, or that have the property of being wrong or of being not-to-be-done. And speakers of Twin English would have to have the same intentions. If they did not, the relevant terms in their language would not be moral terms. If speakers used a term, “twong”, for example, with the intention merely to refer to the same kind of action that most speakers in the linguistic community have in fact been calling “twong”, then “twong” would not be a moral term. It might be a term of etiquette with a meaning similar to “impolite” or a descriptive classificatory term in the way that “skippin” is a descriptive term for a kind of action. But it would not be a moral term. (2000, p.130)

The key point here is that the semantic account ascribed to the realist by H&T treats ‘wrong’ as being like ‘twong’. As Copp points out, this is a mischaracterization of what the realist should say about the referential intentions governing moral discourse. On a proper Putnamian semantic view, the one the realist ought to adopt, moral terms such as ‘morally right’ and ‘morally wrong’ possess a meaning that is determined by what the speakers of a community intend to use those terms to refer to—and not simply what they in fact call ‘morally right’ and ‘morally wrong’. So, if the realist adopts this kind of view, he or she can explain the kind of mistakes that Putnam and Copp describe and also capture the intuitions elicited from the Moral Twin Earth scenario.

67 Of course, the speakers of a community may on occasion make mistakes in how they use their moral terms.
68 Copp likens his position to a claim that Mackie makes about the intentions that speakers have in using moral terms in his Ethics: Inventing Right and Wrong (1977). See Copp 2002, p.130 n42 and Mackie 1977, p.33.
§4.5.5 The Story So Far

So far, Copp has argued that a Putnamian view can capture our intuitions about Moral Twin Earth by viewing the Earther and Twin Earther as having corresponding referential intentions when they use the moral terms of their respective languages. In fact, according to Copp, the two must have corresponding referential intentions if the terms they use are moral terms at all. This must be the case because moral discourse, like any other kind of discourse (e.g. about natural kinds, about etiquette, etc.), presupposes that its terms are used with certain referential intentions. As Copp has explained, what distinguishes moral terms from other kinds of terms is that moral terms are used with the intention to refer to actions that have certain properties, say, of “being right” or “being to-be-done”. Copp’s point is that this stipulation is built into H&T’S scenario. By construction the scenario, Twin Earth moral discourse is uniformly similar in all regards with Earthen moral discourse, so Earthers and Twin Earthers must have corresponding referential intentions when they use their respective moral terms. This much could potentially be shown by supposing that the contrary is true.

Imagine that the Twin Earther uses the putative moral terms of his language with a sort of deviant referential intention: Suppose that the Twin Earther uses the term ‘morally right’ with the intention to refer to whatever actions are of the kind or have the property of requiring exceptional physical fitness. Given that he has this intention when he uses that term, he applies ‘morally right’ to actions such as climbing Mt. Everest, competing in a triathlon, and so on. On Copp’s view, this would imply that the Twin Earther does not genuinely engage in moral discourse when he uses the term ‘morally right’. Whenever the Twin Earther uses this term, he is not speaking about the subject matter of moral discourse. His so-called “moral terms” are not strictly speaking moral terms at all. Now if this Twin Earther were to exchange “moral views” with the Earther described by Horgan and Timmons, we could not reasonably call their exchange a genuine disagreement. Intuitively, the two could only speak at cross purposes whenever they exchange their sentiments about what things are ‘morally right’ in their respective languages: The Earther uses ‘morally right’ with the intention to refer to actions that have the property of being morally right or being to-be-done whereas the Twin Earther uses ‘morally right’ with the intention to refer to actions that have the property of requiring exceptional physical fitness.

§4.5.6 Shared Interests

According to Copp, we can reasonably expect the Earther and Twin Earther to also share the same interests in using moral terms. Remember that for Putnam, what properties are semantically relevant is a matter determined by our general interests in using a word, i.e. in terms of the role we want that word to play. So given whatever interests we may have in using a particular word, we adopt the referential intentions that will satisfy those interests. Given different

69 I preserve the scare quotes that Copp uses when he mentions these properties.
interests, our words would mean different things. When it comes to the moral terms involved in the Moral Twin Earth scenario, Horgan and Timmons presuppose that the Earther and Twin Earther have the same interests when they use the moral terms of their respective languages. In particular, they use moral terms with the aim of playing a practical role—to evaluate actions, deliberate, and make decisions about what to do:

. . . Twin Earthling judgments about what is “right” and “wrong” play the same role in Twin Earthling reasoning, deliberation, and decision making about what to do that Earthling judgments about what is “right” and “wrong” play in the reasoning, deliberation, and decision making of Earthlings. For example, the Earthers and Twin Earthers are normally disposed to do what they take to be right or twin-right, respectively; they normally take considerations about right or twin-right, respectively, to be especially important in deciding what to do, and so on. The Twin Earthers’ beliefs about twin-right and twin-wrong, like the Earthers’ beliefs about right and wrong, are normative beliefs. Moreover, “wrong” is the primary negative term of act evaluation in the Twin Earthers’ moral code just as the Earthling term “wrong” is the primary negative term of act evaluation in the Earthers’ moral code. (2000, p.131)

According to Copp, these features explain the intuition about why the Earther and Twin Earther genuinely disagree in the lying case:

These facts about the example explain, I think, why we have the intuition that the Earthers and Twin Earthers disagree in the lying case about the truth of the proposition that lying is wrong. Moreover, they point us in the direction of the interest speakers have in using the term “wrong” in English, or the corresponding term in Twin English. Their interest is to pick out the kind of action or property of actions, whatever that might be, that is of primary importance morally in deciding which actions to avoid—the kind that is especially to be avoided, or the property of being morally “to-be-avoided” or “not-to-be-done”. (2000, p.131)

Given that the relevant intentions and interests correspond in the right way, Copp claims that, for any coherent Moral Twin Earth, a Putnamian semantics will imply that English and Twin English moral terms mean the same thing. (2000, p.131) After all, in Horgan and Timmons’ scenario, all of the relevant similarities are in place:

. . . by construction of the example, there is no morally relevant difference between the two worlds. People in the different worlds have different moral theories, of course. But Earthers and Twin Earthers are rational to the same degree and have the same abilities, vulnerabilities, needs and desires, and causal laws are the same on the two planets. Hence, any coherent moral theory would say if wrongness is instantiated on the one planet it is instantiated in the other, and vice versa, and that the kinds of actions that are wrong on the one planet are wrong on the other, and vice versa. But if the worlds are the same in all morally relevant respects, and if corresponding terms in the languages spoken on the two worlds are used with relevantly the same referential intentions, and if people’s interests are the same in using these terms, then the terms would be assigned the same meaning in Putnamian semantics (2000, p.132).

This, Copp says, is enough to refute Horgan and Timmons’ argument without explicating either a complete Putnamian story about moral terms or a complete defence of synthetic moral naturalism (2000, p.133). Once we are clear about the referential intentions governing moral discourse, we can see the Earther and Twin Earther as genuinely disagreeing.
In closing, Copp suggests that his Putnamian view can be extended to Twin Earth scenarios involving non-normative terms. To show this, Copp envisions a kind of “Milk Twin Earth”:

It might be useful here to consider a Twin Earth scenario involving a term that is not normative in order to show that the issues raised for Putnamian semantics by the Moral Twin Earth argument can also arise for non-normative terms. Suppose, then, that Earthlings are acquainted with human milk, cow milk and goat milk and call each of these liquids “milk” in English. They are not acquainted with sheep milk. Twin Earthlings are acquainted with human milk, cow milk, and sheep milk and call each of these liquids “milk” in Twin English. But they are not acquainted with goat milk. Both Earthlings and Twin Earthlings are aware that female dogs that have recently given birth secrete a fluid that is drunk by their young. Earthlings say in English that this is “milk” while Twin Earthlings say in Twin English that it is not “milk”. Twin Earthlings deny that it is “milk” because they believe that all twin-milk is white and they have observed that the relevant secretion of female dogs is yellow. Earth and Twin Earth are otherwise as similar to each other and as similar to the actual world as is compatible with the differences already noted. (2000, p.132)

Following suit with our intuitions about the lying example, Copp supposes that we would be inclined to judge that ‘milk’ means the same thing on Earth and Twin Earth, though Earthers and Twin Earthers disagree about whether the liquid secreted by female dogs shortly after birthing is milk or not. According to Copp, his Putnamian view can capture this intuition.

First, we know by construction of the scenario that Earthers and Twin Earthers have corresponding referential intentions and interests when they use the word ‘milk’. Second, we know that the liquids in question (to which ‘milk’ is applied) share no common chemical composition, though they do share the functional and genetic property of being a liquid (roughly) produced by a female mammal after birthing to nourish its young (2000, p.133). This property is the semantically relevant one, Copp says, because it is the property that is the focus of both the Earther and Twin Earthers’ interests in using the term ‘milk’ in their respective languages—as such, this property is the key one that figures into the corresponding referential intentions when Earthers and Twin Earthers use the term ‘milk’ (in English and Twin English). According to Copp, this implies that the terms in English and Twin English mean the same thing. In addition, because the relevant functional/genetic property is in fact possessed by the liquid secreted by female dogs, the Earther is correct to apply the English term ‘milk’ as he does. On the other hand, the Twin Earther is incorrect to apply the Twin English term ‘milk’ as he does. So despite the fact that the Twin Earther believes that this liquid is not milk, because it is yellow and not white, this fact does not rule out the relevant functional/genetic property as the referent of ‘milk’ (2000, p.133). Again, as Copp notes, any coherent semantics must have room for mistaken beliefs about extension and this is the point that causes problems for Horgan and Timmons in their treatment of Moral Twin Earth.
§4.6 The “Copp/Milk” Objection to Meaning Twin Earth

The argument that Copp uses to undermine Horgan and Timmons’ Moral Twin Earth argument might appear to be available to undermine the Meaning Twin Earth argument outlined in Chapter Two. If a suitably similar argument can be drawn, then there would be a way for an anti-normativist to avoid the conclusion that our intuitions about the Meaning Twin Earth scenario are best explained by the normativity of meaning. By using an analogue of Copp’s objection, one could explain the intuitions about the Meaning Twin Earth scenario without assuming that meaning is normative. In brief, an anti-normativist could claim that ‘means’ is like ‘milk’ and that an analogous explanation can be given along the lines that Copp suggests. I will now move on to consider this line of argument.

§4.6.1 ‘Means’ and ‘Milk’

First, it is important to notice that an anti-normativist cannot simply apply Copp’s main line of objection to reject my normativist argument about Meaning Twin Earth. If ‘means’ is treated like ‘good’, then the anti-normativist has already conceded the conclusion that I want to draw, that ‘means’ is a normative term. After all, in Copp’s discussion of Moral Twin Earth, Copp argues that a proper appreciation of the role played by referential intentions in Putnamian semantics leads to the conclusion that moral terms stand for normative properties. Other sorts of terms, like ‘milk’ and ‘jade’ for instance, refer to non-normative properties. Copp does not formally argue this point because he does not need to, given his purposes. Copp’s goal is simply to demonstrate that a Putnamian semantics can vindicate our intuitions about Moral Twin Earth. This he can do without saying anything specific about the normative status of terms like ‘morally right’ or ‘wrong’. Nonetheless he does assume that these terms are normative in kind. As I have mentioned, the anti-normativist and myself are most interested in this feature with respect to meaning, i.e. in whether the term ‘means’, as it is used to ascribe meaning to a speaker and his utterances, is likewise normative. The main point of my argument in Chapter Two is that it is, given that the best explanation of our intuitions about the Meaning Twin Earth scenario involves the idea that meaning is normative. The anti-normativist wants to reject this, but the onus of proof is on him to show that the contrary is true. I have already made a preliminary case for how a normativist view can readily explain our intuitions about Meaning Twin Earth.\footnote{In Chapter Six I develop the case for this point further.}

One way that the anti-normativist about meaning might try to proceed at this point is by employing Copp’s strategy, which as we’ve seen can also be used to explain Twin Earth scenarios about non-normative terms. I will argue that the anti-normativist about meaning cannot do this without begging the question against his normativist opponent.
§4.6.2 ‘Means’ and Dispositionalism

Given the strategy suggested by Copp, the anti-normativist can construct an analogous argument to the effect that any Twin Earth scenario about ‘means’ can be explained without assuming that ‘means’ is normative. Naturally, the anti-normativist can do this by assuming that meaning ascription or ‘means’ refers to some alternative (non-normative) properties analogous to the functional and genetic properties that Copp uses to explain his ‘Milk’ Twin Earth. The best candidate for that role would be some dispositional properties, in line with a dispositionalist account of meaning.\(^{71}\) This would be the best candidate for the anti-normativist, given that Copp assumes that moral terms stand for normative properties—which is in line with the point I eventually want to make and the point that the anti-normativist does not want to concede, given the trouble it may cause for naturalism or other views. If the anti-normativist takes ‘means’ to stand for some dispositional properties, then he can explain a speaker’s meaning in terms of whether those properties are instantiated or not, i.e. whether the speaker in fact possesses a particular disposition to use an expression in a certain way. Of course, there is reason to be sceptical about whether dispositionalism can succeed in meeting the paradigm objections \(\text{vis-a-vis}\) finitude and circularity, but for now let’s assume these worries can be met.\(^{72}\) I argue that, even if we set aside these problems for dispositionalism, the anti-normativist cannot use dispositionalism to foil my thesis via an application of the strategy that Copp suggests for ‘milk’.

§4.6.3 Meeting the Copp-Milk Objection

In order to meet the Copp-Milk objection, it will be crucial to keep an eye on the ground rules \(\text{vis-a-vis}\) what is permissible for my opponent and I to argue without begging the question against one another. In addition, it will be crucial to keep an eye on my general argumentative strategy. Once these are in clear view, I will argue that the best the anti-normativist can do against my argument is to force a stand-off. This stand-off is not sufficient to meet the challenge of my argument and so at this point the anti-normativist does not have access to a successful objection. Here I first gloss my general argument and then explain the key rules that the anti-normativist and I must abide by with respect to this argument.

§4.6.4 General Strategy

In Chapter One, I discussed two constraints on any adequate account of meaning, following KW’s argument for meaning scepticism in Kripke (1982). On the first constraint, any adequate theory of meaning must give an account of how reference is determined, i.e. what makes

\(^{71}\) Again, Copp assumes that moral terms stand for normative properties. The anti-normativist wants to avoid this and so the best alternative, in line with a naturalist framework, is to say that ‘means’ stands for some dispositional properties.

\(^{72}\) The final line of objection, in Kripke (1982) and popular elsewhere, is that dispositionalism cannot capture the normativity of meaning. This line of objection would, in this context, beg the question against the anti-normativist.
it the case that ‘+’ refers to the *addition* function and not the *quaddition* function, as KW’s sceptic proposes. On the second constraint, any adequate theory of meaning must account for the basic normative character of meaning. The overall aim of my thesis is to provide some evidence for the claim that this second constraint is legitimate and imposes a genuine adequacy constraint on accounts of meaning, something that a theory may fail to meet. My anti-normativist opponent claims that the imposition of this second constraint is in fact not justified.

§4.6.5 Ground Rules

First, I cannot simply assume that dispositionalism fails with respect to the first constraint, i.e. the determination of reference. I cannot do so because this would beg the question against dispositionalism. The strategy I use does not need to this because the ultimate upshot of this dissertation is that we have evidence that the second constraint functions as a further constraint on any adequate theory of meaning. Second, my opponent cannot assume that dispositionalism is a plausible semantic theory, given my challenge with respect to the second constraint. If my opponent does so, he begs the question by either assuming that the second constraint is illegitimate or that dispositionalism can meet the second constraint.

At this point, it is open for the anti-normativist to suggest that if dispositionalist can meet the first constraint, then the Copp-Milk line of argument can be used to block my argument with respect to the second constraint. This would be problematic for me, because it would establish that the second constraint, contrary to what I have argued so far, does not impose a check on accounts of meaning above and beyond the first constraint. Fortunately for my argument, this is not the kind of objection that the anti-normativist can gather from Copp’s original discussion. What the anti-normativist can argue is that ‘means’ is like ‘milk’ and that Copp’s objection can therefore be used against my argument. The problem for the anti-normativist is that this assumes that ‘means’ is like ‘milk’ and this assumption is off-limits. The anti-normativist cannot make this assumption without begging the question against my claim about the second constraint, i.e. that the second constraint imposes an additional constraint on accounts of meaning, including dispositionalism. Thus the anti-normativist cannot block my application of Moral Twin Earth to the meaning case by using the Copp-Milk objection without in this way begging the question. So even if Copp’s argument damages the Moral Twin Earth argument proposed by Horgan and Timmons, this same argument cannot damage my application of that argument to the case of meaning via Meaning Twin Earth. Another way to put this point might be as follows.

The anti-normativist can certainly avail himself of the following conditional:

1. If meaning properties are dispositional properties, then Copp’s strategy yields a way for the intuitions about Meaning Twin Earth to be captured without invoking the normativity of meaning.

However, in order for the anti-normativist to detach and reach the conclusion, i.e.:
(3) Copp’s strategy yields a way for the intuitions about Meaning Twin Earth to be captured without invoking the normativity of meaning. The anti-normativist would need (2):

(2) Meaning properties are dispositional properties.

In the present context, the anti-normativist cannot do this without begging the question. The availability of (2) presupposes either that meaning is not normative, or that meaning is normative in a way that dispositionalism can capture.\(^{73}\)

§4.6.6 Connection with Plunkett and Sundell

In closing, it is worth pointing out the similarity between my response to the Copp-Milk objection and my reply to Plunkett and Sundell. In both cases, there are special features about meaning that leave my argument intact even if the relevant objection succeeds against the Moral Twin Earth argument from Horgan and Timmons.

In considering Plunkett and Sundell’s line of argument, I suggested that metalinguistic negotiation cannot capture the disagreement between the Earther and Twin Earther about the meaning of ‘meaning’ because that dispute depends on having a shared notion of ‘meaning’ (at least provisionally) in order to communicate that disagreement. We cannot make this assumption in attempting to extend metalinguistic negotiation to the case of meaning, so any attempt to make this extension will fail. The Earther and Twin Earther must mean the same thing by ‘means’ even if they disagree about whether ‘means’ ought to denote (roughly) an individualist or communitarian theory of meaning—or indeed any other theory of meaning. So though Plunkett and Sundell’s strategy may succeed in capturing numerous other cases of genuine disagreement, Meaning Twin Earth turns out to be a limiting case for that strategy.

In Copp’s case, the analogous argument inspired from the ‘milk’ scenario cannot succeed, given the aim of my thesis and the ground rules that my opponent and I must abide by. Copp’s strategy cannot succeed if it is applied to the meaning case because it would require the question-begging assumption that there is no constraint in addition to the constraint about the determination of reference—one that an account of meaning, like dispositionalism, might fail to meet. This assumption is question begging because it amounts, as I have explained, to a flat rejection of my argument that Kripke’s second constraint (about the normativity of meaning) constitutes a genuine and independent constraint on any theory of meaning. As such, there is no clear way for the anti-normativist to apply Copp’s proposed line of argument against my thesis without begging the question.

\(^{73}\) It may be objected that dispositionalism can view the normativity of meaning as strictly referring to ‘Norm Relativity’, but this would essentially involve a rejection of ‘Normativity’, i.e. the view that meaning properties are essentially prescriptive or action-guiding in the way that the Meaning Twin Earth scenario suggests. See §1.7-1.7.1.
§4.7 Conclusion

In this chapter I have argued that Copp’s line of objection of the Moral Twin Earth argument has no straightforward application to the use of Meaning Twin Earth as a device for providing evidence that meaning is normative. In the next chapter, I move on to consider another prominent line of objection to Moral Twin Earth that is developed by David Merli in his “Return to Moral Twin Earth” (2002).
Chapter Five: Merli ‘Return to Moral Twin Earth’

§5.1 Introduction

In this chapter I evaluate David Merli’s three objections against Horgan and Timmons’ Moral Twin Earth argument from his paper “Return to Moral Twin Earth” (2002). I outline each of these objections and argue that each fail to debunk the Moral Twin Earth argument and by proxy my Meaning Twin Earth argument.

§5.2 Describing Moral Twin Earth

In his first response to Horgan and Timmons (H&T), Merli claims that the intuitive force of the Moral Twin Earth argument hinges on an ‘underdescription’ of the Moral Twin Earth scenario (MTE) and in particular, the workings of twin-moral practice. Once we explore how the scenario can be suitably clarified, Merli says, it turns out that H&T’s argument is faulty. To show this, Merli revisits the MTE argument and highlights what that argument must do in order to succeed. First, recall that the general aim of H&T’s argument is to demonstrate that causal semantic naturalism (CSN), or what Merli calls ‘naturalistic moral realism’ (NMR) is unlikely to succeed. H&T argue that NMR cannot capture our linguistic intuitions about some cases of moral disagreement. In particular, NMR’s semantic theory cannot explain the intuition that the Earther and Twin Earther of the MTE scenario genuinely disagree in spite of the fact that their moral terms, the English ‘morally right’ and Twin English ‘morally right’, refer to distinct natural (functional) properties. NMR implies that these two parties could not disagree; they would simply speak at cross purposes, given that their terms mean different things. This seems patently false, according to H&T. Therefore, our linguistic intuitions about the MTE scenario count as evidence against NMR: the two parties are best viewed as meaning the same thing and as genuinely disagreeing; NMR suggests otherwise and thus we should abandon NMR given that it generates unintuitive results in moral disagreements such as the one showcased in the MTE scenario. According to Merli, the success of this argument against NMR depends on the satisfaction of two key conditions:

1. First, we have to think that we have a real moral dispute (that is, about what’s right) with the twin-moralists. We have to think that our terms mean what theirs do. (If we don’t, the realist’s purported admission of equivocation seems intuitively plausible.)
2. Second, it has to be true that twin-moral terms really are hooked up in the right way (i.e. by causal regulation) to natural properties that don’t stand in this relation to our moral terms. (If not, then the realist can happily agree with our linguistic intuitions, and there’s no objection.) (2002, p.214)
Merli argues that the MTE scenario cannot be explicated in a way that meets both conditions simultaneously.

The main issue, as Merli sees it, is that H&T fail to describe in sufficient detail what twin-moral practice is like:

The problem is that we don’t know very much about what twin-moral practice looks like. We know that twin-moral terms play roughly the same *formal* role as our moral terms: they evaluate actions, their application is connected to patterns of praise and blame, and so on. (2002, p.214)

In particular, we do not know enough about the *content* of twin-moral practice, Merli continues:

What kinds of acts are labelled ‘wrong’ by our alien evaluators, and why? What kind of justifications are offered in defence of these assessments? What broad theoretical commitments inform their reasoning? (2002, p.215)

According to Merli, the answers here (in terms of the requisite details about twin-moral practice) could unseat our initial intuitions about MTE. Can H&T provide these details in a way that leaves our original intuitions intact?

§5.3 Merli’s Dilemma for H&T

There are two ways, Merli says, that H&T can attempt to spell out the requisite details about the content of twin-moral practice: First, H&T can try to meet the first constraint above by ensuring that we have strong intuitions that twin-moral terms mean the same thing as their Earthen moral counterparts and that, as such, the Earther and Twin Earther are most plausibly viewed as genuinely disagreeing. H&T are able to do this by constructing MTE in a fashion that is exactly like Earth in every way possible, including the stipulation that “twin moralists make the same kinds of judgments with ‘right’ that we do, appeal to the same kinds of justifications in defending those judgments, and so on” (2002, p.215). According to Merli, if H&T opt for this route, then “the twin-moral term ‘right’ couldn’t refer to a different (functional) natural property given that the same relation fixes reference in both places” (2002, p.215). If H&T meet one constraint, they cannot meet the other. This is the case because, generally speaking “[a] term will refer to a property in virtue of some fact about the world, the term, and its use; if Moral Twin Earth is just the same as our own planet, then their terms cannot but refer to the same properties” (2002, p.215). This is problematic for H&T because the realist, the proponent of NMR, can say that H&T’s stipulation about difference in reference between moral and twin-moral terms cannot hold. If MTE is identical with Earth in all of the ways mentioned above, which fix the reference of the relevant terms, according to Merli, then those terms must share the same reference and there is then no bar to the realist claiming that the Earther and Twin Earther genuinely disagree. Consequently, H&T cannot meet the first constraint without conceding to the realist that their claim about reference is mistaken.

The second way that H&T can answer to Merli’s worry about the workings of twin-moral practice is by addressing the second constraint. H&T can do this by showing how twin-moral
practice is substantively different from Earthen moral practice (2002, p.215-6). According to Merli, if H&T explain how the content of twin-moral practice differs from that on Earth, then it will be clear how ‘morally right’ and ‘wrong’ etc. can denote different properties, as H&T stipulate. The trouble with this path for H&T is that, according to Merli, stipulating these differences about twin-moral practice will undermine our collective intuition about shared meaning (2002, p.216). If these differences are substantial enough, then Merli says, we may even be inclined to retract our original intuition about twin-moral semantics. After all, one of the main reasons that we have the intuitions that we do about twin-moral semantics is that we imagine Twin Earthers as using terms like ‘morally right’ and ‘wrong’ in roughly the same way that we do (2002, p.216). If we suppose the contrary is true and that the twin-moralizers use their terms quite differently, then H&T’s assessment would lose its intuitive force:

Suppose that we learn that . . . there are significant differences between our uses of the term and theirs (differences, say, not just in the extensions of the terms but the kinds of reasons they give and accept, and so on). Twin moralists apply ‘right’ to acts we’re sure are abhorrent, they offer completely different kinds of justifications for their claims, they see our reasons as irrelevant, and so on. (2002, p.216)

In this case, according to Merli, it would be fair to say that we’d either retract or at least weaken the intuitions that H&T capitalize on:

It seems to me that then we’d either withdraw our initial judgment that we share terms with the twin-moralists, or at least offer it with less conviction. The content of the twin-moral practice will affect our judgments of shared meaning. (2002, p.216)

If twin-moral practice turned out to be radically different from our own, we may be inclined to judge that twin-moral terms do not mean the same thing as our terms.74 In this respect, the second option for H&T does not fare any better than the first. If H&T build the appropriate details into MTE to make twin-moral practice sufficiently different from moral practice on Earth, then it seems that twin-moral terms could not mean the same thing as Earthen moral terms do. So even though H&T can meet Merli’s second constraint, it seems that this must come at the cost of failing the first constraint, and likewise for the converse claim. If H&T meet the first constraint, then it seems that they cannot meet the second. Merli summarizes this dilemma:

The basic strategy is simple: first, we press Horgan and Timmons for a more detailed account of the workings of Moral Twin Earth, and then argue that either (a) the differences in content between moral and twin-moral practice are significant enough to undermine our conviction in the synonymy of moral and twin-moral terms; or (b) the similarities are sufficient to ensure that, according to the realist’s view of reference, twin-moral terms refer to the same properties as our own moral terms. (2002, p.216)

This is the main dilemma for H&T, and the first of the three objections that Merli levels against the MTE argument. In the following I detail how Merli argues that this dilemma cannot be met

74 This could potentially be shown even if the terms play a roughly similar role. The translation of the twin-moral term ‘morally right’ might be a very bad translation of the English ‘morally right’. I made this same point about the notion of meaning deployed by Copp (2002) in his first reply to H&T. The closest thing to ‘best discus thrower’ in the philosophy department might have never picked up a discus before.
by H&T and how this dilemma figures into a wider argument in defence of realism and especially NMR.

§5.4 Merli’s First Objection

Merli delivers his first objection to H&T in two parts: The first part focuses on a concession that Merli says H&T must make in order for their argument to succeed. This concession, according to Merli, is H&T’s assumption that moral discourse on Earth is causally regulated by one set of natural properties, despite the wide range of moral views that Earthers espouse. The second part of Merli’s objection suggests a holistic response to H&T, inspired by other realist responses to familiar worries about realism’s ability to explain moral disputes.

§5.4.1 Merli’s Target: H&T’s Univocity Concession

The main target of Merli’s first objection is H&T’s concession that one set of natural properties causally regulates the use of moral terms by Earthers, despite any variance across the moral views of individuals on Earth. H&T effectively assume that speakers with quite different kinds of normative commitments are, on each planet, “hooked up in the right sort of way to the same cluster of properties, even though they make different moral judgments, offer different sorts of justifications, and so on” (2002, p.217). This, Merli says, amounts to the concession to the realist that, on Earth for example, “the conversations between Kant and Mill, between Christian and cannibal, and between moral communities of various places or times” can all be captured by the semantic theory proposed by NMR (2002, p.217-8). This concession is a generous one and can help NMR explain moral disputes here on Earth (and potentially elsewhere) in that “whatever its details turn out to be, [NMR’s theory of meaning] can preserve univocity between speakers with diverse views on topics such as the divine right of kings, the permissibility of chattel slavery, the obligations to segregate people of different races, and the proper social role of women—or so our assumption lets us say” (2002, p.218). This is an impressive claim, according to Merli, “precisely because of the range of actual-world moral disagreement” (2002, p.218). Merli uses H&T’s concession to reinforce his general worry about H&T’s ability to meet the two conditions he cites for the success of their argument.

§5.4.2 Radical Differences

Merli argues that when we take H&T’s concession seriously, it shows that the two authors are under immense pressure to specify the details of twin-moral discourse in a way that is radically different from our own moralizing if they are to foil NMR’s semantic view. H&T must do this because otherwise NMR can readily capture the MTE scenario in the same manner as moral dispute here on Earth, given H&T’s concession for the sake of argument (i.e. that NMR’s theory of meaning can preserve univocity and capture disagreement across the diverse moral views of individual speakers). H&T need to describe twin-moral discourse in a way that NMR cannot
capture. The only way for H&T to do this, as mentioned, is to make twin-moralizing radically different from our own practice. As Merli puts it:

. . . there’s pressure on Horgan and Timmons to make the details of twin-moralizing much different from those of our own moral discourse; in an important sense it must be like nothing we’ve ever seen before. If it gave us only the views of a past time-slice of our moral discussions, or views that we see as fringe participants in our own conversations, then the realist could reply, plausibly, that the twin-moralists offer no more of a challenge to her view of meaning than do the Earthen interlocutors we’re assuming don’t cause problems. (2002, p.218)

H&T face similar troubles in meeting Merli’s second condition.

§5.4.3 Meeting the Second Condition

If H&T describe twin-moral practice in some way radically different from that on Earth, it is unclear how that description could also meet Merli’s second condition. That is, it is unclear how these details about MTE could also guarantee that twin-moral terms are causally regulated by some natural properties that do not stand in the same relation to the moral terms used by Earthers. 75 In this case, it looks as though H&T are forced onto the other horn of the dilemma. If twin-moral terms are used in such a radically different way, completely alien to the way their Earthen counterparts are used, it would be implausible to view moral terms and twin-moral terms as being univocal:

Imagine our response on encountering some radically foreign culture whose evaluative practices are quite different from our own. We could understand their terms as meaning ‘right’ and ‘wrong’, or we could think that they have different concepts that play roughly the same role that the concepts of right and wrong play in our discussions. We might think that they evaluate action in terms of honor, or fierceness, or some untranslatable notion that (at least generally and for the most part) carries with it the kind of endorsement that rightness-judgments have in our own case. 76 We can still apply our moral notions to their actions—that is, we can judge that they’re doing wrong—and we might think they have reason to take up our way of doing things, but all of this is compatible with thinking that their evaluative terms don’t mean what ours do. (2002, p.218-9)

So though it would be at least prima facie plausible to view twin-moral terms as meaning the same because of some formal resemblance with moral terms on Earth, this resemblance cannot ground the intuition about univocity given how radically different twin-moral practice must be in order to draw the scenario as one that NMR’s semantic theory cannot explain (2002, p.218). Twin-moral practice simply could not bear such a resemblance if it is different in the way required to meet the first condition.

If Merli is right and the evaluative practices of twin-moralizers in the MTE scenario must be so remarkably different from our own, then we cannot reasonably view our counterparts on

75 As Merli says, twin-moral moral terms must be “hooked up in the right way” to properties that don’t stand in that relation to the terms of moral discourse on Earth.

76 The reader may be inclined to think that Earthers and Twin Earthers mean the same thing by ‘right’ and ‘wrong’ etc. but simply have different standards that regulate their uses of those terms, respective to their speaking communities. This is the gist of H&T’s take on dealing with this sort of radical disagreement and we will return to this later in evaluating Merli’s objection.
MTE as speaking about *rightness* and *wrongness* as we understand those concepts in our own moralizing (2002, p.219). Consequently, we could not plausibly view moral and twin-moral terms as having the same meaning. According to Merli, this is enough to show that H&T’s argument does not pose a serious threat or even a unique threat to realism: “... if the realist can account for all, or at least enough, of moral disagreement here at home, then Moral Twin Earth poses no additional threat. At the very least, it adds to the philosophical debt that must be paid if the argument against NMR is to work” (2002, p.220). This debt is formidable enough that, according to Merli, it is unclear how H&T could meet it because H&T must show how the following three claims are consistent:

A. The realist can account for shared meaning across diverse Earth contexts.

B. Moral Twin Earth looks just (or enough) like Earth.

C. The realist cannot account for shared meanings between Earth and Moral Twin Earth.

(2002, p.220)

The prospects for H&T accomplishing this feat are rather dim, according to Merli, and as such, H&T’s argument does not pose a serious threat to realism or NMR. At minimum, there is nothing new for realist to face in H&T’s argument, i.e. there is no additional challenge that MTE brings about in addition to the familiar worries about realism’s ability to explain ‘garden-variety’ cases of disagreement (2002, p.220). In the following sections, I argue that Merli’s first objection is unconvincing.

§5.4.4 Moral Twin Earth is Still a Problem, Even if Based on Familiar Worries

First, it is important to note that H&T’s Moral Twin Earth argument is a formidable challenge that the realist must meet even if the challenge is based on familiar worries about realism’s ability to capture our intuitions about some cases of disagreement. The fact that H&T’s worries are connected to some older ones does not lessen the impact of those worries against the realist. In fact, H&T are forthright about this aspect of their argument. So the ‘discovery’ that H&T’s challenge is not a thoroughly novel one is not itself a point against that challenge. The realist still needs to address the challenge, even if it is a reconfigured version of familiar anti-realist worries.

---

77 H&T claim that their argumentative strategy is designed to revive certain anti-realist worries expressed in classical arguments made in the past. H&T’s argument is that these old problems can re-emerge for modern realism, though in a different form (via the MTE scenario). If H&T are right and these old problems can still be posed against modern realism, then the realist cannot simply say that these old problems have been dealt with already. The realist needs to show how these old problems, in their reconfigured form, are not decisive against modern realist views like CSN. If H&T can show, for example, that something like a revised version of Moore’s OQA can be raised against CSN, then this is a serious problem for CSN—a project developed at least in part to meet the worries embodied by the OQA. If H&T are successful, then realism has failed to address a problem that it, loosely speaking, thinks it has addressed. For an overview of this general argumentative strategy, see H&T 2013.
Second, the realist should be especially concerned with H&T’s challenge insofar as that challenge is specifically designed to call NMR into question. Hare developed an argument based on similar concerns in the past, as pointed out in Chapter Four, but that argument targeted the analytic versions of naturalism of yesteryear and not the synthetic versions popular today, like NMR. H&T’s argument is better equipped to deal with these synthetic renditions of naturalism insofar as the two authors target the sort of semantic theory that underpin these modern versions of NMR (i.e. based around recent developments in the philosophy of language, as outlined in Chapter Two). The most that Merli admits is that H&T succeed insofar as the general worries they cite do indeed bring realism’s prospects for dealing with disagreement here on Earth into question.

§5.4.5 How Can Realism Meet H&T’s Challenge?

In wondering about how the realist might respond to H&T’s challenge, Merli asks whether H&T actually need to make the generous concession that NMR can account for shared meaning across diverse Earth contexts. After all, H&T may only concede this much to realism for the sake of argument. Moreover, Merli wonders whether reflection about MTE really does show that “even Earthian dispute, as the realist construes it, is equivocal” (2002, p.220). If it does, then realism is under pressure to demonstrate that it can “[preserve] univocity in garden-variety cases of moral disagreement” (2002, p.220-221). Merli argues that the realist’s prospects for achieving this are reasonably good, so long as the realist takes a holistic approach to the issue with an eye on some related problems for realism in moral epistemology. This is the second part of Merli’s first objection. Before addressing this part of the objection, we need to carefully consider whether H&T really need to make the concession that Merli cites. I argue that they do not.

§5.4.6 H&T Do Not Need to Make the Univocity Concession to NMR

Ultimately, H&T do not need to concede that NMR can capture the univocity of moral terms across diverse uses of those terms on Earth. As such, there is a way that H&T can avoid Merli’s objection, though potentially at some cost to the intuitive force of their argument. The key point is this: The ‘Earth’ described by H&T is very much like actual Earth, but there is a key difference—the Earthers that inhabit the Earth described by H&T uniformly subscribe to a consequentialist ethical theory. This much is not obvious about actual Earth, though we can imagine (roughly) what Earth would be like if this was true and this is enough for H&T’s purposes; all they need to do is provide a thought experiment that generates intuitions at odds with NMR.

Merli notes that H&T apparently view their argument as delivering a kind of “knockout punch” to CSN, partly because they concede realism can capture the univocity of moral terms on Earth despite the various moral views that individuals espouse. Merli takes H&T to suggest, by making this concession, that they believe their argument will succeed regardless of whatever happens with respect to actual disagreement here on Earth (2002, p.217, fn11). Moreover, Merli takes the concession to imply that H&T view themselves as advancing a new challenge to realism. As we have seen in §5.4.4, this is strictly-speaking not true.

78 Merli notes that H&T apparently view their argument as delivering a kind of “knockout punch” to CSN, partly because they concede realism can capture the univocity of moral terms on Earth despite the various moral views that individuals espouse. Merli takes H&T to suggest, by making this concession, that they believe their argument will succeed regardless of whatever happens with respect to actual disagreement here on Earth (2002, p.217, fn11). Moreover, Merli takes the concession to imply that H&T view themselves as advancing a new challenge to realism. As we have seen in §5.4.4, this is strictly-speaking not true.
According to H&T, this assumption about Earth and the similar assumption about Twin Earth yields intuitions at odds with those about Putnam’s classical scenario, as we saw in Chapter Two. If this is not enough to generate those intuitions, then H&T may be able to revise the scenario in some way that does yield those intuitions.

One of the main reasons that H&T’s argument is so compelling is because it demonstrates that even if NMR can explain univocity on Earth, it cannot explain univocity in the MTE scenario. So though H&T ultimately want to show that NMR is in trouble in that it is unlikely to be able to capture actual disagreement on Earth, the MTE scenario is a powerful way to show this by making the major concession to realism that it does. As I have pointed out, H&T do not need to make the concession that they do, though this concession highlights the fact that does not successfully avoid the kind of objection that Moore raised against its analytic predecessors. That said, there may be some other ways to weaken the force of the concession without also weakening the strength of the argument. The easiest way for H&T to do this would be to revise their scenario in a way that makes less of a concession to the realist, e.g. by imagining less variance in the moral judgments of Earthers. Before we do this much, it is worth emphasizing again that H&T do not take their scenario to be about the actual planet Earth but instead some idealized (albeit very close) version of it. In this respect, the concession that Merli cites about actual disagreement on Earth may not actually reflect the concession that H&T make in the scenario. So given that H&T’s ‘Earth’ isn’t or doesn’t have to be the actual Earth, the concession that Merli mentions simply drops out as irrelevant.

§5.4.7 Disagreement on ‘Earth’ and Revising the MTE Scenario

One avenue of escape for H&T would be to simply drop the sort of variance that they imagine occurs on Earth and Twin Earth. H&T could suppose for instance that Earthers use moral terms in a perfectly uniform way given that they, let’s say, have arrived by consensus at some common moral theory like the consequentialist theory T\(^c\). The difference with the original case is simply that all Earthers have perfect agreement about what moral terms mean and how they are to be applied to various actions, persons, etc. In short, Earthen moral discourse is perfectly regulated by T\(^c\) and as it turns out, there is no deviance from what T\(^c\) prescribes.\(^{79}\) This would weaken the force of the concession to realism in that it would not amount to much to admit that

\(^{79}\) The possibilities are endless here. For instance, H&T might imagine that there is a ‘Planet Singer’, a planet where all of its inhabitants perfectly subscribe to everything that Peter Singer has to say about moral matters, and ‘Planet Korsgaard’ and so on. This would get us more toward the kind of thought experiments that have been popular in the discussions about Putnam’s classical case, i.e. about ‘Mars’ and ‘Twin Mars’ and so on. A good survey of the key papers in this area is Pessin and Goldberg *The Twin Earth Chronicles* (1996). Introducing these more complicated versions may make things more difficult at the level of getting clear intuitions from ordinary folks, but they may nonetheless be able to fend off realist objections like the one Merli proposes here.
realism can capture the univocity of moral terms across a discourse that is perfectly uniform. Now if we assume that there is a Twin Earth with exactly the same kind of phenomena, except for the fact that Twin Earthers use of ‘morally right’ is causally regulated by a property whose functional essence is captured by the deontological theory T, then the realist is still in trouble. In this case, H&T have not conceded that realism can explain univocity on actual Earth, where people accept various moral views, though they do concede that realism can explain univocity on this imagined Earth, where Earthers uniformly accept exactly one moral view. The force of this argument is not as strong as the original one proposed by H&T, though it can still generate the same problem for NMR in that NMR’s semantic theory is not equipped to explain how an Earther and Twin Earther could mean the same thing and disagree about what is ‘morally right’ in spite of the different properties that the English ‘morally right’ and the Twin English ‘morally right’ refer to.

Though the above argument may still work against NMR, this comes at some penalty to the strength of the argument, i.e. by conceding less to NMR and by stretching the imagination with the assumption that Earthers and Twin Earthers have perfect consensus about moral matters on each of their planets. As I have suggested, H&T can re-imagine the scenario to be about two planets that are substantially different from Earth, but this comes at the cost that our intuitions may not be as clear as they are in the original scenario.

The reason that Putnam’s classical Twin Earth scenario is so powerful is because it imagines Twin Earth as part of our actual world, albeit a distant part of it that we are not yet familiar with. The intuitions about that case are compelling largely because Putnam relies on familiar data about Earth. The same can be said of H&T’s Moral Twin Earth. The reason that H&T’s scenario is powerful is because we have basic convictions about how moral disagreement works here on Earth. If H&T stray too far from Earth, it becomes less clear how those convictions determine the assessment of the scenario. Nonetheless, there may be some simple ways that H&T can reimagine the scenario without stretching our intuitions too much, as I have suggested above, though it is unclear whether this is really necessary. We have already seen that Merli simply misinterprets H&T’s original scenario as being about actual Earth. The result of the scenario certainly causes some problems for realism, but this result does not necessarily hinge, as we have seen, on a scenario that is about Earth as we know it. With some slight alterations—or maybe just clarifications—the scenario can still work against NMR without conceding anything about actual disagreement here on Earth.

---

80 Granted, there is a way to view this scenario as weakening the challenge against naturalistic realism in that it reconstructs an actual moral dispute—something that isn’t a new challenge for the realist. However, merely pointing this out does not constitute an answer to the challenge.

81 There may still be a better formulation out there that fares better than the scenario I have quickly sketched here. As I have noted, there are infinite ways for making minor alterations to the scenario and so there may be some scenario that can work against CSN at no (at least obvious) penalty to the force of our intuitions.
In the second part of the objection, Merli argues that realism can meet H&T’s challenge even if the two authors do not concede that realism can explain the univocity of moral terms on Earth. According to him, the realist can meet H&T’s challenge without this concession because realism’s prospects for preserving univocity are, by its own lights, reasonably good, given how realism can address some related problems in moral epistemology.

§5.4.8 A Holistic Strategy for Preserving Univocity

Merli argues that any realist answer to H&T’s challenge needs to target the general worries behind that challenge (i.e., worries about the ability of NMR’s semantic theory to explain moral disagreement). The realist can address these worries, Merli says, by taking a more holistic approach with an eye on three problems that embody these worries. The first he calls the ‘univocity problem’: “how can a naturalistic account of our moral discourse earn the right to say that our terms mean the same thing in the mouths of speakers with such serious differences?” i.e. between two speakers who accept radically different moral views, like consequentialism and deontology, and use ‘morally right’ to apply to widely different things (2002, p. 221). The second problem Merli calls the ‘practical role problem’: “how can the [realist] answer to the univocity problem—in terms of a view of reference, for example—allow for the distinct practical and evaluative role of moral discourse?” (2002, p. 221). The third and final problem Merli calls the ‘moral inquiry problem’: “the problem of showing how an account of the univocity of moral terms can preserve and legitimize (at least parts of) the procedures we use in our moral inquiry” (2002, p. 221). These three issues are closely connected, Merli says, in that “[e]ven if NMR can give an answer to the univocity problem, it hasn’t really made much progress until it can show how its answer is compatible with preserving the practical role of moral discourse and giving us a satisfying picture of moral investigation” (2002, p. 221). The best way that the realist can address the three, Merli claims, is by addressing the “underlying unease that drives them,” by showing how a suitable realist theory of reference can preserve basic facets of our moralizing (2002, p. 221). In effect, Merli is advocating a kind of holistic approach for the realist, whereby the semantic, psychological, and epistemological questions above must be answered simultaneously by one kind of package theory.

§5.4.9 Picking a Suitable Theory of Reference

In order to highlight the importance of treating the above three problems simultaneously, Merli imagines how a simple theory of reference could fail to generally address the underlying worry embodied by them. For example:

If the reference of moral terms were fixed by some kind of initial baptism, then we wouldn’t be able to entertain certain seemingly important questions—for example, were the ancestors wrong to apply the term as they did initially? A view like this seems to undermine our means of ethical inquiry, since the answers to moral questions seem to be decided by causal-historical facts that appear to be irrelevant. (2002, p. 221)

Moreover, a similar issue crops up with respect to the practical role problem:
Suppose that we learn that our terms stand in the right sort of relationship with a single cluster of natural properties. This discovery, by itself, doesn’t do anything to vindicate the evaluative or practical role of these terms. In fact, that role is undermined by the thought that our moral terms are held hostage to historical contingencies that seem not to matter in our moral deliberations. (2002, p.221)

The realist needs to pick a semantic theory that addresses these three issues at the same time and which yields the right answers, as dictated by the intuitive facets of our moralizing: “whatever approach we take to the question of reference must not settle the wrong questions, or dictate the wrong means of investigating these questions” (2002, p.222). Moreover, the realist must pick a theory that allows “room for an account of why these properties (the ones picked out by moral terms, according to the account of reference under scrutiny) are interesting and relevant to our practical decision-making” (2002, p.222). In particular, Merli says, “it can’t preclude adequate answers to the practical role problem or the moral inquiry problem” (2002, p.222). Any adequate realist semantic theory must, in addition to addressing the univocity problem, keep a closer eye on the “practical and evaluative role of moral discourse” and “the procedures we use in moral inquiry” in order to address the general worry behind H&T’s challenge (2002, p.221). Merli believes that the prospects for such a theory are good, but what would such a theory look like?

§5.4.10 Convergence and an End-of-the-Day Moral Theory

The realist can begin to construct the kind of package theory that Merli suggests by aiming at issues about reference in connection with questions about how we deliberate about moral matters and about what our moral theory would look like at the end of inquiry, i.e. in terms of an idealized end-of-the-day moral theory. According to Merli, this connection is important because it could helpfully explain “how natural properties could play certain action-guiding roles, and because it helps to preserve the relevance of our procedures of moral inquiry” (2002, p.223). This would in turn help in addressing the univocity problem because, according to Merli, the univocity of moral terms can be grounded by the findings of these continued procedures:

If the correct account of our moral properties—the referents of our moral terms—were given by the end-of-the-day moral theory, then the issue of shared meaning comes down to, or at least essentially involves, the question of convergence. On such a view, we share meanings in virtue of the kinds of shared canons of evidence and argument, and deference to future theory, that would, under slightly idealized conditions, result in a convergence of our moral opinions. (2002, p.223)

So ultimately, according to Merli, the univocity of moral terms can be explained by the realist in terms of the properties our end-of-the-day moral theory would arrive at after undergoing prolonged and serious moral inquiry “under slightly idealized conditions” (2002, p.223). Let’s unpack this a bit.

First, Merli is essentially saying that when we use moral terms, we are actually referring to whatever properties those terms would refer to at the end of moral inquiry (in line with whatever

82 My emphasis.
theory we would have at the end of inquiry). How does this help with the univocity problem and stave off worries about realism’s ability to explain moral disputes? Let’s take an example: Suppose that Peter Singer and Christine Korsgaard currently disagree about why certain sorts of actions are ‘morally right’, given the different ethical theories they accept. On Merli’s suggested view, the realist can view Singer and Korsgaard as meaning the same thing by ‘morally right’ insofar as the two would ultimately converge (along with everyone else) on a common moral theory at the end of inquiry. According to Merli, the realist can be reasonably optimistic about the prospects for this convergence and consequently, he or she can legitimately treat ‘morally right’ as used by Singer and Korsgaard (or indeed anyone else) to mean the same thing, despite the radically different moral views they currently espouse.\(^83\)

Second, the sort of serious moral inquiry that Merli has in mind is “our prolonged attempts at reasoned inquiry into pressing questions” about what we ought to do, e.g. how we ought to treat other human beings and animals, whether it is ever permissible to tell a lie that would increase overall human happiness, etc. (2002, p.222). According to Merli, the realist can treat moral inquiry as a process in which we eventually arrive, in time, at some properties that ultimately ground our collective judgments about the sort of actions that are morally right and wrong. So in retrospect, Merli says, these properties are the ones that “we’ve been referring to all along” (2002, p.222). This strategy may be familiar to the reader in that this is what Boyd ultimately opts for in his semantic theory: “[i]t’s not by accident that Boyd proposes a version of the causal theory of reference that invokes what we end up saying rather than some initial baptism” (2002, p.222). The advantage of this approach is that it can readily address the practical role question, as suggested above, by leaving “room for an account of why these properties (the ones picked out by moral terms, according to the account of reference under scrutiny) are interesting and relevant to our practical decision-making” (2002, p.222). A semantic theory along these lines can capture the practical role of moralizing “by taking our deliberations about them to be part of the process that reveals . . . the reference of moral terms” (2002, p.222). As Merli notes, there is no guarantee that the end-of-the-day properties (identified with rightness etc.) will always be relevant to our deliberations or will always motivate agents to do what they judge to be right, etc., but in any case, “it makes it likely that these [moral] judgements will continue to play their distinct practical role in our deliberations” (2002, p.222-3).

§5.4.11 Two Advantages of Merli’s Strategy

There are two further advantages to the above strategy, according to Merli. The first is that it “fits neatly into common and appealing ways of thinking about shared meaning in cases of nonmoral terms” (2002, p.223). For example, for these nonmoral cases “we think that speakers share terms across changes in theories, or despite nonstandard beliefs involving the term. The

---

\(^83\) I expand on Merli’s reasons for this optimism in §5.16.
main reason for this, according to Merli, is that these cases involve “the idea of deferring to the right authorities, or to the right kinds of reasons and evidence” (2002, p.223). Suppose for example that someone—let’s say Frank—does not know that water is essentially composed of the chemical compound H\textsubscript{2}O and yet is otherwise familiar with water as the stuff that fills lakes and rivers, etc.. Now imagine that an atomic scientist, Maria, has an exchange with Frank about whether the liquid in a glass in front of Frank is water or not. Maria and Frank each use the term ‘water’ in discussing whether it is water in the glass or not, though Maria is very familiar with the fact that ‘water’ is in fact constituted by the compound H\textsubscript{2}O. Maria and Frank mean the same thing by ‘water’ in that they both are referring to the same thing, the liquid composed of the compound H\textsubscript{2}O. In this respect, they may perfectly well disagree about whether the liquid is ‘water’ or not without talking past one another. After some time, Frank is likely to defer to Maria’s judgment, say, by learning that she is an atomic scientist. He is likely to do this because Maria is the right kind of authority—she knows what constitutes water and how to go about testing whether the liquid is in fact water or not. According to Merli, things go similarly for moral terms and for moral disputes. We use terms like ‘morally right’ across various theories and nonstandard beliefs in our moral thinking. Nonetheless we mean the same thing when we use ‘morally right’ and other moral terminology in that we use those terms to ultimately refer to what, at the end of the day, our moral theory would consist in. We can explain the univocity of moral terms by appealing to moral convergence because our end-of-the-day moral theory plays the role of informing our decision-making in moral matters. Consequently, Merli says, there is a “continuity between the moral and nonmoral cases,” which fits nicely with NMR’s view that moral terms and nonmoral terms are far less different than H&T and other anti-realists suggest (2002, p.223).

The second advantage of Merli’s proposed strategy is that it “takes some of the intuitive sting out of judgements of equivocality” (2002, p.223). Merli argues that this is the case because on his suggested tack, “disputes that looked as though they were genuine turn out to be merely verbal only when it turns out that the parties involved would not, under suitably idealized circumstances, move toward agreement” (2002, p.223). According to Merli, this is exactly the kind of outcome that should call into question our initial assessments about some disputes that may have appeared as univocal and genuine. Let’s look back toward the dispute imagine by H&T in their MTE scenario. In H&T’s case, we may be willing to drop our initial conviction that the Earther and Twin Earther mean the same thing by ‘morally right’ and so genuinely disagree if, for instance, it turned out that the two parties would not move toward agreement in their moral thinking. So using this strategy, the realist may be able to neutralize H&T’s argument by showing how the assumed univocity between the Earther and Twin Earther is on shaky ground. \footnote{This issue is centre stage in Merli’s second objection and we will focus on it later in evaluating that objection.} Alternatively, if the Earther and Twin Earther are likely to converge, then the realist can treat the
two as meaning the same thing by ‘morally right’, despite their differing current views about the things that are ‘morally right’, because eventually they will converge on what ‘morally right’ applies to. Given that they will converge on this theory at the end of inquiry, their uses of ‘morally right’ in each of their respective languages actually refer to the properties settled on by that end-of-the-day idealized theory.

§5.4.12 Why Merli’s Strategy is the Best Defence for NMR

Merli’s suggested route here is incomplete in that it is only, as he says, a gesture toward the best kind of semantic view for the realist. Nonetheless, it is the most promising direction the realist can go in:

NMR’s most promising strategy is to invoke the limit of moral inquiry as a resource to use in answering the problems we’ve been discussing. This offers us a way of preserving the univocity of moral terms across disputes while retaining recognizable forms of moral inquiry and making room for the practical role of moral judgment. If this is the right approach to the univocity problem, then the realist’s hope for account for shared moral terms rests with the convergence of moral inquiry under suitably idealized conditions. (2002, p.223)

If the realist focuses on the prospects of an end-of-the-day idealized moral theory, Merli says, things should go rather well. Merli bases this optimism on two premises. First, he says moral inquiry is relatively young, and second, there are some examples of moral progress. Of course, Merli qualifies, we cannot reasonably expect perfect convergence or consensus across moralizers:

Just as the views of flat-earthers or creation scientists don’t count against the idea of scientific progress, the obdurate moral commitment of, say, white supremacists may not be evidence against the prospects of moral progress. In addition, if we had independent reason to think that moral properties had causal powers, then we might think that convergence is made more likely by the influence these have on our inquiries. (2002, p.224)

Naturally, some philosophers are rather pessimistic about our prospects for moral convergence and settlement on an end-of-the-day moral theory. After all, some topics of moral dispute are deeply entrenched. Merli concedes this but claims that, in time, moral scrutiny will push us toward convergence “because of the appeal of some well-worked-out moral views[,]” even if that process takes a very long time (2002, p.224-5).

With all of the preceding bits in place, the reader should have at least a rough idea of how Merli’s rather broad first objection goes. In order to get crystal clear on how this objection goes, let’s step back a moment and gloss what Merli has argued and where that leaves H&T’s argument against NMR. Merli’s main charge in the first objection is that H&T fail to describe in adequate detail the workings of twin-moral discourse in their Moral Twin Earth scenario. According to Merli, the argument collapses if we remedy this issue. When we fill in the details, we can preserve univocity only at the expense of having the Moral Earther and Moral Twin Earther refer to the same properties by their moral terms, while if we make these properties radically different we lose the intuition that those terms are univocal between the Moral Earther and Moral Twin Earther.
Merli advances this argument further by suggesting that tying reference to idealized moral inquiry can give the realist a credible way of capturing genuine moral disagreement. According to Merli, NMR has good prospects for expecting convergence on an end-of-the-day idealized moral theory. So though any exchange between Earthers and Twin Earthers would be rather taxing at first, we could expect the two populations to converge toward a common end-of-the-day moral theory, even if that theory looks unlikely and distant today. In fact, as Merli has noted, this should sound quite familiar given the state of moral matters here on Earth, where we are split between a few deeply entrenched ethical views, each backed by (roughly) equally plausible and strong intuitions about why certain kinds of actions are morally right to perform. So, at minimum, there does not seem to be, as Merli argues, any new problem for realism presented in H&T’s scenario. The way the realist can face H&T’s challenge is, as suggested here, by viewing the Earther and Twin Earther as using terms that are univocal in virtue of the fact that the two parties are apt to, in time, converge in their moral theories so that their use of moral terminology is actually regulated by one set of properties, specified in whatever shared moral theory they would both assent to at the end of inquiry.

§5.4.13 Responding to the Second Part of Merli’s Objection

I will now argue that the second part of Merli’s objection cannot succeed in that it implies the highly questionable view that some radical disagreements are cases in which the parties involved will converge in their moral views. This is highly questionable in that there are no good independent reasons for viewing the speakers involved in such disputes as likely to converge. In fact, H&T anticipate and effectively dispel this strategy in their 1996 paper, “From Moral Realism to Moral Relativism in One Easy Step”.

In their 1996, H&T argue that moral realism is untenable because it is, according to them, committed to moral relativism and so cannot explain moral disagreement. One of the examples that H&T cite in support of their case is a dispute described by Putnam in his 1981. In the case that Putnam describes, Putnam and his colleague Robert Nozick have a political disagreement about whether welfare spending is morally permissible or not:

One of my colleagues is a well-known advocate of the view that all government spending on ‘welfare’ is morally impermissible. On his view, even the public school system is

---

85 At the least, Merli believes there are no good arguments against convergence. See Merli 2002, p.224-225.
87 The reader may suspect that Merli depends too heavily on this idealization. Merli attempts to forestall this worry by qualifying that “we need make our inquirers only idealized enough to consider fully the reasons for competing moral views, actual and imaginable” and so there is nothing far-fetched in his idealization, or at least nothing more far-fetched than the scenario itself, which imagines that the relevant inquirers have considered other moral views, like the twin-moralists’ theory Td, and rejected or accepted them (2002, p.228 fn21).

Merli also ignores any problems that come with the notion of a “limit of inquiry” or “theoretical end of the day,” though he is confident that there is a defensible account that could fend off these problems, at least for the work he does here (2002, p.228 fn21).
morally wrong. If the public school system were abolished, along with the compulsory education law (which, I believe, he also regards as an impermissible government interference with individual liberty), then the poorer families could not afford to send their children to school and would opt for letting the children grow up illiterate; but this, on his view, is a problem to be solved by private charity. If people would not be charitable enough to prevent mass illiteracy (or mass starvation of old people, etc.) that is very bad, but it does not legitimize government action. In my view, his fundamental premises—the absoluteness of the right to property, for example—are counter-intuitive and not supported by sufficient argument. On his view I am in the grip of a ‘paternalistic’ philosophy which he regards as insensitive to individual rights. This is an extreme disagreement, and it is a disagreement in 'political philosophy' rather than merely a ‘political disagreement’. (1996, p.25)

In H&T’s assessment of this dispute, the two argue that Putnam and Nozick cannot reasonably be viewed as likely to converge in their views about the moral permissibility of spending welfare—or indeed about the permissibility of many other political practices. The two philosophers simply accept radically different standards for what counts as ‘morally permissible’ and even under idealized conditions of convergence more generally, there is no reason to expect the two to converge. So even though Putnam and Nozick are clearly competent peers with all of the relevant knowledge and though they each aim at having maximally coherent moral and political views, we still could not expect them to converge in the way that, for instance, Merli suggests that Earthers and Twin Earthers might in the MTE scenario. According to H&T, this sort of move would not only be desperate but also ad hoc. This is their assessment of the scenario:

Here we have two philosophers who have apparently thought through the implications of their own moral outlooks, have gotten clear about relevant (available) factual information, aiming at unity and coherence among their respective desires and attitudes, and are at loggerheads over the issue of welfare spending. Of course it would be hard to show decisively that Putnam and Nozick have both reached a state of reflective equilibrium; perhaps they don't have all relevant information (if one includes 'information' that is not available for whatever reason). Nevertheless, when one reflects on this sort of case, involving a very high level of philosophical sophistication and moral sensibility, it just looks very unlikely that any movement in the direction of yet greater unity and coherence will bring the two together on this issue. Although one could always insist that one or both parties have not yet achieved equilibrium and that they would converge if they ever were to reach, this looks like a particularly desperate thing to say here. Rather, it seems much more plausible to attribute their differences to significantly different moral standards—so different that, insofar as each philosopher's standards are systematizable, Putnam's standards conform to a different normative moral theory than do Nozick’s (as in the imaginary human/Martian scenario). This diagnosis is basically the one Putnam himself gives of the situation. (Horgan and Timmons 1996, p.25-26)

As H&T propose, the same can be said about the dispute described in the MTE scenario.

Given that the Earther and Twin Earther have such radically different views about what sorts of things are ‘morally right’ and ‘wrong’, NMR’s semantic view cannot reasonably use convergence as a basis for viewing the two as meaning the same thing by those terms. We simply do not have any good independent reasons—other than preserving NMR—for viewing the two parties as converging on some common theory at the end of inquiry. So what this discussion

---

88 My emphasis.
reveals is that, ultimately, our linguistic intuitions about the MTE scenario cannot be reasonably explained by a realist theory of reference insofar as that theory requires convergence in order to explain the univocity of moral terms. As such, this second part of Merli’s first objection cannot succeed. So on both fronts, Merli does not have a successful objection against H&T’s challenge. Let’s see if the second objection fares any better.

§5.5 Merli’s Second Objection: “Revising Our Intuitions”

In Merli’s second objection, Merli argues that the realist can meet H&T’s challenge by demonstrating that our initial conviction about the univocity of twin-moral and moral terms should be retracted. This second objection is a departure from the first insofar as it explains how this conviction about univocity is misleading and that ultimately, Earthers and Twin Earthers should be viewed as having a merely verbal dispute, i.e. as simply speaking at cross purposes. All of this, Merli says, can be done on purely realist grounds by focusing on how the Earther and Twin Earther arrive at their respective moral theories. Merli motivates this second objection in two steps: The first step of the objection introduces a slightly modified MTE scenario that imagines how things would pan out between Earthers and Twin Earthers if moral inquiry on their planets was continued into the distant future. The key issue here is whether a set of idealized and better-informed Earthers and Twin Earthers would view their moral terms as univocal or not: In the second step, Merli says that the two would be inclined to view their moral terms as not being univocal. Merli argues that we should defer to our judgements to these idealized and better-informed individuals and so, given their view in this extended scenario, we should go back and re-evaluate our own intuitions about H&T’s scenario. According to Merli, doing so will reveal that these initial convictions about H&T’s scenario are faulty. In the next few sections, I spell out Merli’s second objection in detail and then argue that this objection cannot succeed because Merli is unreasonably optimistic about the kind of sway that some idealized thinkers could have on us.

§5.5.1 Merli’s Second Objection: Step One

In the first step of the objection, Merli extrapolates how things would go in H&T’s scenario if moral inquiry on Earth and Twin Earth was continued in earnest. In short, Merli says, imagine that equally rational Earthers and Twin Earthers continue to “make efforts to engage in discussion with [their] peers . . . discard moral commitments at odds with more fundamental and better-supported views, and so on,” and likewise, “our [Twin Earthen] counterparts do the same” (2002, p.226). Naturally, the MTE argument can only work if Earthers and Twin Earthers come to different verdicts in their moral inquiries, otherwise the two could not be readily viewed as genuinely disagreeing about the sorts of actions that count as ‘morally right’ and ‘wrong’, etc. This is guaranteed by the scenario in that Earthers and Twin Earthers must end up saying
something to the effect that their respective moral theories, \( T^c \) and \( T^d \), are the best available accounts of morality\(^89\):

\[\ldots \text{it must be true that twin moralists end up saying something like ‘well, it seems that deontological theory } T^d \text{ is true; this is the best account of morality’ while the moralists here on Earth say the same thing of their consequentialist theory.} (2002, p.226)\]

This much is assumed by H&T’s original scenario. Merli pushes this a bit further and imagines that, down the line, the results of moral inquiry on Earth and Twin Earth end up sharing the same status as scientific discoveries like the discovery that the watery substance on Earth is in fact composed of the molecular compound \( \text{H}_2\text{O} \).\(^90\) This would be how things should naturally go on Earth and Twin Earth if moralizing is carried out in the way described, Merli says:

\[
\text{Earth and Moral Twin Earth [would be] alike in having a maximally stable, well-justified, and widely accepted moral (or twin-moral) theory. This, of course makes them different from the Earth of today, since, here and now, we’re stuck with a discourse that seems torn between several incompatible theories with roughly equal backing from firm intuitions and plausible justifications.} (2002, p.226)
\]

Now suppose that Earthers and Twin Earthers were to encounter each other for the first time and come to exchange their findings.

In this case, Earthers and Twin Earthers would likely be rather surprised to discover that they have arrived at radically different results in their moral thinking. In particular, Earthers would be shocked to learn about the sorts of things that twin-moralizers classify as ‘morally right’.\(^91\) For example, Earthers might be shocked to learn that, let’s imagine, “twin-moralists [would] favour a policy of executing murderers, a practice that they think should be condemned” (2002, p.227). Naturally, Earthers would be rather uncomfortable with these sorts of judgments—judgments that they would find objectionable given the findings of their own prolonged and serious moral inquiry.

So in discovering this about their new Twin Earthen peers, Earthers might be inclined to offer objections based on those findings. Earthers might, for example, “raise objections based on the minimal (or non-existent) deterrent value of executions, the economic cost of such a policy, its contribution to scepticism about the justice system, and so on—that is, they appeal to the ‘forward-looking’ reasons for being suspicious of the death penalty” (2002, p.227). Unfortunately for Earthers, their efforts at convincing Twin Earthers would likely be poorly received, given that equally well-grounded findings regulate Twin Earthen moral thinking and that Twin Earthers and Earthers are equally rational beings. After all, Twin Earthers end up being deontologists in the end of their moral inquiry (in line with their theory \( T^d \)), so it would be very unlikely for Earthers to convince these Twin Earthers by giving the kind of reasons that Earthers are inclined to give,

---

\(^89\) Each party believes in earnest that their current working moral theory is the best available one, i.e. as far as their serious moral inquiry is concerned so far.

\(^90\) As Merli puts it: “\ldots imagine that the results of these [moral] inquiries [on Earth and Twin Earth] are widely accepted; these claims, and some of the theoretical apparatus surrounding them, become part of our standard lore, just as claims like ‘water is \( \text{H}_2\text{O} \)’ have” (2002, p.226).

\(^91\) Twin Earthers would have the same kind of reaction about the sorts of actions that their new Earthen peers apply their term ‘morally right’ to.
in line with their consequentialist theory T. These kinds of reasons will simply not convince the Twin Earthers. Twin Earthers might say something like, “We know all that . . . but what has that got to do with whether it’s right or not?” (2002, p.227). What can these two parties do in order to break the state of being at loggerheads they have found themselves in? Merli imagines how the conversation could go from here:

It’s plausible to imagine that moralists and twin-moralists would attempt to engage in serious inquiry to resolve their differences; after all, each group is faced with an equally insightful, equally conscientious body of thinkers who have reached different results. In the spirit of epistemic humility, our interlocutors should be inspired to re-examine their old views and to consider the positions and arguments offered by their new colleagues. Suppose that moralists and twin-moralists decided to sit down together to hash out their (putative) disagreements. Such an inquiry might result in changes in view on the part of one or both parties, or it might leave them both completely unmoved. (2002, p.227)

Let’s look at both possibilities in detail.

§5.5.2 What if Earthers and Twin Earthers Change their Views?

Suppose that, after some heated discussion, Earthers and Twin Earthers decide to change their moral views by adopting one or the other’s view, or else by adopting some synthesis of the two (2002, p.227). In any case, there would be some change at the level of the moral theories that Earthers and Twin Earthers accept, and consequently, there would be some corresponding changes in the way moral semantics works in those groups. For instance, Earthers might subtly change the way they apply ‘morally wrong’, by applying it to certain extreme or egregious examples of individuals being sacrificed for the good of others. Likewise, Twin Earthers might subtly change the way they apply ‘morally right’, by applying it to some extreme cases where a great number of people would massively benefit at the expense of one individual. So going forward, the two populations might end up with some kind of hybrid view, influenced by the moral inquiries of each. Merli argues that we could not reasonably believe that Earthers and Twin Earthers would undergo this kind of “conversion experience” (2002, p.227).

According to Merli, it would be more plausible to believe that, if Earthers and Twin Earthers end up changing their views, either one group or both groups did not ultimately arrive at a fully-coherent theory at the end of moral inquiry. This would be a more plausible view, Merli says, in that if either group ended up changing their view, then they must have been less than coherent or fully justified, i.e. “precisely because they were susceptible to the justifications for some alternative view” (2002, p.227). So if indeed one or both groups found some rationale for a different view given their encounter, then Merli says, this is “evidence that their pre-encounter position didn’t meet the stipulations of the case, insofar as it wasn’t maximally coherent, stable, and so on. A change in view just shows that we weren’t really at the end of the day after all” (2002, p.227-228). Let’s see if things go any better if we suppose that our idealized moralizers do not change their views when they come into contact with one another.
§5.5.3 What if Earthers and Twin Earthers Don’t Change their Views?

Suppose that, after some prolonged discussion, Earthers and Twin Earthers elect not to change their moral views. If this was the case, Merli says, it would be reasonable to expect Earthers and Twin Earthers to view one another as using moral terms that do not mean the same thing. So though their conversation looked at first pass to be a case of the two genuinely disagreeing, that conversation now looks more like a case of merely verbal dispute, i.e. the two groups simply speak at cross purposes whenever they use terms like ‘morally right’ to express their views to one another. Whatever “disagreement” there was, that disagreement now looks to be at cross purposes (2002, p.228). So going forward, Earthers and Twin Earthers are likely to admit that they simply mean different things in using their moral terminology and so have an intractable disagreement in spite of possessing all of the relevant knowledge about the various kinds of actions they call ‘morally right’ and ‘wrong’. As Merli puts it, the two parties cannot resolve their disagreement “despite knowledge of all of the relevant facts of the case, they can’t really engage in meaningful argument with one another, and they each think the other has offered irrelevant reasons for its position” (2002, p.228). From here, Merli explains that one “interpretive option” is to view the two parties as simply dismissing one another’s views (2002, p.228). Can we reasonably expect that the two parties to treat one another as incompetent, i.e. as not actually being peers?

We could suppose that Earthers believe Twin Earthers are “deeply confused or simply incompetent” with moral concepts and likewise, Twin Earthers believe the same kind of thing about Earthers, that Earthers must be simply confused or incompetent in using moral terminology as they do (2002, p.228). This option is not a feasible one, Merli argues, in that it simply ignores the fact that we have assumed (like H&T) that Earthers and Twin Earthers are equally rational beings. So in this case, it would be more reasonable and indeed more charitable “for each group to think that the other is simply talking about something else” (2002, p.229).\(^{92}\) The only sensible thing to do at this point then, according to Merli, is to expect the moral dispute between Earthers and Twin Earthers to boil down to an equivocation between the meaning of their respective moral terms (2002, p.229). So ultimately, Merli says, Earthers and Twin Earthers need to dismiss the idea that ‘right’ and ‘wrong’ mean the same thing in their two languages (2002, p.229).

§5.5.4 The First Step Summed Up

The upshot of what we’ve seen so far is that the Earthers and Twin Earthers of Merli’s extended scenario cannot be reasonably expected to have a “fruitful discussion” when they exchange their moral views with one another (2002, p.229). As we’ve seen, there are at least three possible ways the exchange between the Earther and Twin Earther could go. Let’s begin with the

\(^{92}\) Merli notes that he does argue that they do have a disagreement, but that it is “not a disagreement about what’s right,” as we have supposed (2002, p.229).
third and least attractive possibility we have looked at, ‘possibility (3)’ we’ll call it. First, possibility (3) assumes that Earthers and Twin Earthers do not change their views when they encounter one another and exchange the moral views that they have come to in their inquiries. One way that this might happen is if Earthers and Twin Earthers end up viewing one another as either incompetent or confused in the use of moral concepts. This option cannot be a live one, Merli argues, because we have already stipulated that the two groups in Merli’s scenario are equally rational beings. Given that stipulation, we can at least expect each group to be charitable to one another. Earthers might think, for instance, that their Twin Earthen counterparts are simply talking about something else when they use the terms ‘morally right’ and ‘wrong’ to apply to various things. In other words, Earthers might have the initial impression that the Twin Earthen term ‘morally right’ means the same thing as the Earthen term ‘morally right’, i.e. in the way it is pronounced and how it is written, etc. Ultimately, it must mean something else given the way that Twin Earthers apply this word. As Merli recognizes, it would not take long for the Earthers and Twin Earthers to figure out this much about one another. In this case then, possibility (3) is not a live possibility. The root of this, as I have explained it that possibility (3) ignores the stipulation that the two groups are equally competent and rational. Let’s turn to the two other possibilities, possibilities (1) and (2).

The first option we looked at, possibility (1), suggests that we interpret Earthers and Twin Earthers as undergoing a kind of “conversion experience” when they two encounter one another. In undergoing this “conversion experience” the two groups would, Merli supposes, end up modifying each of their views or else adopt some kind of hybrid theory informed by both of the group’s moral thinking (2002, p.227). One ready explanation for this, Merli says, would be that the two groups actually arrived at the metaphorical table with less than “maximally coherent and stable end-of-the-day” moral theories (2002, p.229). On Merli’s scenario though, we have already stipulated that Earthers and Twin Earthers have maximally coherent and stable end-of-the-day moral theories. Again, this is built into the scenario, so like possibility (3) above, possibility (1) here cannot be a possible way the conversation might go between Earthers and Twin Earthers. Let’s look at our final possibility and see why Merli believes that this is the best way to expect things to go between the two groups.

In the second possibility that Merli considers, possibility (2), we are to suppose that Earthers and Twin Earthers do not end up changing their views when they come to exchange words with one another. We are to assume, in line with the scenario, that the two groups did indeed come to the table with “fully coherent and stable end-of-the-day stable” moral theories (2002, p.229). In this case, Merli explains, Earthers and Twin Earthers would be disposed to admit that their respective moral terms are not in fact univocal. After all, the two groups use their moral terminology, as we have assumed, with radically different sorts of properties in mind, in line with each of their respective moral theories, Tc andTd. It is no surprise then, at least on Merli’s view, that the two groups cannot end up having a genuine disagreement insofar as that disagreement
hinges on meaning the same thing by ‘morally right’ and ‘wrong’ etc. So in this case, Merli says, whatever discussion there is between the two groups about moral matters must ultimately be spurious. So even on the best interpretation of Merli’s scenario, Earthers and Twin Earthers end up admitting that they mean different things by their moral terms and so cannot really disagree in the kind of way that H&T imagine in their scenario. This completes the first step of Merli’s second objection and sets us up nicely for the second.

§5.5.5 Merli’s Second Objection: Step Two

The second step of Merli’s objection focuses on how the scenario he imagines above should influence our current thinking about moral matters and especially about the status of H&T’s scenario. According to Merli, we must take the idealized thinkers that he describes in his scenario seriously because “the judgments of the idealized moralists (concerning the equivocation involved in conversations with the Moral Twin Earthers) give us reason to reject the significance of our original intuitions about univocity” (2002, p.229). Merli argues here that because, under the idealized conditions above, our better-informed selves would reject the idea that twin-moral terms and our moral terms are univocal, then we too should reject that idea and dispose of our initial conviction that those terms mean the same thing. This, Merli says, is the most sensible way forward because it is reasonable to “grant more authority to one’s moral views after this process of inquiry than to one’s current views,” and as such, we have good reason to view those terms as indeed equivocal (2002, p.229).

At this point, the reader may object that Merli cannot simply help himself to the kind of expertise that our idealized and better-informed selves would have after a prolonged period of serious moral inquiry. In particular, the reader may be suspicious that our idealized selves could not have any better semantic expertise than us and so would know no better whether ‘right’ and ‘wrong’ are equivocal between our moral terms and those of the twin-moralists. Merli heads off this objection by insisting that the expertise of our idealized selves is only idealized with respect to the kind of moral expertise that they possess. These moralists are better-informed only insofar as they have already undergone the process of inquiry suggested by Merli and so they have a better grasp of how to apply moral terms correctly:

Just as experts with some concept in the natural sciences are better positioned to see which disputes about that concept are genuine, and which are spurious, our moral experts’ full understanding of our concepts—the ones we currently don’t fully understand—allows them to see more accurately which moral disputes are genuine. (2002, p.230)

As Merli explains, our idealized experts are only experts insofar as they better understand how are moral concepts can be correctly applied and this is enough for us to take their idealized

---

93 Merli assumes that genuine moral disagreement requires univocal meaning and so appears to disagree with the line developed by Plunkett and Sundell that was considered in Chapter Three.
judgments seriously in thinking about MTE. So to review, Merli’s argument takes the following format:

[I]dealized moralists and twin-moralists will see their putative disagreements as equivocal. But these idealized interlocutors have the same concepts (or mean the same things with their terms) as their present-day counterparts; they’re just better-informed about them. Because of this, we have reason to take seriously the judgments of the idealized future moralists. Hence we have reason to think that our initial judgment about the Moral Twin Earth case—that the disputes are univocal—is faulty. That is, we have reason to conclude that, contrary to initial appearances, we didn’t mean what the twin-moralists meant with their terms (2002, p.230).

Once this is clear, according to Merli, it should be clear why we have the faulty intuition that twin-moral terms must mean the same thing as our moral terms.

H&T design Moral Twin Earth to be so similar to Earth, especially vis-a-vis the function of moral practice, that the univocity of moral terms between the two groups looks guaranteed. Once we pay sufficient attention to the details though, Merli says, this guarantee about univocity becomes dubious. In particular, if we reflect on how things could proceed for our idealized selves (and the idealized twin-moralists), it is clear that these better-informed future moralists would reject the claimed univocity. Given this information and given that our better selves would judge that twin-moral terms do not mean the same thing as our moral terms, we should retract our initial judgment that those terms mean the same thing. We have good reason to make this adjustment, Merli argues, because our idealized selves would know better about whether moral terms mean the same thing or not in a given dispute, like the one imagined by H&T, given their (post-inquiry) moral expertise:

. . . our first reaction to the Horgan-Timmons case is driven by details of the scenario that will, ex hypothesi, be rejected as superficial or misleading by careful inquiry. This, I think, explains why we have the intuitions that we do in a way that lets us conclude that such intuitions are less significant than they appear to be. (2002, p.231)

§5.5.6 Why Merli’s Second Objection Cannot Succeed

I argue that Merli’s second objection cannot succeed because the extended scenario that Merli describes cannot be used to undermine the intuitions that H&T glean from their original MTE scenario, i.e. that Earthers and Twin Earthers mean the same thing by ‘morally right’ etc. and so genuinely disagree. The main reason that Merli’s scenario cannot do the requisite work for Merli is that, even if we accept everything he says up to the second step of the objection (i.e. about how his scenario plays out), Merli does not motivate why the idealized moralizers he imagines are in any better of a position to tell us how we should intuit H&T’s original MTE scenario. In particular, Merli does not provide any reasons for why his idealized moralists and twin-moralists possess a certain kind of moral expertise that we do not. As such, Merli does not adequately justify why his idealized thinkers would do any better in thinking about the intuitions we ought to have about H&T’s scenario—aside from simply begging the question and assuming that these idealized thinkers are correct about everything. When Merli brings in the so-called “authority”
that these idealized thinkers provide, this does all of the work of the argument. Conveniently, Merli does not argue for why these “better-informed” thinkers are better-informed at all, i.e. other than assuming that they possess the best end-of-the-day theory and so use moral terms perfectly. Moreover, Merli does not tell us anything about the end of inquiry either—about the kind of things that idealized moralizers approve of, etc.94

Merli argues that we should defer to these idealized moralizers, but it seems that there is no way to specify what this group of idealized moralizers would ultimately judge without begging the question. First, we cannot reliably know (practically speaking) the kind of things that they would judge and thus the sort of moral views they would adopt, and second, there are no good independent reasons for expecting these idealized thinkers’ views to correspond to or regulate our actual moral thinking today. We can only speculate about the sorts of actions that these idealized moralizers would generally commend, and likewise we can only speculate about the kind of theory that they would end up espousing at the end of the day. We simply do not have the kind of epistemic access that Merli seems to assumes we do and so there is serious degree of unwarranted faith on Merli’s part in presuming that his idealized thinkers should ultimately determine, for instance, how we should respond to H&T’s scenario. As such, Merli’s objection is unconvincing insofar as it depends on deferring to the sort of moral judgments these imaginary thinkers allegedly have. Moreover, if Merli’s argument here is cogent, it will also imply that the disagreement between Putnam and Nozick that we considered earlier isn’t a genuine disagreement. This is unpalatable as well and provides further reason to reject Merli’s second objection. Let’s see if Merli’s third and final objection can do any better.

§5.6 Merli’s Third Objection: “Preserving (Nonmoral) Disagreement”

In Merli’s third and final objection, Merli suggests that NMR can relocate the disagreement between Earthers and Twin Earthers in the MTE scenario as a kind of practical disagreement about what to do rather than a strictly-speaking moral disagreement about what is right. Merli argues that by relocating the disagreement in this way NMR can vindicate our collective intuition that the two groups disagree even if NMR cannot view their moral terms as univocal. The realist proponent of NMR must cite good reasons for doing this, given our initial conviction (in line with H&T’s assessment) that Earthers and Twin Earthers have a moral disagreement about whether ‘morally right’ ought to denote one set of properties or another, 94 It may be objected that I have not taken Merli’s main point seriously enough. It would be sensible, one could object, to give greater weight to ‘one’s informed, carefully considered opinions than to one’s hasty, ill-formed judgments’ and to ‘one’s moral views after this process of inquiry than to one’s current views’ (2002, p.229). If indeed, our ‘improved selves’ would ‘reject the idea that moral and twin-moral terms mean the same thing’, then this would be problematic for both H&T’s argument and my own, by proxy (Merli 2002, p.229). The problem with this objection is that it begs the question by simply assuming that, under ideal conditions, our ‘better selves’ would view the relevant terms as meaning different things. I am open to hear how these ideal conditions could be spelled out in a way that does not beg the question, but the current line of objection as it stands is a failure. I am also pessimistic about the prospects for such a specification of the relevant conditions, as I explain in the remainder of this section.
respective to each group’s normative views, Tc and Td. Merli argues that the proponent of NMR
does indeed have good reasons for shifting the disagreement in H&T’s scenario because Earthers
and Twin Earthers are, contrary to our initial conviction, better viewed as having a practical
disagreement. As such, proponents of NMR can “earn the right to say that there is something at
issue in cases [like MTE] where radically different evaluative standards come into play” (2002,
p.232). So even if NMR concedes that Earthers and Twin Earthers are talking about different
things (different properties) when they use ‘morally right’ in their respective languages, they can
still be viewed as genuinely disagreeing about what to do (2002, p.232). In the next sections, I
motivate this final objection from Merli and then argue that this final objection does not pose a
threat to H&T’s argument.

To begin, Merli explains how in general terms NMR can shift the disagreement in H&T’s
scenario even if we grant with H&T that moral and twin-moral terms are equivocal. The first
challenge for NMR is to show how our initial assessment that Earthers and Twin Earthers have a
moral disagreement is not the best assessment of that dispute. This is a serious challenge for NMR
because it is unclear how the two groups could be reasonably viewed as disagreeing if they simply
mean different things by ‘morally right’. Intuitively, the two could only speak at cross purposes.
According to Merli, NMR can concede this assessment and yet still capture the disagreement by
appealing to the different sort of actions that Earthers and Twin Earthers would advocate doing,
driven by the different normative views they possess. For example, Merli says, “[i]nsofar as
they’re committed to doing what’s twin-right, the twin-moralists urge us not to lie [even if it
increases overall human happiness], and good moralists urge just the opposite” (2002, p.232). By
capitalizing on this aspect of the dispute, Merli claims that NMR can view Earthers and Twin
Earthers as disagreeing even though the key terms in that dispute are admittedly equivocal across
the two groups, i.e. ‘morally right’ refers to different things in their respective languages. In fact,
this kind of disagreement is common here on Earth, Merli says:

People sometimes agree about what morality requires, and then go on to disagree about
whether to follow morality’s commands—perhaps because one party thinks that
considerations of prudence or etiquette carry the day. In these cases, the dispute isn’t
about what’s right; it’s about whether the properties of morality, prudence, or etiquette
are the ones that settle what to do. (2002, p.233)

According to Merli, NMR can view MTE as presenting this kind of disagreement, about which
normative standpoint—between Tc and Td—should settle what to do.

§5.6.1 Can Moving the Disagreement Help NMR Defend Against H&T?

Merli argues that NMR is in a good position to revise our initial convictions about MTE
for a few reasons. First, he says “some disagreement is better than none at all” (2002, p.233). So
the first and most basic point for Merli’s final strategy is that it can preserve the idea that Earthers
and Twin Earthers have a genuine disagreement about something, that “there’s something to talk
about when moralists and twin-moralists get together” (2002, p.233). Second, Merli claims that
his strategy answers the difficult question of whether the dispute on MTE is “centered on ‘right’ or somewhere else” and suggests how our initial convictions about the scenario should not be influential in answering that question. Third, Merli’s strategy can explain why “ordinary speakers might have trouble distinguishing between dispute about what’s right and dispute about what to do” (2002, p.233). According to Merli, ordinary speakers are apt to make this kind of mistake and “link a moral term with its broader evaluative accretion” because of the close association between “judgments of rightness and motivations to act accordingly” (2002, p.233). After all, these two matters typically come together. What NMR needs to show is that the dispute described in the MTE scenario between Earthers and Twin Earthers falls under the latter category, i.e. as a practical dispute about what to do and not strictly-speaking a dispute about what is right.95 If NMR can do this, as I have noted above, then it can meet H&T’s challenge by vindicating the intuition that the two groups disagree despite the fact that, on NMR, their moral terms do not mean the same thing. According to Merli, NMR can view the two groups as disagreeing because they have conflicting views about whether Tc orTd determines how to act. As such, the main question that we will be concerned with going forward is this: Can NMR make such a convincing case for shifting the disagreement on MTE? Let’s see.

§5.6.2 Is H&T’s Scenario Better Understood as a Practical Disagreement?

The main objective for NMR is to spell out an understanding of MTE that supports the view that Earthers and Twin Earthers would be best viewed as having a practical disagreement rather than a moral disagreement. Merli outlines a few promising ways this can be done that, according to him, can each defuse H&T’s challenge against NMR (2002, p.233). The first way that Merli proposes takes the two groups as having “clashing attitudes or commitments to action” (2002, p.233). On this proposal, NMR can view Earthers and Twin Earthers as disagreeing in that they possess different attitudes about what to do: Earthers express an attitude of approbation for doing what is in line with their normative theory Tc, whereas Twin Earthers express an attitude of approbation for doing what in line with their normative theory Td (2002, p.233):

Since our two normative theories (Tc andTd) come apart in their assessments, agents who are committed to doing what their theory labels ‘right’ will be motivated to act in contradictory ways. So dedicated moralists and twin-moralists have a kind of disagreement in attitude. (2002, p.233)

This is enough to save the intuition about genuine disagreement, Merli says, in that Earthers and Twin Earthers have a dramatic clash in their different attitudes about what to do. Unfortunately, Merli admits, this suggestion is not very helpful unless we know more about the sort of attitudes that the two groups differ on (2002, p.233). According to him, the difference in attitudes needs to be sufficiently robust to “ensure disagreement” (2002, p.233):

95 One may be apt to object here that disputes about what to do simply are disputes about what is right. I return to this point later in evaluating the objection.
It can’t be simply that we differ in attitude, the way we might differ in blood pressure; this isn’t yet a robust enough notion to ensure disagreement, let alone discussion and deliberation about broader normative questions. In order to capture the kind of incompatibility we’re after, we’d need some kind of attitude with some ‘outward-reaching’ force. My pro-stance toward doing the right thing has to conflict with, not merely be different from, your endorsement of the twin-right act. That is, I need to do more than commit myself to morality’s demands; I need to think that you should do the right (instead of the twin-right) thing, as well. (2002, p.234)

According to Merli, if Earthers and Twin Earthers have such dramatically different attitudes as this, then we can reasonably view the two groups as accepting different judgments about the “last ought before action” (2002, p.234).

§5.6.3 The “Last Ought Before Action”

As we saw earlier, Merli thinks that this kind of disagreement about the “last ought before action” is commonplace on Earth in that actual Earthers disagree about the appropriate evaluative view or standpoint that settles how to conduct ourselves, e.g. whether morality, prudence, or etiquette settles what to do: Morality demands one action, while prudence, etiquette, and so on demand others. In deciding how to act, we need to know not only what’s right, or what’s prudent—we also need to know whether to heed the demands of morality or prudence in those cases where the two come apart. (2002, p.234)

Naturally, people end up saying different things on this topic because they have different takes on which evaluative stance should be decisive in determining what to do: “One person might be committed to doing the right thing, while another is set on the profitable thing, or the polite thing, even if this requires violating moral demands that they recognize” (2002, p.234). Much of our deliberation involves this kind of interplay between competing evaluative standpoints. Even our internal deliberations involve this kind of interplay, e.g. “we wonder if moral reasons are outweighed, in some instances, by other kinds of considerations” (2002, p.234). Once we understand what the various options prescribe, we need to decide what evaluative standpoint is most appropriate for determining what to do. In the MTE scenario, Earthers and Twin Earthers need to decide whetherrightness or twin-rightness ought to guide their conduct in going forward—“Should [they] listen to morality, or twin-morality?” (2002, p.234). This question, Merli notes, “can’t be answered within one [framework] or the other, as both claim to give us the answer of the question of what to do” (2002, p.234).

However, according to Merli, there is room for a potential objection here: Even if we can treat Earthers and Twin Earthers expressing incompatible attitudes about what is to be done, that dispute can also viewed as equivocal and therefore spurious in that the two groups mean different things by ‘to be done’.

119
§5.6.4 Can H&T’s Challenge Be Revised to Target Broader Evaluative Discourse?

H&T’s challenge can rear its head again at this second level for Merli: How can Earthers and Twin Earthers genuinely disagree about whether $T^c$ or $T^d$ should regulate our behaviour if ‘to be done’ refers to different things between the two groups? For instance, if Earthers use ‘to be done’ to refer to some prudential properties and Twin Earthers use ‘to be done’ to refer to some properties of etiquette, then Merli’s suggested strategy is no help—H&T can simply relocate the challenge just as Merli relocates the disagreement. In order to head off this objection, Merli needs to show how the dispute he imagines at the practical level, about what to do, can preserve univocity and thus avoid the return of H&T’s challenge: “If we want to preserve dispute about what to do with our twin-moralist colleagues, we can’t allow for these disputes to be equivocal between, say, ourselves and the residents of some Normative Twin Earth” (2002, p.235). Merli claims that we can be reasonably optimistic on this front because our general evaluative terminology is “not parochial in the way we’re conceding (for the sake of argument) moral terms to be” (2002, p.235). The main reason that we can view our general evaluative terms in this way, according to Merli, is because the wide range of practical considerations that come up in our conversations do not seem to endanger the univocity of those evaluative terms:

After all, our confidence in the catholic nature of our disagreements about the last ought before action is unshaken by the variety of considerations that enter into that discussion. Most of us wonder, at one time or another, about what kinds of reasons to follow (moral, prudential, aesthetic, and so on). All of these have significant pull on our decisions about action. Yet despite the diversity of considerations taken into account in discussing what to do, we aren’t really concerned about the univocity of these conversations. Why should another normative standpoint, that is, twin-moral discourse, pose any additional threat? (2002, p.235)

As Merli says, if our general deliberations about what to do pose no problem for the univocity of general evaluative terms, then it is unclear how MTE could pose some new threat to that effect. Merli says a little bit more about why this challenge cannot arise again in this case.

Merli claims that, in order to stave off the objection that twin-moral discourse poses an unique threat against the above strategy, we need to spell out in more detail “what’s involved in making a judgment of what to do, all-in” (2002, p.235). Merli says there are two options for fleshing these details out: First, NMR can outline “an analogous account of the last ought before action” by, for instance, giving “an account of all-in endorsement that construes these judgments as being about motivationally inert natural properties” (2002, p.235). Alternatively, NMR can “combine [its] view of moral discourse with a different treatment of our talk of what to do” (2002, p.235). Merli goes in for the latter option.
§5.6.5 Toward an Expressivist View of All-in Endorsement

The best way that NMR can give a more detailed account of practical judgment, Merli says, is to combine NMR with “expressivism about all-in endorsement” (2002, p.236). Merli spells this out in a bit more detail:

According to this view, moral rightness is a matter of natural fact, but an answer to the question of what to do—not from a moral or a prudential or perhaps even a rational standpoint, but what to do, all-in—is not a factual judgment but an endorsement of one course of action or one set of reasons for action. When I get behind doing the right thing, I’m expressing my acceptance of certain norms, or urging others to act accordingly, or something along these lines. (2002, p.236)

At first, this move might seem rather bizarre on Merli’s part. How can one adopt NMR and expressivism at the same time? And how could adopting expressivism at the practical level help NMR meet H&T’s challenge? First, Merli says, most expressivists do not endorse global expressivism, i.e. expressivism about every kind of discourse. Instead, most expressivists accept expressivism for some domains and realism for other domains, e.g. about natural kinds. In Merli’s case, Merli accepts realism when it comes to moral discourse but expressivism when it comes to practical or evaluative discourse about all-in endorsement. According to him, this is an appealing view for three reasons:

The first has to do with skepticism about the ability of any substantively constrained ‘ought’ to play the role of the last ought before action. The second, related consideration has to do with the connection between judgments of what to do and motivation. The third reason concerns the differences in descriptive content between moral discourse and talk of what to do, all things considered. (2002, p.237)

Let’s look at each of these in detail.

First, as Merli says, we might be hesitant to identify some evaluative standpoint with the last ought before action (2002, p.237). Merli’s main reason for this may be familiar to the reader as the driving worry behind Chapter Two: “there’s a kind of open question that can be raised about whether or not the deliverances of any of those perspectives settle what to do” (2002, p.237).

As the reader may recall, H&T’s challenge can be viewed as a revised version of Moore’s classical open question argument (OQA), updated to deal with the kind of modern (synthetic) semantic theory that NMR goes in for. Of course, H&T’s challenge is directed at NMR’s treatment of moral discourse; Merli’s worry is that this challenge can be easily shifted to target evaluative discourse. Merli uses an example to bring this out:

Suppose that we think that what an agent has reason to do depends on his (intrinsic) desires, broadly construed. We might admit that a person with wicked desires has reason to do horrid things, and that he wouldn’t be irrational for doing them, but we might still say, sensibly, that he ought not to. What we’re doing is more than merely point out that from the moral point of view, which he cares nothing about, he’s doing something wrong. This is something to which he can assent. We’re also saying that this point of view settles what to do, that he should (all-in) do what’s right instead of acting as he desires. So it seems coherent to think that, when we evaluative others’ actions, we don’t take facts about rational requirement to settle the question of what to do, and this indicates that we think the two questions are distinct. (2002, p.237)
According to Merli, we can repeat this same argument for any “substantively defined evaluative standpoint” and so, he says, we are better off viewing the last ought before action as having a kind of prescriptive rather than descriptive role (2002, p.237).

The second reason that Merli moves toward an expressivist view of all-in endorsement deals with the connection between motivation and action. One of the key challenges for any realist moral theory is to make sense of the connection between an agent’s judging that some action is right and his or her being motivated to act accordingly. As Merli says, realists often make “a great deal of the fact that sometimes moral judgments don’t motivate agents to act accordingly; as they like to point out, the tie between judgment and motive mustn’t be too tight to account for the everyday phenomena of our moral experience” (2002, p.238). By contrast, Merli says, “we might think there’s a tighter connection between judgments of what to do and motivation” (2002, p.238). For instance, we can make better sense of “someone unmotivated to follow through on his moral judgments than we can of someone who invokes the last ought before action and then fails to act” (2002, p.238). In the moral case, Merli says, it is more plausible to identify moral properties with natural properties (i.e. to accept NMR), given the wider gap between moral judgment and motivation. In the more broadly evaluative case, though, Merli says that this gap between judgment and motivation is much tighter and in that case, it is more reasonable to view all-in judgments about what to do as not involving natural properties.

Merli’s third reason for opting for an expressivist view about all-in endorsement centres on the differences between how we talk about moral discourse and how we talk about what to do (2002, p.237). Basically, Merli thinks the former is thickly descriptive while the latter is only thinly descriptive. When it comes to our talk about moral discourse, our moralizing is ‘thickly descriptive’ because we are very specific about the kinds of considerations that figure into our moral judgments (2002, p.238). According to Merli, our “moral judgments can’t be about just anything . . . When we think about morality, we have some rough idea of the kinds of considerations that are relevant (facts about benefits and harms to creatures like us, for examples) and facts that don’t matter (the height of an agent, perhaps)” (2002, p.238). This descriptive thickness is important for Merli in that “it suggests that the way to capture what’s distinct about morality is by appeal to a certain subject matter or content, not a particular set of formal features” (2002, p.238). Unsurprisingly, Merli believes that this aspect of our moralizing should drive us toward viewing moral judgments as being descriptive, in line with NMR. Broad evaluative judgments on the other hand are ‘descriptively thin’ (2002, p.238):

... people can take all sorts of considerations to settle how to act, and, as long as they satisfy certain formal constraints, we can think of them as coherent ... they’re still making a judgment of what to do, unlike the moral eccentric, whose deviance might prevent him from taking part in moral discourse at all. (2002, p.238)

So ultimately, Merli argues, we should opt for an expressivist view of all-in endorsement in that the considerations relevant to our broadly evaluative talk are, unlike moral talk, not limited to thick descriptive content.
§5.6.6 Why Merli’s Third Objection Cannot Succeed

I argue that Merli’s third argument collapses to at least one of two objections. The first objection takes issue with Merli’s interpretation of H&T’s scenario. In this objection, I claim that the Earther and Twin Earther of H&T’s scenario cannot be plausibly viewed, on any interpretation, as engaging in a practical disagreement about what to do, rather than a strictly-speaking moral disagreement about what’s right. Insofar as Merli interprets the scenario in this way, his objection cannot damage H&T’s argument. The second objection I propose assumes that the first objection can be met. In this objection, I argue that even if Merli can view the disagreement as a practical one and bring in expressivism about all-in-endorsement to explain this disagreement, Merli faces a dilemma: First, if expressivism can successfully explain the disagreement between the two parties, then NMR is unmotivated, since one could then propose an expressivist treatment of morality that would be cheaper than NMR on matters of moral metaphysics and epistemology. Second, if expressivism cannot successfully explain the disagreement between the two parties, then NMR cannot meet H&T’s challenge and Merli’s objection fails. Let’s look at each of these two objections in turn.

§5.6.7 Why Merli’s View of the MTE Scenario is Faulty

I argue that Merli’s third objection cannot damage H&T’s argument because Merli’s interpretation of the MTE scenario involves a substantive departure from the one described by H&T. The key point of difference is Merli’s view that Earthers and Twin Earthers stand in a practical disagreement about what to do rather than a moral disagreement about what’s right, i.e. about what sorts of actions are ‘morally right’. If we return to the details of H&T’s original scenario, it is clear that there is no plausible way to interpret the two parties as having this kind of disagreement. The main piece in the scenario that blocks this interpretation is H&T’s stipulation that Earthers and Twin Earthers assign *overriding importance* to their considerations about moral matters in deciding what to do:

. . . the uses of [twin-moral] terms on Moral Twin Earth bear all of the ‘formal’ marks that we take to characterize moral vocabulary and moral practice. In particular, the terms are used to reason about considerations bearing on Moral Twin Earthling well-being; Moral Twin Earthlings are normally disposed to act in certain ways corresponding to judgments about what is ‘good’ and ‘right’; they normally take considerations about what is ‘good’ and ‘right’ to be especially important, even of overriding importance in most cases, in deciding what to do, and so on. (1992, p.164)

Given that their considerations about what is ‘good’ and ‘right’ have this overriding importance, the sort of contrast between moral judgments and judgments of what to do all-in on which Merli’s third objection depends is simply not available given the way H&T describe the MTE scenario. John McDowell makes a similar case against the suggestion that moral reasons might be trumped by other kinds of reasons:

[T]he dictates of virtue, if properly appreciated, are not weighed with other reasons at all, not even on a scale which always tips on their side. If a situation in which virtue imposes
a requirement is genuinely conceived as such, according to this view, then considerations which, in the absence of the requirement, would have constituted reasons for acting otherwise are silenced altogether—not overridden—by the requirement (McDowell 1978, p.26).

Moral reasons are a special breed of practical reasons in that, among other things, they have overriding importance in our deliberations about what to do. McDowell takes this even further in the passage above: moral reasons “silence” other reasons altogether. This key function of our moral thinking is the primary reason that, as H&T emphasize, the intuitive data gleaned from MTE and from Putnam’s case do not match. As such, any plausible interpretation of H&T’s scenario must account for the fact that moral considerations have overriding importance in decision making. Merli simply ignores this function of moral vocabulary and because of this, his interpretation of the MTE scenario cannot be used to bolster an argument against H&T (given that it lacks the key ingredient that yields the linguistic intuitions we have about that case).

§5.6.8 Why Expressivism Cannot Save NMR

Let’s assume for a moment that Merli can meet the response spelled out above and that the Earther and Twin Earther of H&T’s scenario can be plausibly viewed as having a practical disagreement about what to do that contrasts with disagreement over moral rightness. Moreover, let’s assume with Merli that expressivism about all-in-endorsement can explain this disagreement. If we grant all of this, it turns out that Merli’s prescribed explanation of Moral Twin Earth forces a dilemma: On the one hand, if NMR must deploy an expressivist treatment of all-in-endorsement in order to explain moral disagreement, this would force NMR into a precarious position; NMR would need to address the standard objections against expressivism in addition to the standard objections against NMR. If NMR must meet the standard objections against expressivism (objections that NMR is designed to avoid), it is unclear what motivation there could be for accepting NMR over the alternatives. What advantage could NMR have over, say, accepting expressivism and at least avoiding the standard objections to realism? Things are no better for NMR on the other horn of the dilemma: If we suppose that expressivism cannot explain the disagreement between the Earther and Twin Earther, then NMR simply cannot meet H&T’s challenge by using an expressivist view of all-in-endorsement.

§5.7 Conclusion

It seems to me, then, that all of Merli’s objections to H&T’s Moral Twin Earth argument in the end fail, so that they are not available to damage my Meaning Twin Earth argument. In the next chapter, I go on to flesh out the main options for a theory of normative judgment that would explain the intuitions elicited from the Meaning Twin Earth scenario.96

96 It might be argued that I am not justified in presupposing that if Merli’s fails in the moral case it has no validity in the case of meaning. However, given that morality is the paradigm example of a normative area of thought and talk, the H&T-style Twin Earth argument is more likely to work for the case of morality than in any other area. So—in the absence of a detailed argument to the contrary—I can assume that if an
objection to the argument in the moral case fails, it will fail likewise in the case of meaning. This is a defeasible assumption, of course, but the onus at this point is on my opponent to provide the necessary detail.
Chapter Six: In What Sense is Meaning Normative?

§6.1 Introduction

This chapter outlines in broad terms two candidates for a theory of normative judgment that would explain the intuitions elicited from the Meaning Twin Earth scenario of Chapter Two. First, I consider how a theory of this kind is developed in the moral case by returning to R.M. Hare’s argument for prescriptivism, based on the ‘missionary’ thought experiment. I explain that, as Smith (1994) contends, there are effectively two main options for a theory of normative judgment: an expressivist form and a rationalist form. Gibbard (2012) adopts an account of the former type and uses an analogue of the missionary scenario to argue that meaning is normative. Baker (2016) claims that this argument fails. According to him, we simply do not have suitably similar intuitions in this case. Moreover, Baker argues that an analogue of the Moral Twin Earth scenario fails to generate analogous intuitions in the meaning case. I argue that Baker misconstrues each of these scenarios and that once we address this, we do get intuitions in favour of the normativity of meaning (and against reductive dispositionalism). I sketch out Gibbard’s expressivist view of semantical judgment in more detail and then turn to McDowell’s non-reductive realist alternative. I outline some obstacles ahead for each of these accounts and argue that, provisionally, McDowell’s appears to be more promising than Gibbard’s as far as meeting these obstacles is concerned.

§6.2 Return to Hare

Recall that in Hare’s scenario, Hare envisions a missionary who visits an island of cannibals and, in the course of trying to impress his religious and moral views on these cannibals, the two come to exchange their views about the sort of things that they each call ‘good’. The cannibals quickly discover that the missionary uses his term ‘good’ in a peculiar way, to apply to people who are meek and mild. The cannibals find this very strange; they use their term ‘good’ to apply to people who collect as many scalps as possible. On Hare’s view, the missionary and cannibals are “under no misapprehension” about the meaning of ‘good’ and are best viewed as genuinely disagreeing (1952, p.148). Hare uses these intuitions as evidence against a descriptivist moral semantics by showing that, if this semantics was correct, we would end up with highly unintuitive results with regards to the missionary scenario.

A descriptivist moral semantics would imply that the missionary and cannibals do not genuinely disagree on account of their meaning different things by their respective terms, the missionary ‘good’ and the cannibal ‘good’. The missionary uses his term ‘good’ to refer to the property of being meek and mild, and the cannibals use their term ‘good’ to refer to the property
of collecting many scalps. As such, Hare explains, descriptivism would imply that the two simply speak at cross purposes in exchanging their views:

[The missionary and cannibals] would not be disagreeing, because in English (at any rate missionary English), ‘good’ would mean among other things ‘doing no murder’, whereas in the cannibals’ language ‘good’ would mean something quite different, among other things ‘productive of maximum, scalps’. (1952, p.149)

This suggests that we cannot view the missionary ‘good’ and the cannibal ‘good’ as descriptive terms. In order to view the two as disagreeing, Hare says, we need to view the missionary and cannibals terms as having the same meaning on account of their performing the same evaluative role in the language in which those terms are used, as the “most general adjective of commendation”:

It is because in its primary evaluative meaning ‘good’ means neither of these things [“doing no murder” or “productive of maximum scalps”], but is in both languages the most general adjective of commendation, that the missionary can use it to teach the cannibals Christian morals. (1952, p.149)

Moreover, Hare claims, any plausible account of our moral practice must leave room for “moral development” and this essentially involves the idea that moral judgment and thus moral terms are essentially evaluative. The idea that our moral judgments simply refer to a set of standards is too rigid and “the words used in referring to them . . . too dominantly descriptive” (1952, p.150). As Hare explains, these standards are bound to change as circumstances change and the “instrument” of this change is the “evaluative use of value-language” (1952, p.150). Moreover, Hare says this evaluative use of value-language is the cure for moral stagnation:

The remedy, in fact, for moral stagnation and decay is to learn to use our value-language for the purpose for which is it designed; and this involves not merely a lesson in talking, but a lesson in doing that which we commend; for unless we are prepared to do this we are doing no more than pay lip-service to a conventional standard. (1952, p.150)

This brings us to the focal point of Hare’s criticism of descriptivism. Descriptivism, according to Hare, cannot make sense of the idea that moral judgment and moral language are tied to the will. If an agent judges that some action is right, then he or she ought to be “prepared” to comply with that judgment. Smith (1994) develops this idea.

§6.3 Smith: Two Options for a Theory of Normative Judgment

One intuitive feature of our moralizing, according to Smith (1994), is the idea that moral judgments have a ‘practical upshot’. This feature of our moral practice is generally labelled ‘internalism’ and refers to a broad category of claims about the connection between our moral judgments and our having some reasons or motivation to act in accordance with those judgments (1994, p.60). Smith outlines three main internalist claims that vindicate the idea that there is a conceptual connection between moral judgment and the will (1994, p.61). The first and strongest claim is the thesis that an agent’s moral judgments as a matter of conceptual necessity motivate that agent: “If an agent judges that it is right for her to Φ in circumstances C, then she is motivated
to Φ in C” (1994, p.61). It is unlikely that moral judgments are as closely linked to motivation as this first claim suggests because, Smith argues, this requires the abandonment of the idea that our motivation is defeasible, that sometimes we suffer from weakness of the will or the like and so fail to be motivated to comply with our judgments about what is right and wrong (1994, p.61). Consequently, Smith says it is more plausible to opt for the alternative view that moral judgment has a weaker and defeasible connection with the will. Smith embodies this view into a second internalist claim: “If an agent judges that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational” (1994, p.61). Given that we are subject to the “distorting influences of weakness of the will and other similar forms of practical unreason,” we sometimes fail to be motivated to do what is right (1994, p.61). In most cases though, these influences are absent and so we are typically motivated to comply with our moral judgments.

A final way to make sense of the idea that there is a ‘practical upshot’ to our moral judgments is by supposing that there is a conceptual connection between “the content of a moral judgment—the moral facts—and our reasons for action” (1994, p.61). This amounts, on Smith’s view, to the following third internalist claim: “If it is right for agents to Φ in circumstances C, then there is a reason for those agents to Φ in C” (1994, p.62). Alternatively, as Smith puts it, “moral facts are facts about our reasons for action; they are themselves simply requirements of rationality or reason (1994, p.62). Smith argues that this third claim may explain or even entail the second internalist claim above: An agent has a reason to act insofar the agent would be motivated to act accordingly if he or she were rational (1994, p.62). As such, Smith says, the agent who judges that he or she has a reason to act and yet fails to be motivated to do so, must be practically irrational: “if [the agent] is not motivated accordingly, then she fails to be rational by her own lights” (1994, p.62). According to the third form of internalism “the judgment that it is right to act in a certain way is simply equivalent to the judgment that there is a reason to act in that way” and so the third form of internalism entails the second form (1994, p.62). Nonetheless, Smith notes, the reverse claim is not true—the third claim does not follow from the second:

Expressivists, for example, agree that someone who judges it right to act in a certain way is either motivated accordingly or practically irrational in some way, but deny that moral requirements are requirements of rationality or reason. They thus accept the second internalist claim because they think that a moral judgement is the expression of a preference, or perhaps the expression of a disposition to have a preference; but they reject the third because they think that fully rational creatures may yet differ in the preferences that they have, or are disposed to have. (1994, p.62)

So though one must accept the second claim if one accepts the third, one can nonetheless accept the second and yet reject the third. Smith labels the second claim the ‘practicality requirement on moral judgment’ and the third ‘rationalism’. These two claims delineate the main branches of internalism.

In the next sections, I argue that a semantic analogue of Hare’s argument supports the contention that meaning is normative, i.e. that semantic judgment is essentially evaluative or action-guiding in the sense described by Smith. Baker (2016) argues that such an argument cannot
succeed on the grounds that we do not have relevantly similar intuitions in the meaning case. I outline Baker’s argument and then respond to that argument by arguing that the scenario Baker uses is not in fact a suitable analogue of Hare’s missionary scenario. Once this discrepancy is addressed, it turns out that a proper analogue of Hare’s scenario does support the normativity of meaning. I use the Meaning Twin Earth argument proposed in Chapter Two to bolster this argument.

§6.4 Baker: Intuitions about Disagreement Do Not Support the Normativity of Meaning

Baker (2016) develops a metasemantic version of Hare’s thought experiment to test whether ‘means’ shares the special normative status as moral terms like ‘good’. Before he argues how this version fails to secure the right intuitions in support of the normativity of meaning, Baker explains what sets normative terms apart from other sorts of terms when it comes to cases of disagreement. The main thing that makes normative terms unusual in these cases, Baker says, is that they seem to violate the general principle that two terms will possess the same meaning only if those terms are roughly coextensive for competent speakers (2016, p.71). Normative terms apparently break this rule of translation by allowing radical disputes—disputes that involve radically non-coextensive uses of a term—to count as genuine disagreements: “our intuitions seem to allow for the possibility of radical disputes in the case of normative terms” (2016, p.72).

As Baker explains, Hare took this unusual property of moral terminology as evidence for expressivism. According to Hare, we cannot reasonably treat moral terms as possessing meaning in virtue of what they refer to, but instead in terms of the prescriptive attitudes they express. That is, we should not attempt to give ‘good’ a referentialist semantics, but should adopt a form of prescriptivism (expressivism) instead.

The most plausible assessment in Hare’s case involves viewing ‘good’ as meaning the same thing between the missionary and cannibals and then concluding that the two simply have incompatible views about the sort of actions that are ‘good’. As Baker has emphasized, the fact that the dispute is a moral one centred on the term ‘good’ is here responsible for our collective intuition that the missionary and cannibals genuinely disagree, i.e. despite the different extensions that the term ‘good’ has in each of their languages. The same goes for the Moral Twin Earth devised by H&T—the fact that the dispute between the Earther and Twin Earther centres on the term ‘good’ enables us to view their disagreement as genuine. Baker’s principle for translation is violated in both cases. If we look to more ordinary cases, involving any other sort of terms, we tend not to allow the same degree of difference in application between speakers who use a given term to exchange their views with one another, e.g. about ‘water’ or ‘pigeon’. In these cases, Baker says, we are more inclined to say that these speakers simply mean different things by a

97 Baker’s main target in using this test is Gibbard (2012), who argues that an analog of Hare’s scenario can be to show that the concept MEANING is normative.
term if their applications of that term are not broadly coextensive (in line with the general principle for translation). According to Baker, the analogue for Hare’s scenario in the case of meaning suggests that ‘means’ is less like ‘good’ and more like other ordinary sorts of terms. As such, Baker says, this analogue inspired by Hare’s argument cannot support the view that ‘means’ is a normative term.

The semantic analogue Baker develops from Hare’s scenario is designed to be a test: If our intuitions in the meaning case match those in the moral case, then we have evidence to think that meaning is normative. If not, then then we have no such evidence. The result of this test, Baker says, is the latter; Hare’s scenario cannot be used in the case of ‘means’ to generate evidence for the normativity of meaning. Meaning may still be normative, Baker says, but the analogue from Hare’s argument does not provide evidence for that thesis. Let’s look at Baker’s scenario and see why he thinks the scenario fails to generate the same results as the moral case originally devised by Hare.

§6.4.1 Baker’s Scenario

Baker reimagines Hare’s thought experiment as one that focuses on a metasemantic dispute. In this new scenario, Baker borrows two characters, Jerry and Tyler, from Gibbard’s presentation of Hare’s argument in his Meaning and Normativity (2012). Jerry, we are to imagine, is an individualist about meaning, whereas Tyler is a communitarian about meaning. Baker imagines that Jerry and Tyler travel to an island inhabited by cannibal metasemanticists. These metasemanticists have access to very advanced technology, including a giant machine that enables the three parties to access all of the non-normative and non-meaning facts about the world, “all of the facts which potentially form part of meaning’s supervenience base” (2016, p.72). Jerry, Tyler, and one of the cannibals use this giant machine to look at a woman named Jenny. The machine tells us that “Jenny is disposed to apply ‘green’ to all and only green objects, but people in her community are disposed to apply the term ‘green’ to green objects before t and blue objects after” (2016, p.72). Moreover, the machine tells us that “[n]o one is disposed to apply the term to red objects” (2016, p.72). Jerry, Tyler, and the cannibal say the following things about Jenny:

Jerry says, “‘Green’ uttered by Jenny means GREEN’.

Tyler says, “‘Green’ uttered by Jenny means GRUE’.

But the cannibal says, “‘Green’ uttered by Jenny means RED’ (2016, p.73)

From here, the three assign meanings to all of Jenny’s other terms and the general pattern above continues: “Jerry and Tyler actually agree a lot . . . The cannibal never agrees.” (2016, p.73). Jerry and Tyler question the cannibal on this front. In particular, they ask “if he realizes that on his theory of meaning, neither individuals nor communities ever use their own concepts correctly” (2016, p.73). The cannibal answers in the affirmative, that “we agree about the subvening facts, we just disagree about what meanings supervene on them” (2016, p.73).
Baker claims that our intuition in this scenario is that Jerry and Tyler genuinely disagree and the cannibal simply speaks at cross purposes with everyone else. Whatever the cannibal means by ‘meaning’ must be something other than what Jerry and Tyler are talking about. Baker justifies this assessment by using the principle for translation he introduces to explain Hare’s original scenario. Recall that in Hare’s scenario, the term ‘good’ can be plausibly understood as meaning the same thing between the missionary and cannibals of that scenario because ‘good’ has the unusual property of permitting radical disputes: “Even if two speakers apply the term ‘good’ in radically different ways, we are still able to hear disagreements between them as genuine” (2016, p.73). In the case of ‘means’ though, Baker argues that the threshold for disagreement is lower: “. . . in the case of ‘meaning’, there is a much more significant limit to how radical disagreement can be before we hear it as merely verbal” (2016, p.73). Jerry and Tyler disagree about Jenny, but they also agree about a wide background of other cases. In this respect, Jenny’s case is anomalous (2016, p.73). As we know, Jerry and Tyler ascribe to quite different research programs. Consequently, the two disagree about the sort of linguistic phenomena that are central and those that are peripheral in determining meaning (2016, p.73). With this difference in mind, we can reasonably expect cases like Jenny to bring the two apart in that the two will have competing explanations for what counts as Jenny’s meaning something one thing or another. The cannibal, on the other hand, does not have any explanatory overlap with Jerry and Tyler and moreover does not offer any explanation for his judgment that Jenny means RED by ‘green’. For this reason, the cannibal cannot be viewed as genuinely disagreeing as Jerry and Tyler do.

The main failure on the cannibal’s part is that his theory of meaning fails the “non-negotiable desideratum on [any] theory of meaning . . . that it explains linguistic behaviour: why people speak and judge as they do” (2016, p.73). Jerry and Tyler’s theories abide by this desideratum in that they offer competing explanations for what Jenny means vis-à-vis the different linguistic phenomena they take to be relevant in determining what she means. For any theory of meaning, Baker explains, “there is a class of phenomena that intuitively make up the data for a theory of meaning to explain, and which both sides agree their theories must by and large explain, with some small remainder to be explained with auxiliary hypotheses, and an even smaller remainder to be at the very least explained away” (2016, p.74). Baker claims that it is this explanatory constraint that sets the meaning case apart from the moral case and crucially, it is this constraint that the cannibal of Baker’s scenario fails to meet.

According to Baker, the key feature responsible for the different results between Hare’s case and his own is that ‘meaning’ must play an explanatory role whereas ‘good’ does not: “there is no requirement that a theory of goodness will explain actual behaviour” (2016, p.75). In Hare’s case, this explains why we can treat the missionary and cannibals as disagreeing even though they prescribe radically different kinds of behaviour when they use the term ‘good’ in each of their languages. Given that an account of ‘meaning’ must be sufficiently explanatory, Baker’s case fares differently. The explanatoriness requirement puts a limit on how far we can interpret ‘means’
as meaning the same thing between two parties and thereby, on how far we can interpret two parties as genuinely disagreeing. This drives Baker to conclude that ‘means’ is not marked by the characteristics that allows a Hare-style argument to be run for ‘good’ and other normative terms.

§6.4.2 Reply to Baker’s Objection to Gibbard

I reply to Baker’s argument in two stages. In the first stage, I take issue with the scenario that Baker develops for his argument. I argue that this scenario is not a suitable analogue of Hare’s original thought experiment. If we repair this scenario, with some inspiration from the Meaning Twin Earth scenario of Chapter Two, it does yield evidence for the normativity of meaning. With this scenario in place, I turn to the second stage of my reply to Baker. In this stage, I object to Baker’s focus on the explanatory constraint on any account of meaning. I argue that this is a red herring on Baker’s part. The real issue, as my amended scenario shows, has to do with another constraint on any account of meaning, namely a *normativity constraint*.

§6.4.3 Why Baker’s Scenario Is Inadequate

Baker’s scenario is not a suitable analogue of Hare’s original thought experiment insofar as Baker mislocates the appropriate disagreement. The main reason that Baker’s scenario fails in this respect is because he assumes the cannibals have no principled reasons for viewing Jenny as meaning RED by ‘green’. It is this assumption that does all of the work for Baker’s argument in that naturally, this assumption about the cannibals preclude them from having any kind of genuine disagreement with the other parties, Jerry and Tyler. This marks a serious departure from Hare’s scenario. The cannibals described by Hare do have principled reasons for using ‘good’ as they do, to commend people who collect many scalps. The cannibals have a moral outlook that commends people who do this among other things. Baker does not assume anything of the sort about the metasemantic cannibals of his scenario, save for the fact that they think semantic predicates work in a radically different way from what Jerry and Tyler’s theories suggest. So ultimately, Baker chooses a rather inopportune thought experiment and this leads him to conclude that the scenario does not yield analogous intuitions in favour of the normativity of meaning. I argue that if the right changes are made to Baker’s argument, with an eye on Meaning Twin Earth, an analogue of Hare’s argument does support the normativity of meaning.

§6.4.4 How to Recalibrate Baker’s Scenario

I argue that Baker’s test, with some modification, does yield intuitions for the normativity of meaning if we use a proper analogue of Hare’s thought experiment. As I have mentioned, we can get such an analogue by combining Baker’s scenario with some of the features from my Meaning Twin Earth scenario. The main thing that this analogue aims to do is locate the right disagreement in the case of ‘meaning’ and then determine whether this disagreement provides evidence for the normativity of meaning. Baker failed to do this properly by using a scenario that does not conform to the one described by Hare. Once we rectify this mistake, I explain that
Baker’s focus on the explanatory constraint on any account of meaning turns out to be a red herring. In the next sections, I formulate a new scenario that is closer to the one described by Hare and then assess that scenario in line with Baker’s test for normativity.

§6.4.5 Baker’s Scenario Revised

Imagine that Jerry*, an individualist about meaning, visits an island inhabited by cannibal metasemanticists. Jerry* aims to convert these cannibals to his view about meaning. In the process of trying to convert these cannibals, Jerry* learns that the cannibals have access to very advanced technology, including a giant machine that enables Jerry* and the cannibals to access all of the non-normative and non-meaning facts about the world. Jerry* comes to exchange his views with one of the cannibals, Tyler*, who explains their communitarian view about meaning to Jerry*. When Jerry* informs Tyler* about his individualist theory of meaning, Tyler* and the rest of the cannibals find this rather odd in that Jerry* is willing to ascribe meaning to speakers that their communitarian theory rules out as meaning anything at all. Jerry* and Tyler* step up to the giant machine that allows access to all non-normative facts. They use this machine to look at a border case that Tyler* takes as a counterexample to the individualist theory espoused by Jerry*. Tyler* asks Jerry* to observe, using the machine, a solitary individual Jenny* who inhabits a nearby island. The machine reveals that Jenny* was at a very young age shipwrecked on this island and managed to survive in spite of her solitary existence by hunting and foraging food on her own. One of the things that Jenny* does, the machine shows, is exclaim “Blackberry!” whenever she encounters one of the berries that are a staple in her diet. Jerry* and Tyler* exchange their views about what Jenny* means by her words:

Jerry* says, “‘Blackberry!’ uttered by Jenny* means BLACKBERRY.”

Tyler* says, “No, ‘Blackberry!’ uttered by Jenny* does not mean BLACKBERRY nor anything else!!”.

The two assign meanings to the rest of Jenny’s terms and the general pattern continues. Jerry* insists that all of the utterances that Jenny* speaks are meaningful and Tyler* uniformly disagrees with this. Tyler* takes Jenny* to not mean anything by any of her terms. Tyler* presses Jerry* on this topic.

Tyler* is especially sceptical about how an individualist theory, like the one espoused by Jerry*, can preserve the idea that there is a distinction between what seems right and what is right with respect to a speaker’s utterances. Tyler* uses the example of Jenny* to bring this problem out against Jerry*. According to Tyler*, his communitarian theory can make sense of this distinction by appealing to whether a speaker’s use of an expression conforms to the use of that expression by his or her speaking community. Jenny* has no such background, Tyler* explains, and so Jenny* cannot mean anything by her utterances. In fact, Tyler* and the other cannibals are

It may have been more appropriate to use Jerry* and Saul* here, given that we will discuss KW’s communitarian view later in the chapter, but I stick to Tyler* here for the purposes of responding to Baker.
in part motivated to accept their chosen theory of meaning in light of cases like this, where the seems right/is right distinction is inapplicable. When pressed about this, Jerry* concedes the need for the distinction between what seems right and what is right. Nonetheless, Jerry* explains, this distinction can be drawn by citing some system that Jenny* has for her uttering “Blackberry!” whenever she encounters such a berry. For example, Jerry* may cite that Jenny* associates a specific smell and taste with these berries and so knows when to exclaim “Blackberry!” . As such, there is room to view Jenny* as meeting the requirements for drawing a distinction between what seems right and what is right in virtue of her having this system. For this reason, Jerry* claims that Tyler* is wrong in saying that Jenny* does not mean anything by “Blackberry!”.

§6.4.6 Assessing the Revised Scenario

Intuitively, Jerry* and Tyler* genuinely disagree, though this disagreement does not hang on the explanatory role of their theories. Jerry* and Tyler* both have adequate reasons for their views about Jenny*. In other words, both parties accept the requirement that ascriptions of meaning are sufficiently explanatory vis-a-vis a speaker’s behaviour. The heart of the disagreement between Jerry* and Tyler* has to do with the applicability of the seems right/is right distinction, which serves as a constraint on any account of what Jenny* means (and more generally, any account of a speaker’s meaning full stop). Jerry* and Tyler* disagree because they disagree about whether Jenny* can mean anything by her utterances in light of this distinction. For Jerry*, Jenny* can mean something by her utterances and for Tyler*, she cannot.\textsuperscript{99} For Tyler*, the requirement to preserve the seems right/is right distinction trumps the explanatory considerations.\textsuperscript{100} If we drop the constraint embodied by the distinction between what seems right and what is right, then it is unclear how we can plausibly explain the disagreement between Jerry* and Tyler*. Therefore, there seems to be similar scope in the semantic case for interpreting ‘meaning’ as having the same status as ‘good’, as permitting some radical disputes to be plausibly viewed as genuine disagreements, and crucially, this is because meaning predicates perform the same evaluative role for Jerry* and Tyler* and their semantic judgements about Jenny* (even though they apply their respective meaning predicates differently). This is why I have referred to this further constraint on any account of Jenny*’s meaning a normative one. (Note that since Jerry* and Tyler* have different first-order semantic theories about the constitution of correctness conditions, it will not be possible to recover a locus of disagreement between them simply in

\textsuperscript{99} One may object by pointing out that Jerry* and Tyler* will actually agree most of the time when they apply their terms to individuals within communities. Baker emphasised this point about Jerry and Tyler in his scenario and I believe that similarly, Jerry* and Tyler* are likely to agree most of the time. Nonetheless, this does not bear on the disagreement over Jenny*. Even if Jerry* and Tyler* uniformly apply their terms for meaning ascription in most other cases, they are still polarized in their judgments about Jenny*. I have tried to emphasize this point by framing Jenny*’s case as something that Tyler* and the other cannibals use as a counterexample to Jerry*’s theory of meaning.

\textsuperscript{100} Tyler* needn’t deny that there is some explanation of Jenny*’s behaviour, just that this explanation can involve ascriptions of meaning to Jenny*’s utterances.
terms of the first of the two sorts of normativity distinguished by Hattiangadi (norm-relativity). On this suggestion, Jerry* and Tyler* will simply talk past one another in the manner of Hare’s missionaries and cannibals)

In order to meet Baker’s argument, I have shown that there is an analogue of Hare’s argument that does suggest that ‘meaning’ is normative—it just happens that Baker uses an inopportune analogue in his own argument. In the scenario I describe, this problem is resolved by using an appropriate analogue and by focusing on the right kind of constraint. The most plausible treatment of this scenario involved the distinction between what seems right and what is right, which operates as a further constraint on any account of meaning that is more important than that requiring meaning to play an explanatory role. Consequently, this scenario highlights the fact that Baker’s focus on explanatoriness is a red herring. Though an account of meaning does aspire to be explanatory in the way that Baker suggests, this is not the only constraint on any adequate account of meaning. In the case of Jerry* and Tyler*, the two both accept that an account of meaning should aspire to be explanatory but nonetheless disagree as to whether an account of Jenny*’s meaning can abide by the seems right/is right constraint, so that in this case the need to preserve the seems right/is right distinction trumps matters relating to explanation.

§6.4.7 Baker’s Objection to Moral Twin Earth

Baker argues that, like Hare’s case, there is no plausible semantic analogue of Moral Twin Earth that supports the normativity of meaning. We simply do not have analogous intuitions in the meaning case, Baker says, because semantic judgments do not have the same overriding importance that moral judgments do. According to Baker, this feature of moral judgment drives the collective intuition that the Earther and Twin Earther of the Moral Twin Earth scenario use the term ‘good’ with the same meaning and so genuinely disagree in spite of the fact that their applications of that term diverge in the way that they do. Baker claims that we cannot plausibly view ‘meaning’ as being univocal under similar circumstances, e.g. between Jerry and Tyler in his original scenario.

Baker argues that Jerry and Tyler are more plausibly viewed as having a merely verbal dispute insofar as their use of the term ‘meaning’ equivocates between PRIVATE MEANING and PUBLIC MEANING, respectively (2016, p.78). Jerry argues that “the social aspects of language are explained by the linguistic and cognitive dispositions of individuals” and Tyler argues that “individual usage is parasitic on collective practice” (2016, p.78). Both parties have a theory with “grounds to argue that either public meaning or private meaning is a derivative property,” yet nonetheless “it could turn out that a theory positing distinct, explanatorily independent properties of public meaning and private meaning actually provides a more elegant and powerful account of the data” (2016, p.78). Consequently, Baker says, we can view ‘meaning’ as equivocating between PRIVATE MEANING or PUBLIC MEANING given that these
concepts are “explanatorily independent” of one another. In the moral case, Baker claims that ‘good’ (or ‘morally right’) could not be similarly equivocal.

Baker argues that ‘good’ cannot be plausibly viewed as being equivocal in moral contexts. Even if we suppose, for example, that two parties in a dispute view ‘good’ as referring to CONSEQUENTIALIST RIGHT and DEONTOLOGICAL RIGHT respectively, we are more inclined to view one party as wrong rather than treat ‘good’ as ambiguous between the two competing concepts (2016, p.79). Moral Twin Earth is designed to give credence to the view that ‘good’ functions this way; in that scenario, our intuitions suggest that ‘good’ means the same thing in English and Twin English and because of this, we still hear the Earther and Twin Earther as disagreeing in spite of their different uses of that term: “Even when we imagine linguistic communities on two different planets, each using ‘right’ to express one of the proposed precisifications of the term, we still hear disagreement” (2016, p.79). Things cannot go similarly in the meaning case, Baker claims, because we cannot plausibly view judgments about ‘meaning’ as having the same action-guiding force as judgments about ‘good’.

For a suitably-similar ‘Semantic Twin Earth’, Baker argues, we cannot plausibly view Earthers and Twin Earthers as using ‘meaning’ univocally and genuinely disagreeing across radically different views about what meaning consists in, e.g. PRIVATE MEANING or PUBLIC MEANING. Even if we assume that semantic judgment is action-guiding, judgments of this kind could only be weakly action-guiding in the sense that they are readily overridden by other sorts of considerations (2016, p.79). After all, Baker says, this is our intuitive view of semantic judgment here on Earth. We do not typically take our judgments about ‘meaning’ to have overriding importance in deciding what to do, whereas we do take our judgments about ‘good’ to have such importance (2016, p.79). The limit for interpreting ‘meaning’ as univocal in a dispute is considerably lower than it is for ‘good’. Consequently, there is a lower limit on viewing disputes about ‘meaning’ as substantive disagreements, compared to similar disputes about ‘good’. Given that we assign overriding importance to our judgments about ‘good’, there is wider scope for interpreting two speakers as disagreeing, even if they use ‘good’ in radically different ways. Moral Twin Earth is designed to illustrate this unusual feature of our thinking about ‘good’:

The utility of Moral Twin Earth as a thought experiment depends in part on the fact that judgments about ‘right’ are typically treated as overriding by normal earhlings and their counterparts, thus we can imagine without strain significant and noticeable departures in behaviour on Moral Twin Earth. (2016, p.79)

Judgments about ‘meaning’ are simply not overriding in this way, Baker says, so we cannot reasonably expect to hear genuine disagreement in radical disputes about ‘meaning’:

Even if MEANING is normative, we cannot expect similar departures on a Semantic Twin Earth. (2016, p.79)

This suggests, Baker says, that ‘meaning’ does not function like ‘good’ in the sense that the limit for univocity (and so substantive disagreement) is comparatively low—‘meaning’ is more like a
“standard referring expression” and so we do not get results on Semantic Twin Earth similar to those on Moral Twin Earth (2016, p.79).

§6.4.8 Reply to Baker’s Moral Twin Earth Objection

Baker’s focus on the overridingness of moral judgments is misleading. Admittedly, this feature of our moral judging is partly responsible for our intuitions about Moral Twin Earth, but this is only one part of the broader stipulation that ‘good’ plays the same role on Earth and Twin Earth. The main reason that the Earther and Twin Earther are most plausibly viewed as disagreeing in the scenario envisioned by H&T is that each of the parties use their respective terms, the English ‘good’ and Twin English ‘good’, to fulfill a distinctive role in guiding the moral judgments of the two communities. Of course, one of the distinguishing marks of this sort of discourse, shared between the two populations, is that judgments about ‘good’ are overriding relative to other sorts of considerations. Nonetheless, this alone does not carry our intuition that the Earther and Twin Earther genuinely disagree about ‘good’ in spite of the widely different ways they apply that term. What does carry that intuition is the fact that, as I have mentioned, Earthers and Twin Earthers use ‘good’ in their respective languages in the way that they do—to deliberate about what ought to be done and whether various actions are to be commended or permissible, etc. Given that ‘good’ plays this role between the two languages, there is room to view the Earther and Twin Earther as disagreeing with respect to what theory ought to regulate the use of ‘good’, i.e. whether a consequentialist or a deontological theory should play the role of guiding our thinking about what to do. As we know, the Earther and Twin Earther each have robust explanations for their views on this topic. The Earther vouches for his consequentialist theory and the Twin Earther for his deontological theory. Whether or not they assign overriding importance to judgments about ‘good’ is beside the point so long as in each case the degree of importance attached to the relevant judgments is the same on the two planets. We can readily assume, for instance, that the two parties do not assign overriding importance to these judgments without damaging the intuition about disagreement. The same can be said in the case of meaning.

Baker seems to assume that if meaning is normative, then our judgments about meaning must have overriding force over other sorts of considerations; this is simply false. The thesis that meaning is normative is perfectly compatible with the idea that semantic judgments are not overriding over other judgments, such as those about ‘good’. Baker looks to concede this, though he ultimately concludes that because meaning is not normative in this overriding sense that a Semantic Twin Earth cannot succeed. This is too quick on Baker’s part. Even if our judgments about ‘good’ and ‘means’ come apart in that ‘means’ has a lower threshold for univocity and semantic judgments are not overriding, this does not preclude a Semantic Twin Earth scenario from generating intuitions for the normativity of meaning. The main reason for this is that we can assume that, in line with the discussion above, that Semantic Twin Earth involves two parties that use ‘means’ in their respective languages to perform the same role. So even though ‘means’ does
not have overriding force over either party’s other judgments, about ‘good’ etc., this does not have any bearing on the fact that they genuinely disagree in the Semantic Twin Earth scenario.

In the next sections I consider how a robust theory of normative (semantic) judgment could explain our intuitions in cases such as those discussed in the preceding sections. I field two candidates for such a theory. The first is Allan Gibbard’s expressivist theory, advanced in his *Meaning and Normativity* (2012). According to Gibbard, the concept MEANING involves plans for belief and so is fundamentally normative. The second candidate I consider is a realist alternative that takes ‘meaning’ to refer to normative properties, suggested by John McDowell. After some scrutiny, I make a provisional ruling about which account is likely to do best in meeting the obstacles that each of these theories must meet in going forward.

§6.5  Gibbard: Meaning is a Normative Concept

Gibbard (2012) argues that the normativity of meaning derives from the fact that MEANING is a normative concept. According to Gibbard, meaning is essentially normative because certain categorical ‘oughts’ are built into the concept MEANING. This is not to say that the meaning of all terms is normative, but instead that the *meaning of ‘meaning’ is normative*. So, though terms like ‘rock’ and ‘tree’ etc. are clearly non-normative in that they don’t prescribe any categorical ‘oughts’, the term ‘meaning’ is special in the same way that ‘good’ is—both terms express normative concepts, i.e. MEANING and GOOD are normative concepts:

The thesis that “meaning is normative” concerns not meaning itself but the concept of meaning. Equivalently, it concerns the meaning of our term ‘meaning’, or the content of our thoughts of what a person is thinking. Just as the concept GOOD is a normative concept, goes the claim, so is the concept MEANING (2012, p.6).

Gibbard argues that MEANING is a normative concept because, in line with Kripke (1982), if a speaker means something by a term, that seems to have immediate implications for what that speaker should do when he or she uses that term (2012, p.10). In this respect, the slogan that “meaning is normative” might be better put, according to Gibbard, as “*means* implies *ought*” (2012, p.10). As we saw in Chapter One, Kripke’s Wittgenstein (KW) used this assumption about our intuitive notion of meaning to attempt to rule out dispositionalism about meaning as an adequate answer to the sceptical challenge posed in his 1982. The key assumption for Kripke is that it is because I mean *addition* by ‘+’ that I *ought* to answer ‘125’ to the query ‘58 + 67 = ?’. It is in virtue of the meaning that I ascribe to the sign ‘+’ that I ought to answer as I do in this new case, even if in practice I may fail to do so:

From statements saying what I mean by the plus sign and other arithmetic terms and constructions, it will follow that I *ought* to answer “125” when I ask myself “What’s 58 + 67?” Not that I necessarily will answer “125”, but I ought to answer that way (2012, p.10).

Gibbard admits that all of this is rather sketchy, but the terms ‘normative’ and ‘normativity’ are after all not terms that “we learned at our parents’ knees” (2012, p.10). Instead these terms have an at-best “shallow” characterization of being about what we ought to do: “They
are judgments that are, as Wilfrid Sellars put it, “fraught with ought”. They are judgments that move within the “space of reasons” (2012, p.10). These sorts of considerations are characteristic of our talk about ‘good’, and according to Gibbard, about ‘meaning’. So though we only have a rather general and vague idea of the kind of considerations that count as ‘normative’, this is enough for ‘normative’ to be used as a general umbrella term for our talk about ‘ought’ judgments. In the case of ‘meaning’, Gibbard says, there is a “primitive” and “idealized” kind of normative ought that issues from a speaker’s use of the term ‘means’ (2012, p.16).

§6.5.1 Meaning Ascriptions Imply ‘Ought’ Statements

Meaning is normative, Gibbard argues, insofar as meaning ascriptions imply straight ‘ought’ ascriptions that do not derive analytically from purely naturalistic facts (2012, p.11). Gibbard calls this the weak thesis, i.e. the view that \( \text{means} \) implies \( \text{ought} \). As we saw above, something like this thesis motivates KW’s normative constraint on any plausible story about what constitutes meaning facts. The stronger thesis, Gibbard says, accepts both that \( \text{means} \) implies \( \text{ought} \) and, roughly, its converse—“for every means, there is an ought that implies it” (2012, p.12). The latter we can set aside for the moment. For now, let’s consider whether the former—the weak thesis—is plausible.

First, Gibbard explains, it is unclear why we should accept the weak thesis. In particular, it is unclear why we should view meaning as normative simply because, for example, the fact that I mean \( \text{addition} \) by ‘plus’ implies that I \( \text{ought} \) to answer ‘125’ to KW’s sceptic. This example is surely a normative matter in that it has normative implications about what I should do, but this alone does not necessarily reveal anything about the concept of MEANING itself. As Gibbard puts it:

If I mean PLUS by ‘+’, Kripke says, then I should answer 125. But what does this show? If it’s going to rain I should carry an umbrella, but that doesn’t make the concept of rain a normative concept. Not everything with normative import is itself normative. Mightn’t the tie of meaning PLUS to answering 125 be like the tie of rain to taking an umbrella? Oughts, in that case, aren’t built into the concept of meaning PLUS, any more than they are built into the concept of rain. (2012, p.12)

If we drop the thesis that \( \text{means} \) implies \( \text{ought} \), can we explain these two examples? Given that the concept of meaning PLUS and the concept RAIN do not seem to necessarily imply any straight ‘ought’ ascriptions, there must be some alternative to Gibbard’s weak normativity thesis. Gibbard imagines how such an alternative might go:

---

101 Gibbard is careful to qualify that not all statements involving the term ‘ought’ are genuinely normative; some statements of this sort are simply degenerate (2012, p.11). For example, Gibbard explains, the claim “You ought to do whatever you ought to do” is a normative statement, albeit a degenerate one: “Any statement entails this one vacuously—and obviously, this fact doesn’t render all statements normative” (2012, p.11). The \( \text{ought} \) involved in genuine normative statements is a straight ascription that cannot derive analytically from “the purely naturalistic layout of things” (2012, p.11). We can roughly define genuine normative or ‘ought’ statements, Gibbard says, as “those thoughts or sentences that are ought-infused in a way that keeps them from following analytically from anything purely naturalistic” (2012, p.11). Gibbard claims that meaning is normative on this shallow characterization.
It does tend to be a good thing to have an umbrella when it rains, but that’s because an umbrella can keep you dry and—usually—you ought to keep dry. It’s also a good thing not to say or think that $58 + 67 = 5$, but mightn’t that too just be a matter of consequences? If you do things like answering 5, banks dishonor your checks, buildings you design fall down, and all sorts of other bad things tend to happen. Or perhaps it’s also a matter of intrinsic value: to accept an answer 5 would be to believe falsely, and it is desirable to believe the truth. (2012, p.12-13)

Gibbard argues that we need to adopt a normativist view of meaning because we must explain why certain ‘oughts’ follow from meaning. For example, Gibbard says, we need a thesis that explains why I ought to answer ‘125’ rather than ‘5’ to KW’s sceptic if I use ‘plus’ to mean addition and not quaddition (2012, p.13). The best way to explain this unusual quality of meaning is via an expressivist view akin to Ayer’s emotivism or Hare’s universal prescriptivism, according to Gibbard (2012, p.19).

Gibbard argues that the concept MEANING is normative insofar as it essentially involves states of planning, where Gibbard uses the term ‘plan’ in a somewhat technical way, to refer to plans for belief as opposed to, strictly-speaking, plans for action. On Gibbard’s view, to accept that a word has a meaning involves a plan to accept and reject sentences in one’s language. Gibbard claims that we can elucidate the concept MEANING in this “oblique” way by explicating the sort of planning state that a speaker is in when he or she takes an expression to have a certain meaning, e.g. that ‘Dogs bark’ means DOGS BARK. As Gibbard puts it:

A sentence of mine that means DOGS BARK is ‘Dogs bark.’ The small caps convention, after all, is that ‘DOGS BARK’ just is whatever the sentence ‘Dogs bark’ means in my mouth (2012, p.179).

Gibbard realizes that this construal of meaning, as it stands, is not robust enough. If we construe meaning in terms of a speaker’s plan for the use of an expression, it is unclear how we could capture shared meaning across languages or across, say, that same speaker in the past or future:

The problem now is to extend this to other languages, beyond the sentences of my own language right now. What of Pierre, who speaks French but no English? What of myself at another time, when we can at least raise the question of whether my meanings have changed? Pierre’s believing DOGS BARK consists in accepting a sentence in his French that means DOGS BARK, and I am convinced that ‘Les chiens aboient’ is such a sentence. What does it mean, though, I am now asking, to say that his sentence ‘Les chiens aboient,’ means DOGS BARK? (2012, p.179)

We can address this worry, Gibbard says, by further explicating the “state of mind that believing a meaning claim amounts to” (2012, p.179). In this case, for instance, we can cite the state of mind that a speaker would be in if he or she has the belief that Pierre’s sentence ‘Les chiens aboient,’ means DOGS BARK?

§6.5.2 Synonymy

Gibbard argues that we can make progress towards such an account by citing the state of mind that a speaker is in when he or she takes two expressions to be synonymous, e.g. the state of mind of a speaker with the belief that ‘Les chiens aboient’ and ‘Dogs bark’ are synonymous.
If a speaker believes two sentences are synonymous, on Gibbard’s view, then that speaker takes those sentences to consist in the *same plan for belief*, i.e. they consist in the same plan for their use under various circumstances (whether that be to accept or reject them by Gibbard’s terminology). By extension, if two speakers mean the same thing by a sentence, then they will have the same plan for accepting and rejecting those sentences. Pierre and Gibbard, for instance, will mean the same thing by ‘Les chiens aboient’ and ‘Dogs bark’ insofar as they plan to accept or reject those sentences under the same conditions. So given the same circumstances, Pierre and Gibbard will use the sentences in their respective languages in the same way: Pierre will accept his sentence whenever Gibbard does, Gibbard will reject his sentence whenever Pierre does, and so on. A similar point can be made for other sentences in Pierre and Gibbard’s languages: “The epistemic circumstances for which I myself ought to accept ‘Here’s a dog’ are those for which Pierre ought to accept ‘Voici un chien’ . . . This is what it means to say that Pierre’s sentence ‘Voici un chien’ is synonymous with my sentence ‘Here’s a dog’” (2012, p.180). Gibbard puts this more formally:

1. For epistemic circumstances E, I ought to accept my sentence ‘Here’s a dog’.
2. For epistemic circumstances E, Pierre ought to accept his sentence ‘Voici un chien’.

(2012, p.180)

If Pierre and Gibbard uniformly agree in their plans for using those sentences for any circumstances E, then we can conclude that the two mean the same thing; both use their sentences to express *HERE’S A DOG*.

This is only an initial sketch of Gibbard’s expressivist account of meaning and there is a lot of detail that I am putting to one side for the purposes of my discussion here. The key takeaway for the purposes of this chapter is there is an expressivist analysis that respects the normativity of meaning. For Gibbard, the normativity of meaning is simply built into the concept MEANING. The best way to characterize this feature of meaning is to say that MEANING essentially involves some OUGHT beliefs, or plans, that regulate speaker behaviour. Pierre’s belief that ‘Voici un chien’ *means* HERE’S A DOG is an OUGHT belief in the sense that Pierre’s meaning determines how he ought to go on, to apply ‘Voici un chien’ only in circumstances where there is a dog. In the next section, I cast some doubt on this expressivist account of meaning.

§6.5.3 Meaning Non-Factualism and scepticism

I argue that Gibbard’s version of non-factualism about meaning is unlikely to provide a plausible vehicle for capturing the normativity of meaning. The main reason for this, beyond any of the paradigmatic problems that non-factualism faces in other domains (e.g. the Frege-Geach problem\(^\text{102}\)), is that non-factualism about meaning has a special problem in addressing the sceptical line of argument proposed in Kripke (1982) and discussed in Chapter One. Gibbard’s

\(^{102}\) See Geach (1965).
position may readily accommodate the idea that meaning is normative, but this is all for nothing if it cannot face the more general problem at the heart of the sceptical argument. I argue that, in the end, Gibbard’s position is unlikely to solve this problem and so we may be better off in opting for an alternative account of meaning that takes meaning ascriptions to perform some fact-stating role.

In the next sections, I first introduce the basic scheme of this more general objection to non-factualism about meaning by briefly explaining how it motivates a recent argument by Anandi Hattiangadi. I then argue that this objection can be delivered in a more straightforward way by following the strategy deployed in Miller (2010a). Once the strategy of this general objection to non-factualism about meaning is clear, I explain how Gibbard’s view from Meaning and Normativity (2012), as a kind of non-factualism about meaning, is under serious pressure to meet this objection. If this objection is not met, then Gibbard’s non-factualist view is forced toward the radical sceptical conclusion that all language is meaningless—something that any account of meaning ought to avoid.

§6.5.4 Hattiangadi: The Limits of Expressivism

In her recent paper “The Limits of Expressivism” (2015), Anandi Hattiangadi argues against Gibbard (2012) that non-factualism about meaning is both unprofitable and self-defeating in its treatment of the intentionality of language and thought—the power of language “to represent, to be about, or to stand for something or other” (2015, p.225). According to Hattiangadi, this feature of language essentially involves semantic properties such as “meaning, reference, content, extension, intension, truth conditions, and the like” (2015, p.225). On a factualist view about meaning, i.e. that views meaning ascriptions as performing some fact-stating role, this feature can be readily explained. For example, Hattiangadi says:

Suppose you sincerely utter the sentence ‘Nehru was the first Prime Minister of India’. This sentence of your language means that Nehru was the first Prime Minister of India, and the belief that ensures your sincerity has the content that Nehru was the first Prime Minister of India. Your belief and your assertion are true iff Nehru was the first Prime Minister of India. The name ‘Nehru’ and its mental analogue the concept NEHRU refer to Nehru. These are all examples of semantic properties of language and thought. (2015, p.225)

Hattiangadi argues that non-factualism cannot explain the intentionality of language by using, as Gibbard suggests, the idea that there are states of affairs that accord or fail to accord with a plan. This move is blocked, Hattiangadi says, because a non-factualist cannot take the notion of a plan for granted: in assuming that there are states of affairs that accord or fail to accord with a plan we would be assuming the existence of the very sort of semantic or intentional fact whose existence is denied by the non-factualist. This objection can be given in a more straightforward way, without getting into the details of Gibbard’s view.
§6.5.5 Meaning Scepticism

Recall that KW’s sceptical argument is designed to show that there is are no facts in virtue of which ascriptions of meaning, such as Jones means magpie by ‘magpie’, are true or false. KW motivates this argument by envisioning a sceptic who claims that there is no fact that would determine what I meant in using the ‘plus’ sign in the past and as such, nothing determines that ‘125’ would be the only answer I ought to give to when prompted with ‘68 + 57 = ?’, where ‘68 + 57 = ?’ is an arithmetic operation I have never performed before. KW’s sceptic hypothesizes that by ‘plus’, I did not mean addition but instead a non-standard function quaddition, where quaddition can be defined as the following function:

\[ x \oplus y = \begin{cases} x + y & \text{if } x, y < 57 \\ 5 & \text{otherwise} \end{cases} \]

As such the sceptic claims that I ought to answer ‘5’ to the query ‘68 + 57 = ?’. According to him, there is no fact that would secure ‘125’ rather than ‘5’ as the answer that I ought to give in this case: “no fact about my past history—nothing that was ever in my mind, or in my external behaviour—establishes that I meant plus rather than quas” (Kripke 1982, p.13). Given that there is no such fact about my past usage, as to whether I meant plus or quas in the past, there is nothing in my past history that dictates ‘125’ rather than ‘5’ as the answer I ought to give to the current query ‘68 + 57 = ?’. In other words, I have no justification for answering one way rather than the other. This challenge, the sceptic says, can be applied more generally against any of my other words (including words from all of the main syntactic categories).

The sceptic’s challenge is apparently radical and absurd, but what fact could there be that would determine that ‘125’ is the unique answer I ought to give when prompted with ‘68 + 57 = ?’? Any adequate answer to KW’s sceptic must meet two constraints:

1. It must offer a candidate fact that would constitute my meaning plus and not quas.
2. It must explain how that fact justifies my giving the answer ‘125’ rather than the answer ‘5’. (Miller 2007, p.168)

The first constraint is driven by the idea that the meaning of a term dictates a kind of (weak) normative standard for the correct use of that term, i.e. and so whether various applications of that term are correct or incorrect. Intuitively, the term ‘plus’ is used correctly if applied to the function addition, and incorrect if applied to any other function, including quaddition. As such, any adequate account of my meaning must specify addition (and only addition) as regulating my past use of the term ‘plus’ and thus rule out the sceptical hypothesis that I meant quaddition all along.

The second constraint is driven by the idea that any fact about my past meaning must imply that ‘125’ is also the only answer I ought to give to the query ‘68 +57 = ?’.

KW enumerates a series of candidate facts that would secure my past meaning, but as we saw in Chapter One, each apparently fails to meet at least one of the above constraints. This leads KW to conclude that there are no facts that would secure a determinate meaning for my past self, for ‘plus’ or indeed for any other term. Consequently, KW says, the notion of meaning seems to
“vanish into thin air” in the wake of the sceptic’s challenge. Nonetheless, KW argues that we can revive the notion of meaning by means of a non-factualist ‘sceptical solution’.\(^\text{103}\)

§6.5.6 Non-Factualism and the Sceptical Problem

KW claims that we can avoid the “intolerable” conclusion of the sceptic’s challenge by opting for a ‘sceptical solution’ that concedes the “sceptic’s negative assertions are unanswerable” and yet revives the notion of meaning by supposing that ascriptions of meaning perform some non-fact-stating role (Miller 2010a, p.5). We can get clear on how this would go in the meaning case by looking at a basic non-factualist theory in the moral case. For example:

Suppose that for one reason or another you have reached the conclusion that there are no such things as moral facts, so no facts in virtue of which claims such as “Stealing is wrong” can be true or false. Then, you might worry that moral thought and talk is liable to be convicted of a massive and systematic error: since moral claims purport to describe moral facts, and there are no such facts, the conclusion that all (positive, atomic) moral judgments are false appears to follow, threatening us with a form of moral nihilism according to which the notions of right and wrong “vanish into thin air”. (2010a, p.5)

Emotivism, a basic kind of non-factualist moral view, tries to avoid this radical conclusion by abandoning the claim that the function of moral claims is to state moral facts (2010a, p.5). The emotivist argues that moral claims instead perform under a non-fact-stating role of expressing emotion, e.g. of expressing a speaker’s feeling of disapprobation towards a kind of action, such as stealing. The most famous proponent of this view was A.J. Ayer:

If I say to someone, “You acted wrongly in stealing that money”, I am not stating anything more than if I had simply said, “You stole that money” In adding that this action is wrong, I am not making any further statement about it. I am simply evincing my moral disapproval about it. It is as if I had said, “You stole that money”, in a peculiar tone of horror, or written with the addition of some special exclamation marks. The tone, or the exclamation marks, adds nothing to the literal meaning of the sentence. It merely serves to show that the expression of it is attended by certain feelings in the speaker … If I now generalize my previous statement and say, “Stealing money is wrong”, I produce a sentence which has no factual meaning - that is, expresses no proposition that can be either true or false. (Miller 2010a, p.5)

The sceptical solution outlined by KW takes an analogous approach to the case about meaning. The main idea behind KW’s solution is that we can treat ascriptions of meaning as performing some non-fact-stating role, e.g. of expressing some degree of confidence that, say, Smith’s use of “+” will accord with the use that expression has in our speech community, “or perhaps as marking our acceptance of Smith as a member of that community” (2010a, p.5). This, KW argues, avoids the spiral into meaning nihilism from the sceptical argument because we can revive the notion of meaning on non-factualist grounds, where meaning ascription plays a function that is non-fact-

\(^{103}\) Most but not all commentators view the sceptical solution as non-factualist. Wright (1984), McGinn (1984), Blackburn (1984), and Boghossian (1989) view the sceptical solution as non-factualist, whereas Wilson (1994) views the solution as factualist. For a critique of Wilson’s interpretation of the sceptical solution, see Miller (2010b).
stating and yet can make sense of our judgments concerning a speaker’s going right and wrong with respect to his use of a word.¹⁰⁴

Arguably the kind of non-factualism proposed in KW’s sceptical solution begs the question against the sceptical argument: “The ascription of a non-fact-stating function to a type of statement still presupposes that there is a distinction between correct and incorrect uses of statements of that type” (2010a, p.6). For example, Miller explains:

. . . suppose that statements of the form “X is bad” have the function of expressing moral disapproval of X, while statements of the form “X is yummy” have the function of expressing a desire to eat X. Suppose that I intend to express moral disapproval of George Bush, and I utter, “George Bush is yummy”. Then we can say that this use of the sentence is incorrect, unlike “George Bush is bad”, which, modulo my intention, would have been correct. In other words, a serious non-factualist account of the semantic function of a type of statement requires a distinction between correct and incorrect uses of statements of that type, on pain of collapsing into a form of nihilism about the type of statement in question (according to which utterances of statements of that type would be mere noise or mere “sounding off”, and hence not assessable in terms of correctness and incorrectness). (2010a, p.6)

Even if non-factualism may have some purchase in explaining the function of moral judgments, it faces a special difficulty in the case of meaning. Non-factualism must account for the platitude that there is a distinction between the correct and incorrect use of a meaningful word or expression, even if semantic judgments (ascriptions of meaning) are non fact-stating. In the moral case, we can drop the assumption that moral language is descriptive and still make sense of moral judgments as playing a non-factual role and expressing non-cognitive states, e.g. emotions or sentiments. In the meaning case, this cannot be done because we cannot drop the idea that ascriptions of meaning involve standards for correctness.

A consequence of this is that the sceptical argument applies to both factualist and non-factualist semantic functions and so meaning non-factualism cannot escape nihilism in the same way that moral non-factualism attempts to avoid moral nihilism:

. . . the sceptical argument targets the notion of a rule per se, not just the notion of a rule governing the use of expressions with descriptive semantic functions: even non fact-stating language is rule-governed, and hence susceptible to the argument of the rule-following sceptic. (2010a, p.6)

This argument can be applied against non-factualism more generally by following the emotivist model of moral judgment glossed above:

Suppose for the sake of argument that we hold an emotivist account of judgments about goodness according to which the semantic function of “X is good” is to express moral approval of X. Suppose at time t I intend to express moral approval of Fidel. Then, it

¹⁰⁴ KW expands this view into a communitarian account of rule-following and meaning. According to KW, we can only make sense of a speaker’s following a rule or speaking meaningfully insofar as he or she is part of a community of rule-followers or speakers. On KW’s view, the community provides the requisite check on the speaker’s accordance with a rule or meaningful use of an expression in the sense that, in light of that community’s background of use, we can judge whether the speaker in question does right or wrong by it (in that he or she either accords or fails to accord with the rule or the meaning of an expression as dictated by the community’s use).
seems, I ought to utter “Fidel is good”. However, according to the rule-following sceptic, I should not utter “Fidel is good” because the rule-governing “good” is not

R1: Utter “X is good” if and only if you intend to express moral approval of X

But rather

R2: Utter “X is good” if and only if (a) it is time t* < t and you intend to express moral approval of X or (b) it is time t** ≥ t and you intend to express moral disapproval of X. (2010a, p.7)

If R1 governs the use of ‘good’ then the utterance “Fidel is good” is correct at time t; if R2 governs the use of ‘good’ then the utterance “Fidel is good” is incorrect at time t (2010, p.7). In other words, “one might say: according to R1, ‘good’ expresses moral approval, whereas according to R2, ‘good’ expresses moral schnapproval” (2010a, p.7).

Suppose that a bizarre sceptic challenges whether my use of ‘good’ is governed by R2 rather than R1, that really I mean moral schnapproval rather than approval by my use of ‘good’. According to this sceptic, no fact can determine that my use of ‘good’ is governed by R1 rather than R2. In this case, non-factualism is no better off in that it must simply presuppose that there are some predetermined facts governing correctness and incorrectness that would secure my meaning R1 rather than R2 as the rule governing the use of the expression:

. . . any form of non-factualism that is more sophisticated than e.g. a very crude form of emotivism that views moral judgments as mere “sounding off” presupposes determinate facts about the semantic functions of linguistic expressions, or the rules governing their correct use, irrespective of whether those functions are conceived to be descriptive or non-descriptive or whether the rules govern description or some non-descriptive linguistic practice. A generalized version of KW’s argument establishes that there are no such facts. (2010a, p.7)

Any form of non-factualism that simply assumes there are some such correctness standards for the expressions of a language does not provide a legitimate response to the challenge posed by KW’s sceptic (2010a, p.7).

The important point for our present purposes is that the preceding objection applies also to non-factualist accounts that characterize the function of meaning ascription in terms of prescriptions, i.e. as orders or commands (2010a, p.7). For example:

. . . suppose that I intend to get Jones to open the window at time t. So I utter “Open the window!” Intuitively, this utterance is correct, modulo my intention to get Jones to open the window. But the sceptic can argue that it is incorrect, since actually “window”, as I use it, means windows before time t but doors at time t and thereafter. (In fact, “window” actually means qwindow …). Thus despite the fact that “Open the window!” is an imperative rather than a description, it presupposes facts about the meanings of its constituent expressions. (2010a, p.7-8)

Consequently, any version of non-factualism that presupposes that there are standards governing correct use for various expressions will fall to KW’s sceptical argument. In other words, any sceptical solution along the non-factualist lines sketched by KW will fail as a response to the sceptical argument about meaning.
§6.5.7 Why Gibbard’s Sceptical Solution Cannot Succeed

I will now argue that Gibbard’s theory of meaning, as a version of meaning non-factualism, is equally susceptible to the sceptical line of argument delivered in the second chapter of Kripke (1982) and revived in Miller (2010a) against the sceptical solution suggested in Kripke’s third chapter. The key reason for this is that by bringing in the notion of a semantic plan, Gibbard begs the question by assuming that there are facts about which uses of expressions do and do not accord with the dictates of the plan. By using this notion, Gibbard smuggles the idea that one can go right or wrong in the sense that one can fulfil or fail to fulfil a plan for using an expression. Consequently, Gibbard’s account of meaning is no improvement on the factualist alternatives, i.e. ‘straight solutions’, to KW’s sceptical challenge in that his account must also presuppose that there are some facts that determine when various expressions are correctly and incorrectly applied. This point is made especially clear by re-applying the sceptical line of argument to a scenario where a speaker’s plan for using an expression (in line with Gibbard’s view) is the centre of the sceptic’s attention. When we apply the sceptical challenge in this scenario, it turns out that the notion of a plan has no more traction than any of the alternatives outlined by Miller. Gibbard’s view does no better in addressing the sceptical challenge because we can draw the sceptical challenge again at the level of a speaker’s plan. As noted above, the only way that a non-factualist view like Gibbard’s can escape here is by presupposing that there are some predetermined facts that fix correctness standards for meaning. However, this move is blocked for any account of meaning that is truly non-factualist. As such, there is no clear way to adopt non-factualism and at the same time meet the requirements for addressing the sceptical argument proposed by KW. This can be shown by using the following example.

Suppose that I have a plan to use the term ‘magpie’ in accordance with the meaning I take that word to have, as a term that is applied correctly only insofar as it is applied to magpies and not finches, sparrows, or anything else.105 From time to time, I might fail to execute this plan. For instance, I might mistakenly use the term ‘magpie’ to describe a seagull—perhaps it is dusk and I am not wearing my glasses. By the lights of my plan for using ‘magpie’ in accordance with its meaning, I have made a mistake in that I have used that term to apply to something other than a magpie. Most of the time though, I do abide by my semantic plan and use the term ‘magpie’ accordingly. In fact, I believe that the term ‘magpie’ possesses a meaning that corresponds to this plan for applying the term, i.e. to magpies and nothing else. Insofar as I have this plan for the use of the term ‘magpie’, I use that term meaningfully: the plan is what governs my use of the expression, and determines which uses are correct and incorrect.

Imagine that a stranger takes a seat next to me while I am enjoying a view of the Dunedin harbour. The stranger says nothing at first. He and I simply relax and observe the birds circling

105 The specific details of the plans are simplified here and are not necessarily identical to the plans deployed by Gibbard—but this simplification makes no difference to the plausibility of my argument.
the harbour. After some time, let’s say time t, I notice a magpie among the seagulls that
dominantly populate the harbour area. I point out this unusual sight to the stranger by gesturing
at the bird and by saying to him “Hey, that is a magpie!” in a tone of surprise. The stranger, who
reveals himself to be a studied philosopher, poses a challenge to me in response to my statement;
he says that the term “magpie” as I use it is governed by a deviant rule that dictates that I utter
“magpie” to refer to magpies before time t and seagulls after time t. More formally, he argues that
R2 rather than R1 dictates the correct application of the expression ‘magpie’ in this case:

R1: Utter “x is a magpie!” if and only if x is a magpie.
R2: Utter “x is a magpie!” if and only if (a) it is time t* < t and x is a magpie or (b) it is
time t** ≥ t and X is a seagull.

If R2 is the rule that governs my use of the term ‘magpie’, then my utterance “That is a magpie!”
was incorrect insofar as I was speaking and gesturing at the unusual bird, a magpie and not a
seagull, at time t. According to the stranger, my utterance would have been correct (modulo the
intention to conform to R2) if applied to one of the many seagulls that dominantly populate the
harbour. Minimally, the stranger argues, there is no fact that could determine that my use of the
term ‘magpie’ is governed by R1 rather than R2.

In order to meet this sceptical challenge from the stranger, I concede that indeed there are
no facts that could determine that my use of the term ‘magpie’ is governed by R1 rather than R2.
Nonetheless, I argue that I was correct to use ‘magpie’ as I did insofar as I have a plan for using
that term in line with R1 and not R2 or any other plan. So in response to the stranger, I specify
my plan for using the term ‘magpie’ in accordance with the meaning I take that term to have, in
line with R1 and as a term that is only applied correctly if used to refer to magpies and no other
birds (or anything else for that matter). Given that I have this plan for the use of the term ‘magpie’,
I argue in response to the stranger that I used the term ‘magpie’ correctly in referring to the
unusual bird that appeared over the harbour at time t. I insist that this plan dictates that my
utterance “There’s a magpie” was correct under the circumstances. The stranger argues that this
is not enough. He demonstrates this by generalizing his sceptical challenge to target my plan for
using the term ‘magpie’.

The stranger recalibrates his sceptical challenge and suggests that my plan to apply
‘magpie’ only to magpies consists in P2 rather than P1, where P1 and P2 are the following
semantic plans:

P1: Utter “x is a magpie!” if and only if x is a magpie.
P2: Utter “x is a magpie!” if and only if (a) it is time t* < t and x is a magpie or (b) it is
time t** ≥ t and x is a seagull.

At this stage, the stranger argues that there is nothing that could specify on purely non-factualist
grounds that P1 rather than P2 regulates my behaviour for using ‘magpie’ and that uttering “That
is a magpie!” at time t was correct, modulo the intention to comply with P1.
Under this pressure from the stranger, it is unclear what kind of answer I could offer other than to insist that ‘magpie’ possesses a meaning that dictates that that expression is only correctly applied to magpies and not seagulls, or anything else. Invariably, this involves assuming that ‘magpie’ already has a meaning that is specified in terms of some fact about how that term is correctly and incorrectly applied. So given the sceptical considerations raised in Miller’s argument, it seems that Gibbard’s account of meaning can do no better here. The notion of a semantic plan for applying an expression must, at its core, assume that the expression in question already has a determinate meaning insofar as it assumes that there are facts about which uses of the expression do and do not accord with the dictates of the plan.

The main problem with Gibbard’s construal of the concept MEANING is that in using the notion of a semantic plan, Gibbard helps himself to the notion of accord and effectively ignores the sceptical problem about the determination of reference, embodied in the challenge posed by KW’s sceptic. The problem with this move on Gibbard’s part is that his position is supposed to offer a solution to this challenge in the sense that it is supposed to explain what meaning consists in and part of that is explaining how reference is determined, e.g. what makes it the case that I mean *addition* rather than *quaddition* by ‘plus’. Gibbard does not have an answer to this sceptical challenge, but instead assumes that the notion of a plan includes the idea that one can accord or fail to accord with a plan, i.e. in terms of whether a plan is executed or not. This is not an adequate answer to the challenge posed by KW’s sceptic insofar as the success or failure of a speaker’s attempt to execute a plan is ultimately measured by relying on predetermined facts about what uses of an expression do and do not accord with a plan. As a non-factualist account of meaning, this move is both self-defeating and ultimately no more successful than the factualist alternatives.106

Overall, then, Gibbard’s non-factualist account of ascriptions of meaning as expressing ‘plan-laden’ states, does not provide us with an adequate vehicle for capturing the intuitions elicited by the Meaning Twin Earth scenario.

§6.6 McDowell: Non-Reductionism about Meaning

In the next sections I motivate McDowell’s non-reductionist account of meaning as a vehicle for the normativity of meaning and hence for a normativist explanation of the Meaning Twin Earth scenario. I introduce McDowell’s view by detailing his response to KW’s sceptical argument in his 1984 and 1992 essays on the topic and then expand on this by appealing to McDowell’s remarks on meaning and understanding in *Mind and World* (1994). I motivate one objection against McDowell, raised by Fodor (1995), and then consider whether McDowell’s non-reductionism does any better than Gibbard’s non-factualism. In concluding, I argue that, at least

106 For similar arguments to the effect that meaning non-factualism is in some sense incoherent or self-defeating, see Wright (1984), Boghossian (1989), and especially Boghossian “The Status of Content” (1990).
provisionally, McDowell’s view fares better in that the most serious objection for non-factualism is a knock-down objection whereas the most serious objection for McDowell’s non-reductive factualism is a challenge, albeit a formidable one.

§6.6.1 Wittgenstein’s ‘Straight Solution’ and the Master Thesis

McDowell argues that the sceptical argument at the heart of Kripke (1982) rests on an assumption that Wittgenstein rightly rejects. The crux of this is that Kripke’s Wittgenstein (KW) mistakenly accepts the conclusion that there are no meaning facts in the face of the sceptical paradox posed by Wittgenstein in *Investigations* §201:

This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here. (Wittgenstein 1953/2009)

According to McDowell, Wittgenstein rejects the sceptical conclusion of the paradox. This is evidenced by Wittgenstein’s proposal in the second paragraph of §201:

It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it. What this shews is that there is a way of grasping a rule which is not an interpretation, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases. (1953/2009)

McDowell points out that, contrary to Kripke’s reading, Wittgenstein here proposes a ‘straight solution’ to the paradox, one that specifies a fault in the reasoning that led to the paradox. Kripke neglects this second paragraph of §201 and reads Wittgenstein as embracing the sceptical conclusion and subsequently, offering a ‘sceptical solution’. This misconstrues the point of §201, according to McDowell:

. . . what Wittgenstein clearly claims, in the second paragraph of §201, is that the reasoning [that yields the sceptical paradox] is vitiated by “a misunderstanding”. The right response to the paradox, Wittgenstein in effect tells us, is not to accept it but to correct the misunderstanding on which it depends: that is, to realize “that there is a way of grasping a rule which is not an interpretation.” (McDowell 1984, p.331)

The paradox of §201, McDowell says, is the first horn of a dilemma generated by the mistaken idea that grasping a rule is always an interpretation (1984 p.331). McDowell attributes this ‘misunderstanding’ to the acceptance of what he calls the ‘master thesis’ in his 1992 paper “Meaning and Intentionality in Wittgenstein’s Later Philosophy”:

. . . the thesis that whatever a person has in her mind, it is only by virtue of being interpreted in one of various possible ways that it can impose a sorting of extra-mental items into those that accord with it and those that do not. (1992, p.45)

McDowell argues that Kripke’s tacit acceptance of this thesis drives his embrace of the sceptical paradox, i.e. the first horn of the dilemma. Kripke assumes that a mental representation only has a content (and so can sort items into those that accord and fail to accord with it) insofar as that mental representation is given an interpretation.
Kripke’s reading of how the regress of interpretations threatens the very idea of understanding turns on this thesis: ‘no matter what is in my mind at a given time, I am free in the future to interpret it in different ways’ [Kripke 1982: 107]. This presupposes that whatever is in a person’s mind at any time, it needs interpretation if it is to sort items outside the mind into those that are in accord with it and those that are not. There are always other possible interpretations, and a different interpretation, imposing a different sorting, may be adopted at a different time. Considered in themselves, that is, in abstraction from any interpretations, things in the mind just ‘stand there’. (1992, p.44)

The master thesis assumes that we can draw a distinction between a mental state and its content, something that requires interpretation in order to have a content at all according to the master thesis. Hattiangadi (2007) teases these two apart by drawing a contrast between a kind of ‘vehicle’ and its content:

The vehicle is thought to be a non-intentional state—such as a brain state, or a dispositional state. The vehicle is as devoid of intentional content as a stone, or an uninterpreted word. In order for the vehicle to have a content it needs to be interpreted. Without an interpretation, we cannot make sense of the idea of the vehicle applying correctly or incorrectly.” (2007, p.168)

On this thesis, interpreting a rule R is a matter of applying a further rule R* that tells what R requires in a given situation. So if we accept the master thesis and thus the idea that grasping a rule always involves interpretation, it turns out that there are no facts about the correct and incorrect application of the rule. As such, the first horn of the dilemma (generated by the master thesis) threatens to dissolve the notion of meaning into thin air. In the absence of some fact that would constitute “my having given some expression an interpretation with which only certain uses of it would conform,” the notion of meaning begins to appear an illusion (McDowell 1984, p.331-2).

The other horn of the dilemma is a “familiar mythology of meaning and understanding,” that meaning is like a sign that is not even potentially subject to interpretation:

[the idea that] coming to mean the expression in the way I do (my ‘grasping the rule’) must be my arriving at an interpretation; but it must be an interpretation that is not susceptible to the movement of thought in the sceptical line of reasoning—not such as to content us only until we think of another interpretation standing behind it. (1984, p.332)

McDowell cites Wittgenstein as rejecting this idea:

What one wants to say is: “Every sign is capable of interpretation; but the meaning mustn’t be capable of interpretation. It is the last interpretation.” (Wittgenstein, Blue Book, p.34)

Wittgenstein claims that ‘what one wants to say’ is based on a tacit acceptance of the master thesis, which is taken to an equally intolerable extreme at the other end of the spectrum in the idea that

---

107 There is room for the non-factualist to explain why the notion of meaning is not in danger of nihilism, but the most serious challenge for this view, as I have argued, is that meaning non-factualism looks committed to the self-defeating assumption that there are predetermined facts determining which uses of expressions are correct or incorrect(irrespective of whether those expressions themselves have a non-factualist semantics). Consequently, it is hard to see how non-factualism can avoid the intolerable conclusion that meaning is illusory without begging the question. McDowell has his own reservations to this effect, which he raises in §3 and §10 of McDowell 1984, p.328-30, 345-7.
following a rule is akin to the operation of a *rigid mechanism*. This is the second horn of the dilemma:

How queer: It looks as if a physical (mechanical) form of guidance could misfire and let in something unforeseen, but not a rule! As if a rule were, so to speak, the only reliable form of guidance. (Wittgenstein, *Zettel* §296)

Wittgenstein attacks this ‘mechanical’ picture of meaning and understanding as well, though this attack is, contrary to Kripke’s reading, not an argument for the acceptance of the sceptical conclusion (impaling oneself on the first horn of the dilemma). Given that Kripke accepts the master thesis, he assumes that the dilemma is compulsory, that we must choose between admitting the intolerable conclusion of the sceptical paradox or accept the “fantastic mythology” that there is a kind of final sign-post, a *last interpretation*:

[Kripke would be right about Wittgenstein] if the dilemma were compulsory; but the point of the second paragraph of §201 is precisely that it is not . . . the attack on the mythology is not support for the paradox, but rather constitutes, in conjunction with the fact that the paradox is intolerable, an argument against the misunderstanding. (McDowell 1984, p.332)

For McDowell’s Wittgenstein, the dilemma is not compulsory; he sees a course between the two horns. According to McDowell, Wittgenstein introduces the dilemma to mock, in a rather dramatic way, the misunderstanding at its root—the idea that understanding a rule or the meaning of an expression is always an interpretation.

§6.6.2 Steering the Middle Course: Introducing the Notion of a Practice

If we abandon the master thesis, then as Wittgenstein suggests in the second paragraph of §201, this opens up a way of understanding meaning that both avoids the paradox and the absurd conception of meaning as consisting in a sign that cannot be itself interpreted; there is a way of grasping a rule that is not an interpretation. Wittgenstein gestures at this solution in §201:

And hence also ‘obeying a rule’ is a practice. And to *think* one is obeying a rule is not to obey a rule. Hence it not possible to obey a rule ‘privately’: otherwise thinking one was obeying a rule would be the same thing as obeying it. (1953/2009, p.87)

For Wittgenstein, the notion of a practice or a custom enables an escape from the dilemma. The makings of this escape, McDowell says, are already present in §198:

“Then can whatever I do be brought into accord with the rule?” - Let me ask this: what has the expression of a rule - say a sign-post - got to do with my actions? What sort of connexion is there here? - Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it.

“But that is only to give a causal connexion: to tell how it has come about that we go by the sign-post; not what this going-by-the-sign really consists in.” - On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom. (1953/2009, p.86)

Wittgenstein introduces the analogy with reading a sign-post in §198, McDowell says, “in order to adumbrate the diagnosis that he is going to state more explicitly in §201”: 152
When I follow a sign-post, the connection between it and my action is not mediated by an interpretation of sign-posts that I acquired when I was trained in their use. I simply act how I was trained to.\textsuperscript{108} (1984, p.339)

This, McDowell claims, prompts an objection analogous to the one advanced by the interlocutor in §198—something like:

‘Nothing in what you have said shows that what you have described is a case of following a rule; you have only told us how to give a causal explanation of certain bits of (what might tell as well be for all that you have said) mere behaviour.’ (1984, p.339)

Wittgenstein’s answer to this objection, McDowell argues, corresponds to the first sentence of §201:

\textit{. . . the training in question is initiation into a custom. If it were not that, then the account of the connection between sign-post and action would indeed look like an account of nothing more than brute movement and its causal explanation; our picture would not contain the materials to entitle us to speak of following (going by) a sign-post.} (1984, p.339)

By introducing the notion of a custom Wittgenstein is able to, in one fell swoop, block the regress of interpretations before it begins and rule out the conception of meaning and understanding as “nothing more than brute movement and its causal explanation,” i.e. the conclusion that there is no such thing as following a rule or being guided by one’s understanding of a word. McDowell articulates a more nuanced version of this reply on Wittgenstein’s behalf in his 1992:

When one follows an ordinary sign-post, one is not acting on an interpretation. That gives an overly cerebral cast to such routine behaviour. Ordinary cases of following a sign-post involve simply acting in the way that comes naturally to one in such circumstances, in consequence of some training that one underwent in one’s upbringing. (Compare §506: “The absent-minded man who at the order ‘Right turn!’ turns left, and then, clutching his forehead, says ‘Oh! right turn’ and does a right turn.—What has struck him? An interpretation?”). (1992, p.50)

If we overstate the role of interpretation, McDowell explains, then we force our intuitive conception of understanding into danger by ‘overbalancing’ the notion of grasping a rule toward something like a brute causal mechanism:

\textit{. . . if we give this corrective to an over-mentalizing of the behaviour, perpetrated by giving the concept of interpretation an unwarranted role in our conception of it, we run the risk that we shall be taken to overbalance in the opposite direction, into under-mentalizing the behaviour-adopting a picture in which notions like that of accord cannot be in play, because the behaviour is understood as nothing but the outcome of a causal mechanism set up by the training. Such a picture might fit an acquired automatism, in which there is no question of acting on an understanding of the sign-post’s instructions at all. This risk, that if we exploit the concept of training to exorcize the idea of interpretation, we shall lose our entitlement to the idea of understanding as well, is averted by adding another bit of common sense, that the training is initiation into a custom.} (1992, p.50)

\textsuperscript{108} That is, you do not need some further rule R that allows you to work out what the sign-post says.
As McDowell recognizes, we can only protect the notions of meaning and understanding by abandoning the master thesis and thus the idea that the grasp of an expression’s meaning is always up to an interpretation.

The upshot of McDowell’s response to Kripke’s reading of Wittgenstein is that Kripke makes a serious mistake in locating Wittgenstein on the first horn of the dilemma, as embracing the conclusion that there are no facts about meaning. As we have seen, Wittgenstein views both horns of the dilemma as unacceptable—the paradox and the rigid mechanism—and suggests an escape by way of jettisoning the view that got us into the dilemma in the first place, the ‘master thesis’ and the idea that understanding can be assimilated to interpretation. Once we are rid of this misunderstanding, a middle course between the horns of the dilemma becomes possible: There is a way of grasping a rule that is not an interpretation and the notion of a custom provides this.  

§6.6.3 Second Nature and ‘Bildung’

McDowell argues for a non-reductive factualist view of meaning, according to which there are facts about meaning that are essentially normative. In light of this normative character, McDowell argues that such facts cannot be reduced to natural facts as they are standardly conceived under the restrictive purview of what McDowell calls “bald naturalism”, according to which natural facts are understood on a narrow scientistic view to include only the subject matter of the empirical sciences. McDowell rejects the standard interpretation of naturalism. According to him, we should expand naturalism, with some inspiration from Aristotle and Wittgenstein, to include both natural facts as standardly conceived plus content beyond the scope of the empirical sciences—in particular, moral facts and meaning facts at the level of the space of reasons (McDowell 1994, p.109). According to McDowell, facts of this sort should be admitted by a proper naturalist view in that they are ‘second nature’ to rational beings like us with a distinctive capacity for reasoning and reflection. McDowell takes natural facts (standardly conceived) to

---

109 Some readers of Wittgenstein, such as Malcolm (1986), take the notion of a custom as a constructive philosophical thesis, one that answers the question “How is meaning possible?” (1992, p.49). McDowell argues that this is a misreading:

. . . there is no reason to credit Wittgenstein with any sympathy for this style of philosophy. When he says “What has to be accepted, the given, is—so one could say—forms of life” (p.226), his point is not to adumbrate a philosophical response, on such lines, to supposedly good questions about the possibility of meaning and understanding, or intentionality generally, but to remind us of something we take in the proper way only after we are equipped to see that such questions are based on a mistake (1992, p.50-51).

Wittgenstein’s uses the concept of a custom not as a positive thesis but instead as a tool for halting the regress of interpretations before it begins—to show that the regress is based on a misunderstanding:

What made the notion of accord seem problematic was the regress of interpretations, and the first move in the passage [of §198], the appeal to training, has ensured that we need not begin on the regress of interpretations. The point of the appeal to custom is just to make sure that that first move is not misunderstood in such a way as to eliminate accord, and with it understanding, altogether (1992, p.50).
constitute only part of our natural world—what he calls the ‘realm of law’. The move toward a less restrictive naturalism is necessary, McDowell says, because an important part of our natural history as rational beings cannot be understood from within the realm of law. The primary example for this is our intuitive notions of meaning and understanding, which have a “spontaneity” that fits them outside the realm of law. For this reason, McDowell locates these notions under a distinctive part of our natural world, the ‘space of reasons’.

In order to forestall the worry that his view is committed to an untenable kind of supernaturalism about meaning, McDowell qualifies that his view does not simply equate nature with the realm of law. Instead, McDowell’s view widens the conception of nature to include both the realm of law and the space of reasons. According to McDowell, this move enables him to escape the charge about supernaturalism and deny the idea that meaning facts are part of the realm of law. McDowell calls this view a ‘naturalism of second nature’ (Miller 2013, p.253).

McDowell argues that meaning facts are a special species of facts about reasons for action. According to him, we develop an appreciation for facts of this sort by having a certain upbringing and undergoing a certain kind of education—what McDowell calls a process of ‘Bildung’. By bringing in the process of Bildung, McDowell argues that we can treat meaning facts as ‘natural’ in spite of their residing in the space of reasons rather than the realm of law (Miller 2013 p.254). Moreover, McDowell claims that construing meaning facts as the result of this process enables him to avoid the charge that his non-naturalism is disrespectful of the tenets of natural science (Miller 2013 p.254).

The initiation into a custom which avoids the master thesis is an example of Bildung that enables our access to normative facts in the space of reasons. So, by introducing the notion of a custom and of Bildung, McDowell is able to both reject the master thesis and maintain the view that there are facts about meaning, facts that we have access to via our second nature and Bildung and that are essentially normative in kind. Though this account looks especially promising, especially in light of the apparent failure of Gibbard’s non-factualist account of meaning, McDowell’s theory must face a formidable challenge: How can we account for our access to the normative facts of the space of reasons if they are facts of a distinct kind (that is, aside from saying they are a natural part of Bildung)?

§6.6.4 Fodor’s Objection

In response to McDowell’s analogous account for non-naturalism about moral properties, Jerry Fodor objects that McDowell’s view requires an epistemology that he simply cannot provide.

Having situated . . . the ethical . . . outside the realm of law, McDowell needs to face the embarrassing question how, by any natural process, do we ever manage to get at it? (Fodor 1995, p.11)

The best that McDowell can do to assuage this worry is by “naturalizing” his moral epistemology to something that human beings develop in the course of their upbringing, i.e. via the process of Bildung:
The picture is that ethics involves requirements of reason that are there whether we know it or not, and our eyes are opened to them by acquisition of 'practical wisdom'. (1994, p.79)

The rational demands of ethics are not alien to the contingencies of our life as human beings. Even though it is not supposed that we could explain the relevant idea of demandingness in terms of independently intelligible facts about human beings, still ordinary upbringing can shape the actions and thoughts of human beings in a way that brings these demands into view. (1994, p.83)

The result of this process of education is our second nature, “[a] conception of nature that includes a capacity to resonate to the structure of the space of reasons” and thereby to moral properties (1994, p.109).

Fodor argues that McDowell’s use of metaphor is unhelpful and still leaves the nature of moral epistemology mysterious:

‘Bringing into view’ is a metaphor; only what is in Nature [on the usual characterization] can literally be viewed. And ‘resonating’ is also just a metaphor; only what is in Nature [on the usual characterization] can be literally attuned to. (1995, p.11)

Again, the best that McDowell can do to meet Fodor’s worry is to insist that human beings realize what he calls their ‘distinctive potential’ for exercising ‘practical wisdom’ in the normal course of their upbringing. According to McDowell, there is nothing mysterious in our upbringing that brings about this awareness of the demands of morality and so likewise there is nothing ‘occult’ in the idea that there is a distinctive space of reasons that we have access to via this matured capacity. McDowell attempts to clarify how this process goes by drawing an analogy with our capacity to learn language:

This transformation [the realization of the ‘distinctive potential’] risks looking mysterious. But we can take it in our stride if, in our conception of the Bildung that is a central element in the natural maturation of human beings, we give pride of place to language. In being initiated into a language, a human being is introduced into something that already embodies putatively rational linkages between concepts, putatively constitutive of the layout of the space of reasons, before she comes on the scene. This is a picture of initiation into the space of reasons as an already going concern; there is no problem about how something describable could emancipate a human individual from a merely animal mode of living into being a full-fledged subject, open to the world. A mere animal, moved only by the sorts of things that move mere animals, could not single-handedly emancipate itself into possession of understanding. (McDowell 1994, p.125)

Arguably, the analogy with language-learning does not get McDowell out of trouble. First, McDowell helps himself to the idea that human beings have a ‘distinctive potential’ for accessing the space of reasons. McDowell does not help his case by claiming that our responsiveness to moral reasons is a responsiveness we have in virtue of some ‘second nature’ of ours, external to the realm of laws, and then explaining how we develop this ‘second nature’ capacity:

If one found the idea that we can responsive to moral reasons underexplained when neither the relevant facts about reasons nor the capacities allegedly responsive to them are conceived of as belonging to the realm of law, it helps little to be told that acquiring the relevant capacities is a way of realizing our ‘distinctive potential’. To say all of that is simply to paraphrase the claim that human beings are capable of making ethical judgments: it is not something that can even begin to help us understand what ethical
judgments are or how it might be possible for us to engage in a practice of making them. (Miller 2013, p.256)

Second, the analogy with language-learning is likewise unhelpful:

We are trying to understand how it could be that human beings are capable of speaking intelligibly where this is conceived of as a matter of acquiring a capacity to be responsive to facts about meaning, where neither the capacity nor the facts about meaning can be rendered intelligible from the viewpoint of the realm of law. Does it help to be reminded that humans, but not, say gerbils, can acquire the capacity to converse intelligibly? Of course not.110 (Miller 2013, p.256)

McDowell does not alleviate the ‘fog of mystery’ that surrounds the transformation from ‘distinctive potential’ to ‘awareness of moral reasons’ by using the language analogy. For this reason, McDowell’s notions of Bildung and ‘second nature’ are no improvement on the non-naturalist theories of yesteryear, such as that proposed by G.E. Moore in his Principia Ethica (2013, p.256).111 Given that they don’t work in the moral case, they are unlikely to work in the case of the normative meaning properties either.

§6.7 Review and Provisional Conclusion

In this chapter, I fielded two potential candidates for a robust account of the normativity of meaning. The first candidate was Allan Gibbard’s non-factualist view of meaning, which took the concept MEANING to be normative insofar as it is plan-laden, i.e. consists in states of planning. The principal appeal of this non-factualist view is that it can readily make sense of the idea that meaning ascription is normative by appealing to the non-cognitive ‘planning states’ of speakers. In spite of this appeal, non-factualism about meaning faces a serious challenge from a line of objection advanced in Hattiangadi (2015) and Miller (2010a). According to this objection, non-factualism looks committed to supposing that, on pain of being self-defeating and begging the question, there are determinate facts about the meanings of expressions according to which the use of those expressions are either correct or incorrect even if those expressions are afforded a non-factualist semantics.

The second candidate I considered for an account of normative judgment was John McDowell’s non-reductive factualist view about meaning. According to McDowell, there are facts about meaning that are both essentially normative and cannot be reduced to natural facts (under the standard interpretation). The inspiration for this view is Wittgenstein’s reflections on rule-following and meaning in his Philosophical Investigations. McDowell’s view escapes the line of objection posed against Gibbard’s non-factualist view, though McDowell’s view has a problem of its own, pointed out by Jerry Fodor. Fodor argues that McDowell cannot provide an

110 Miller attributes this reply to McCulloch 1996, who seconds Fodor’s argument that McDowell needs an epistemology for the moral facts of ‘second nature’ that he cannot provide.
111 Moreover, there is a certain circularity in McDowell’s argument: McDowell uses the language analogy to defend the notion of ‘second nature’ in the moral case, but referring to ‘second nature’ in the moral case will not help assuage any worries about its cogency in the case of language and meaning.
adequate epistemology for how we could access irreducible meaning facts, at the level of what McDowell called the ‘space of reasons’.

I suggest that, though McDowell is under pressure to provide an adequate epistemology for his view of meaning facts, this problem is less serious than the one that Gibbard faces with respect to non-factualism. In Gibbard’s case, there is no clear escape from the challenge that non-factualism smuggles in the notion of accord by assuming that there are predetermined facts about the meaningful use of language. In McDowell’s case, McDowell can at least argue that our commitment to the standard ‘bald’ naturalist metaphysics is unreasonable, given that the notions of meaning and understanding simply cannot be captured by a narrow scientistic view of content. Moreover, McDowell can argue that his view does not need to provide a robust epistemology with respect to our access to meaning facts. McDowell may simply dig his heels in here and insist that we are perfectly familiar with the sort of facts that ground ascriptions of meaning and so Fodor and other critics are demanding more than McDowell needs to provide on his quietist view.\footnote{There are additional challenges for non-reductionism about meaning and content that I do not address here. For example, Boghossian (1989) outlines some further problems for McDowell’s non-reductionist view:}

It is sometimes said that an anti-reductionist conception is too facile a response to the problem about meaning. It is hard not to sympathize with this sentiment. But if the considerations canvassed against the alternatives are correct, and if it is true that the ‘rule-following’ considerations leave an anti-reductionist conception untouched, it is hard, ultimately, also to agree with it. Meaning properties appear to be neither eliminable, nor reducible. Perhaps it is time that we learned to live with that fact. (1989, p.548)

Admittedly, the claim that non-reductionist factualism is a better vehicle for normativism than non-factualism would require a full defence in a separate thesis that would no doubt be longer than the present one. Outlining this defence is a task for another day.

\footnote{Boghossian explains that these problems for McDowell’s view echo those from Wittgenstein: Though see his remarks—which I am afraid I do not understand—on a ‘linguistic community [that] is conceived as bound together, not by a match in mere externals (facts accessible to just anyone), but by a meeting of minds’. McDowell's problems here echo, I think, Wittgenstein's own. The main difficulty confronting a would-be interpreter of Wittgenstein is how to reconcile his rejection of substantive constitutive accounts—especially of meaning, see Zettel #16: ‘The mistake is to say that there is anything that meaning something consists’—with the obvious constitutive and transcendental pretensions of the rule-following considerations. It is fashionable to soft-pedal the rejection of constitutive questions, representing it as displaying a mere 'distrust' on Wittgenstein’s part. But this ignores the fact that the rejection of analyses and necessary and sufficient conditions is tied to extremely important first-order theses about meaning, including, most centrally, the family-resemblance view of concepts. (1989, p.544 fn66)}
Conclusion

In the course of the dissertation I have attempted to provide some new evidence in favour of the claim that meaning is normative—specifically, for the claim that semantic judgments or ascriptions of meaning are action-guiding.

In Chapter One, I outlined KW’s sceptical argument that there are no meaning facts in virtue of which ascriptions of meaning, such as ‘Molly means blackbird by ‘blackbird’’ are true or false. One of the crucial components necessary for this argument to succeed with the kind of generality that KW intended it to have (against any putative candidate for a meaning fact), is the assumption that meaning is normative—or what we called the ‘normativity constraint’.

In Chapter Two, I explained how KW’s main argument against dispositionalist accounts of meaning can be viewed as an application of Moore’s open question argument to the case of meaning. However, construed this way, the argument failed because modern reductive dispositionalist accounts of meaning do not accept the principle that meaning facts are reducible to dispositional facts as a matter of conceptual or analytic necessity. Proponents of dispositionalism about meaning view the reduction of meaning facts to dispositional facts as being an *a posteriori* or synthetic reduction on par with the identification of water with H₂O. Nonetheless I argued that a revised open question argument based on an analogue of Horgan and Timmons’ Moral Twin Earth argument does apply against reductive dispositionalist accounts of meaning. According to Horgan and Timmons (H&T), our linguistic intuitions about a Moral Twin Earth scenario suggest that synthetic ethical naturalism is false. Moral terms such as ‘good’ or ‘morally right’ do not possess a meaning determined by the natural properties that causally regulate their use (contrary to the tenets of contemporary forms of synthetic ethical naturalism). The best explanation of this result, according to H&T, involves the idea that moral terms are essentially normative. I argued that a suitably similar Meaning Twin Earth scenario yields analogous results, and that arguably these are best explained by the normativity of meaning.

In Chapter Three, I outlined an objection to H&T’s Moral Twin Earth argument from Plunkett and Sundell (2013). According to Plunkett and Sundell, H&T’s argument crucially turns on the thesis that speakers who genuinely disagree must mean the same thing by the terms used in communicating that disagreement. This is strictly-speaking not true. Some genuine disagreements are non-canonical disputes—disputes in which the conflicting contents of a disagreement are expressed non-literally, e.g. by implicature. So, one way for a naturalist to escape the Moral Twin Earth argument would be to view the Moral Twin Earth scenario as involving a *metalinguistic negotiation*, a special kind of non-canonical dispute where the parties involved use terms with different meanings precisely because they disagree about what those
terms ought to mean. I argued that the Meaning Twin Earth is a limiting case for this method of explaining disagreement. That is, Meaning Twin Earth cannot be plausibly viewed as being about a metalinguistic negotiation, and therefore, Plunkett and Sundell’s strategy cannot be used to block the Meaning Twin Earth argument.

In Chapter Four, I discussed two objections advanced by David Copp (2000) against the Moral Twin Earth argument. In his first reply, Copp argued that we can view the Earther and Twin Earther of the Moral Twin Earth scenario as meaning roughly the same thing and as genuinely disagreeing, even if they ultimately mean different things in the philosophically preferred sense of referring to the same properties. I argued that that this first reply is unconvincing for two reasons. First, the reply involved a disposal of the general principle that meaning determines reference, something that would seriously limit the options for giving an account of meaning, and in particular, for capturing the compositionality of meaning and reference. Second, the reply does not meet the requirements for actually viewing the Earther and Twin Earther as disagreeing, but only as if they disagree. In his second reply, Copp objected that Horgan and Timmons fail to respect the details of a proper realist semantic theory. On a more complete Putnamian moral semantics, Copp argued, a naturalist can vindicate the intuitions that Horgan and Timmons elicit from the Moral Twin Earth scenario. I argued that even if this argument can successfully meet the Moral Twin Earth argument, an analogous argument cannot be used against my Meaning Twin Earth argument without begging the question.

In Chapter Five, I outlined three objections against the Moral Twin Earth argument advanced by Merli (2002). In his first objection, Merli argued that the Moral Twin Earth argument depends on an underdescription of the Moral Twin Earth scenario. Once these details are suitably clarified, it turned out that the Moral Twin Earth argument is faulty. I argued that, in this first objection, Merli makes a crucial misstep by viewing the ‘Earth’ of the Moral Twin Earth scenario as actual Earth. Given that Horgan and Timmons are using a thought experiment, they can subtly alter the details to meet this objection. In his second objection, Merli argued that naturalistic moral realism’s prospects for preserving the univocity of moral terms are good if it addresses the issue holistically, i.e. by appealing to realist solutions to similar problems in moral epistemology. The main reason that the realist can be optimistic about preserving univocity, Merli claimed, is that we have good reason to expect a convergence in moral theory and so disputes like the one showcased in the Moral Twin Earth scenario are likely to result in convergence. So, according to Merli, the realist can respond to the Moral Twin Earth argument by treating the Moral Earther and Moral Twin Earther as referring to the properties that their moral communities would both refer to at the end of inquiry. I argued that this second objection is unconvincing because we cannot plausibly view some radical disagreements as trending toward convergence. I used an example of a political dispute between Putnam and Nozick to bring this point out. Merli also suggested that we should defer to idealized moralizers in these cases. I argued that this is likewise no help to the realist’s case given that we can only speculate about what these idealized thinkers
would judge. In his third objection, Merli argued that we can explain the disagreement between the Moral Earther and Moral Twin Earther as a practical disagreement about what to do and that we can explain this disagreement by using an expressivist view of all-in endorsement. I argued that this objection fails because the dispute between the Earther and Twin Earther cannot be viewed as a practical disagreement and that bringing in expressivism to explain this disagreement leaves moral realism either unmotivated or unable to meet the Moral Twin Earth argument. Either way, Merli’s objection cannot be used to damage the Moral Twin Earth argument or, by proxy, the Meaning Twin Earth argument.

In Chapter Six, I considered two broad options for a normativist account of meaning that would explain the intuitions elicited from the Meaning Twin Earth scenario: an expressivist form (Gibbard) and a rationalist form (McDowell). In order to motivate how these two forms of normativism could go, I looked at how an account of normative judgment could be developed in the moral case by looking at Smith (1994) and Hare (1952). I argued against Baker (2016) that an argument for the claim that meaning is normative can be mounted, as Gibbard (2012) suggests, from an analogue of Hare’s ‘missionary’ scenario. I outlined, in brief, Gibbard’s expressivist/non-factualist account of meaning and McDowell’s non-reductive factualist alternative. I concluded by arguing, provisionally, that McDowell’s view faces a less serious challenge than Gibbard’s view in going forward. Gibbard’s view faces a knock-down objection, whereas McDowell’s view faces a challenge, albeit a formidable one.

If successful, the Meaning Twin Earth argument has two important consequences: First, it shows that a revised open question argument might pose a threat to modern reductive dispositionalist accounts of meaning (even if the classical version of that argument fails). Second, it provides some new evidence to think that meaning is normative and thus constitutes a challenge to the anti-normativist to explain the intuitions elicited from the Meaning Twin Earth scenario without invoking the normativity of meaning.

It is important to be clear about some of the limitations of my argument. Firstly, I have attempted to provide some new evidence for the claim that meaning is normative and that ascriptions of meaning are action-guiding. However, I have not attempted to respond to the many serious challenges faced by the idea that meaning is normative. We saw in Chapter One that the claim that meaning is normative could be cashed out as the claim that ‘correct’ is a normative or evaluative term, or that there are specifically semantic prescriptions governing the use of meaningful expressions. However, there are formidable arguments against both of these claims in the literature, including especially Boghossian (2005), Glüer (1999), Glüer and Wikforss (2009), Hattiangadi (2006, 2007, 2009), Miller (2010c, 2012), Verheggen (2011), and Wikforss (2001). In a full defence of the claim that meaning is normative, these arguments would need to be met.

Secondly, although I have discussed and attempted to defuse some of the standard objections to the Moral Twin Earth argument itself, as well as its application in the case of
meaning, this discussion has been far from exhaustive. The Moral Twin Earth argument has generated a huge literature, and among the discussions that a full treatment of the issues would need to cover are Dowell (2016), Freiman (2014), Geirsson (2005, 2014), Gert (2006), Horgan and Timmons (2000a, 2015), Rubin (2008, 2014a, 2014b, 2015a, 2015b), van Roojen (2006), and Sonderholm (2013). Dealing with these issues is a task for future work.

Thirdly, although I have attempted to provide new evidence in favour of the claim that ascriptions of meaning are action-guiding, I have not attempted to provide a detailed account of what the action-guiding nature of ascriptions of meaning consists in. In Chapter Six we looked at two broad ways in which this might be done. As we saw, an expressivist account of meaning ascription along the lines of Gibbard’s arguably falls prey to a generalised version of KW’s sceptical argument. However, a non-reductive normative realist view of meaning such as that defended by John McDowell may fare better. Non-reductive normative realist views face considerable challenges: How do we account for our epistemic access to normative facts and properties? And how we assuage naturalistic worries about their inclusion in our metaphysics? Arguably, though, these are challenges rather than knock-down arguments. In my view, the same cannot be said for the objection facing Gibbard’s expressivist view.

Overall, then, this thesis aspires to make a modest contribution to the case in favour of the claim that meaning is normative, and there is no shortage of serious challenges that will need to be addressed in future work.
Appendix: Frege on Meaning, Reference, and Compositionality

In Chapter Four, we discussed two objections advanced by David Copp (2000) against the Moral Twin Earth argument developed by Horgan and Timmons (1992a). In his first of these two objections, Copp argues that the Earther and Twin Earther of the Moral Twin Earth scenario can be viewed for most practical purposes as genuinely disagreeing even if strictly-speaking they do not mean the same thing, i.e. given that moral and twin-moral terms refer to distinct properties. I argued in response to this objection that this would involve an unpalatable disposal of the principle that the meaning determines reference, something that any adequate account of meaning should aim to preserve. In this brief appendix, I expand on some of the reasons that this general principle is worth preserving by looking at Frege’s account of sense and his introduction of the principle that meaning/sense determines reference/semantic value.

Frege: Sense and Semantic Value

Frege had principled reasons for introducing the notion of sense and the principle that sense determines semantic value. First and foremost, Frege introduces sense as a semantic property over and above the property of semantic value, i.e. the property possessed by an expression that determines whether sentences in which that expression appears are true or false (Miller 2007, p.23). Frege does this in order to deal with three problems that arise if we assume that meaning can simply be identified with semantic value.

Bearerless Names

First, if meaning is identified with semantic value, then we are forced to conclude that bearerless names—names that do not refer to existing persons or objects—are meaningless. This we can call the ‘problem of bearerless names’. Take for example the sentence ‘Harry Potter is a wizard’. Harry Potter is a fictional character and so the name ‘Harry Potter’ does not refer to an actual person. The expression ‘Harry Potter’ is thus a bearerless name. According to Frege, the semantic value of a name is its bearer—the object or person that that name refers to (Miller 2007, pp.12-13). As such, the name ‘Harry Potter’ cannot have a semantic value given that ‘Harry Potter’ has no bearer. Moreover, according to Frege, “the semantic value of a sentence is determined by the semantic value of its parts” (Miller 2007, p.24). Consequently, the sentence ‘Harry Potter is a wizard’ likewise has no semantic value. So if semantic value is the only property that determines meaning, then we would be forced to say that the sentence ‘Harry Potter is a wizard’ is meaningless. Intuitively though, the sentence is perfectly intelligible, like many similar sentences involving bearerless names, e.g. ‘Atticus Finch is a lawyer’, etc. For Frege, this suggests that there
must be some other semantic property that can account for the meaning of bearerless names, in addition to semantic value. As Miller puts it:

. . . names without a reference (semantic value) are not meaningless; so there must be some other semantic property possessed by names in addition to having a reference (semantic value). (2007, p.25)

This additional property is that of possessing a sense. Two further problems point us in this direction, according to Frege.

Substitution and Belief Contexts

A second problem that Frege points out is that semantic value alone cannot make sense of cases in which co-referential terms, such as ‘Clark Kent’ and ‘Superman’, are substituted in belief contexts. This we can call the ‘problem of substitution into belief contexts’. Let’s use an example to bring the problem into clear view: Suppose that Lex Luthor, Superman’s arch nemesis, does not know that Clark Kent is Superman. In other words, Lex Luthor does not know that ‘Clark Kent’ and ‘Superman’ refer to the same person. Lex Luthor is nonetheless very familiar with Superman as the pesky superhero who always foils his plans, can leap buildings in a single bound, etc.. We can safely accept the sentence ‘Lex Luthor believes that Superman is Superman’ as being true. This much is uncontroversial of course, unless Lex has some bizarre understanding of identity (2007, p.25).

Remember that, according to Frege, the semantic value of a name is its bearer. Given that ‘Clark Kent’ and ‘Superman’ have the same bearer (are co-referential), they must have the same semantic value. If we combine this with Frege’s two basic principles about the compositionality of semantic value, i.e. that (1) “the semantic value of a complex expression is determined by the semantic value of its parts” and (2) the “substitution of a constituent of a sentence with another which has the same semantic value will leave the semantic value (i.e. truth-value) of the sentence unchanged,” then we should be able to substitute ‘Clark Kent’ for ‘Superman’ and vice versa without changing the semantic value of any sentence in which those names appear (2007, p.11-12). Let’s see if we can do this in the case of sentences about Lex Luthor’s beliefs.

If we substitute ‘Clark Kent’ for ‘Superman’ in the sentence “Lex Luthor believes that Superman is Superman”, we end up in trouble—for this substitution yields ‘Lex Luthor believes that Clark Kent is Superman.’ If semantic value is identified with meaning and ‘Superman’ and ‘Clark Kent’ are co-referential, then we should have no change in semantic value—we should end up with a sentence that possesses the same truth-value as “Lex Luthor believes that Superman is Superman.” That is, the two sentences should have the same semantic value given that their parts have the same semantic value. But this does not seem to be the case. By stipulation, Lex Luthor does not know that Clark Kent is Superman, so the sentence ‘Lex Luthor believes that Clark Kent is Superman’ is false. Lex Luthor does not have this information; after all, things would go much better for him if he did have this information about the pesky Superman! This example suggests, like the preceding case, that the notion of semantic value alone is insufficient for capturing our
intuitive notion of meaning. In this case, the problem is that a theory of meaning utilizing only
the property of semantic value cannot make sense of how two terms that are co-referential could
make a difference when substituted in belief contexts. Intuitively, there must be some additional
property to semantic value in terms of which we can understand the different contributions of co-
referential names to this sort of case. Before we introduce the property of sense to play that role,
let’s look at one final problem that results from identifying meaning with semantic value.

Informativeness

A third problem that arises from the assumption that meaning can be identified with
semantic value can be called the ‘problem of informativeness’. In order to bring this problem out,
we need to first gloss how the notion of understanding is related to meaning. One way to do this
is as follows: “When someone understands an expression, we say that he knows its meaning:
meaning is that semantic property of an expression which someone with
an understanding of that
expression grasps” (2007, pp.26-7). If we combine this intuitive view of understanding with the
assumption that meaning is identical with semantic value, then we end up in trouble. This can be
demonstrated by returning to Frege’s classic example about the planet Venus:

It took an empirical discovery in astronomy to discover that this planet was both the
celestial object known as the Evening Star and also the celestial object known as the
Morning Star. Consider the state of a competent language speaker before this empirical
discovery (or of a competent language speaker after the discovery who is unaware of it).
Such a person understands the identity statement “The Morning Star is the Evening Star”,
even though they do not know its truth-value. Frege’s point is that if meaning were
identified with semantic value, this would be impossible. (2007, p.27)

As Frege argues, it seems possible for someone to understand the sentence ‘The Morning Star is
the Evening Star’ without knowing that ‘The Morning Star’ and ‘The Evening Star’ have the same
semantic value in that they refer to the same celestial body. If we assume that meaning is identified
with semantic value, this would be impossible because grasping the meaning of those two
expressions would amount to grasping their shared semantic value. This seems wrong because it
is clearly possible for someone to understand the sentence ‘The Morning Star is the Evening Star’
without knowing that that sentence is true. As such, Frege argues (again) that the semantic value
of an expression cannot be the only component of its meaning.

Introducing the Notion of Sense

In order to address the preceding problems outlined against the assumption that meaning
is identical with semantic value, Frege introduces the notion of sense. Frege defines sense as a
semantic property with a few key features, the most important of which is that “The sense of an
expression is that ingredient of its meaning which determines its semantic value” (Miller 2007,
p.28). Consequently, there is a property in addition to semantic value that determines what
semantic value an expression has: “Thus, a name has a reference—stands for a particular
object—and also has sense, some means of determining which particular object this is” (Miller
2007, p.28). Note that this feature of sense corresponds to the principle that Copp rejects, namely, the principle that meaning determines reference. Let’s see how the notion of sense can address the three problems we have outlined so far.

Addressing the Problem of Informativeness

If we return to the classic example that Frege describes concerning Venus, the property of sense enables us to explain how a competent speaker can grasp the meaning of the sentence ‘The Morning Star is the Evening Star’ without knowing that ‘The Morning Star’ and ‘The Evening Star’ actually share the same reference (and therefore without knowing that the sentence is true). A speaker can do this by grasping the sense of ‘The Morning Star’ and the sense of ‘The Evening Star’ where we think of these senses as descriptive conditions associated with the names which an object must satisfy in order to be the referent of the relevant names. In this case, the sense of ‘The Evening Star’ could be something along the lines of ‘that object which appears in such and such a place in the sky at such and such times in the evening’ and mutatis mutandis for ‘The Morning Star’ (Miller 2007, p.28). Thus, a speaker can grasp the meaning of ‘The Morning Star’ and ‘The Evening Star’ vis-a-vis the notion of sense insofar as that speaker grasps the conditions under which an object would be the referents of those names. It is possible then that a speaker could fail to know that the relevant object in the sky (Venus) is in fact the referent of both ‘The Morning Star’ and ‘The Evening Star’ and nonetheless understand the meaning of the sentence ‘The Morning Star is the Evening Star’, given that that speaker knows what objects would count as the referent of each of those names. That is, the speaker knows the senses of those expressions. It is possible to know the sense of an expression without knowing its semantic value. This principle enables us to solve the problem of informativeness: The sentence ‘The Morning Star is the Evening Star’ is informative for the speaker in that he can grasp its sense and go on to discover that ‘The Morning Star’ and ‘The Evening Star’ refer to the same celestial object.

Addressing the Problem of Bearerless Names

Introducing the notion of sense also enables us to explain how bearerless names are meaningful. Remember that, on the assumption that semantic value is the only semantic property behind the notion of meaning, the expression ‘Harry Potter is a wizard’ must be meaningless, given that the name ‘Harry Potter’ has no bearer. If we deploy the notion of sense, we can avoid this unintuitive conclusion. Suppose for instance that the sense of ‘Harry Potter’ is given by the description ‘The protagonist of J.K. Rowling’s most popular fiction series and the arch-nemesis of Lord Voldemort.’ Even though there is no actual object that meets that descriptive condition,

---

113 Frege is not committed to the idea that the sense of a name must be given in terms of its description, though this is one effective and at least initially attractive way to take Frege’s point. I follow Dummett again here. See Miller 2007, p.58.

114 A similar line of argument shows how the sentence ‘The Morning Star is the Morning Star’ would be uninformative for the speaker. See Miller 2007, p.46-47.
a competent speaker could still understand what it would be for an object to meet that condition. As such, it is clear that one could grasp the meaning of the sentence ‘Harry Potter is a wizard’ even though Harry Potter is a fictional character and the name has no actual referent. This amounts to accepting the principle that an expression without a semantic value can nonetheless have a sense. And so names that do not have a semantic value, such as ‘Harry Potter’, are not simply meaningless because they can nonetheless have a sense (Miller 2007, pp.30-31).

Addressing the Problem of Substitution in Belief Contexts

Finally, the notion of sense allows us to explain how substituting apparently co-referential names can lead to a change the truth-value of sentences in certain belief contexts. Let’s return to our example about Lex Luthor. Remember that Lex Luthor does not know, by stipulation, that Clark Kent is Superman. That is, he does not know that ‘Clark Kent’ and ‘Superman’ are co-referential terms in that they refer to the same person. In this, we can safely judge that Lex Luthor believes that Superman is Superman or even that Lex Luthor believes that Clark Kent is Clark Kent, but we cannot judge that Lex Luthor believes that Clark Kent is Superman—this is clearly false in that given our description of the scenario. If we stick to the assumption that semantic value is the only property involved in meaning, then we cannot make sense of the change in truth-value in this context between ‘Lex Luthor believes that Superman is Superman’ and ‘Lex Luthor believes that Clark Kent is Superman’. As we know, the semantic value of ‘Clark Kent’ and ‘Superman’ are the same, given that the semantic value of a name is its bearer and the two names share the same bearer. Again, Lex Luthor does not know this; he does not know that the names ‘Clark Kent’ and ‘Superman’ have the same bearer. As such, substituting the former for the latter in ‘Lex Luthor believes that Clark Kent is Superman’ takes us from the a truth to a falsehood. How can the notion of sense help here?

Frege addresses the problem about substitution in belief contexts by introducing the notion of customary sense. Under ordinary circumstances, i.e. outside of belief contexts, Frege admits that ‘Clark Kent’ and ‘Superman’ refer to the same person. Frege is committed to admit this much given his aim to preserve the compositionality of semantic value, which he encapsulates in two principles mentioned above, that (1) “the semantic value of a complex expression is determined by the semantic value of its parts” and (2) the “substitution of a constituent of a sentence with another which has the same semantic value will leave the semantic value (i.e. truth-value) of the sentence unchanged” (Miller 2007, pp.11-12). Frege wants to preserve these two intuitive principles about meaning and so he must say that, for any sentence involving a name such as ‘Clark Kent’ or ‘Superman’, the semantic value (or truth-value) of that sentence will be preserved when a co-referential term is exchanged for that name, e.g. if ‘Superman’ is substituted for ‘Clark Kent’. In belief contexts, things go differently, according to Frege.

For Frege, some terms that are co-referential outside of belief contexts can fail to be co-referential within belief contexts. In these cases, apparently co-referential terms like ‘Clark Kent’
and ‘Superman’ do not refer to the same person, but instead to what Frege calls the ‘customary sense’ or ‘indirect reference’ of those names. The customary sense of ‘Clark Kent’ is given by the description ‘the mild-mannered reporter of the Daily Planet’ and the customary sense of ‘Superman’ is given by the description ‘the pesky superhero who always foils Lex Luthor’s plans and can leap buildings in a single bound’. So though the names ‘Clark Kent’ and ‘Superman’ are co-referential outside of belief contexts, they are not so within those contexts, i.e. provided that the customary senses attached to those names are different. So by introducing the idea that the customary sense of an expression is its indirect reference, Frege can preserve the aforementioned principles about the compositionality of semantic value in spite of apparent counter-examples like the one we have focused on about Lex Luthor’s putative belief that Clark Kent is Superman.

Sense as an ‘Ingredient in Meaning’

We must be clear that Frege’s notion of sense is an ingredient in meaning in addition to semantic value, where ‘ingredient in meaning’ has a semi-technical meaning borrowed from Dummett:

> What we are going to understand as a possible ingredient in meaning will be something which it is plausible to say constitutes part of what someone who understands the word or expression implicitly grasps, and in his grasp of which his understanding in part consists. (Dummett 1973, p.93)

Sense is an epistemic notion insofar as it is that ingredient of an expression’s meaning that constitutes what someone grasps (implicitly or otherwise) whenever they understand the meaning of that expression. The key contrast here with semantic value is that semantic value is not an ingredient of an expression’s meaning (whereas sense is such an ingredient). Consequently, the notion of semantic value is not epistemic in the way that the notion of sense is. We can spell this out as the principle that “[t]he semantic value of an expression is no part of what someone who understands the expression grasps” (2007, p.35). We can see that this aspect of Frege’s view also plays an important role in his solution to the problem of informativeness.

Suppose for a moment that the contrary is true, that semantic value is an ingredient of meaning and is part of what a speaker grasps when they grasp the meaning of an expression. Now if a speaker were to grasp the semantic value of an expression in grasping its meaning, then it would be impossible, for instance, for a speaker to understand the sentence ‘The Morning Star is the Evening Star’ and then go on to make the empirical discovery that ‘The Morning Star’ and ‘The Evening Star’ in fact refer to the same celestial body. If a speaker already grasps the semantic value of both ‘The Morning Star’ and ‘The Evening Star’ in grasping the meaning of those expressions, then he would already know that the Morning Star is indeed the Evening Star and consequently the truth-value of the sentence ‘The Morning Star is the Evening Star’. Thus, dealing with this problem about informativeness moves Frege to rule out semantic value as an ingredient in meaning—as something a speaker must grasp in order to understand the meaning of an expression.
What then is the place of semantic value in the theory of meaning if it is not an ingredient of meaning? Dummett clears this up for us:

To say that reference [semantic value] is not an ingredient in meaning is not to deny that reference [semantic value] is a consequence of meaning, or that the notion of reference [semantic value] has a vital role to play in the general theory of meaning: it is only to say that the understanding which a speaker of a language has of a word in that language . . . can never consist merely in his associating a certain thing with it as its referent [semantic value]; there must be some particular means by which this association is effected, the knowledge of which constitutes his grasp of its sense. (Dummett 1973, p.93)

So as Dummett points out, semantic value still plays a key role with respect to the meaning of an expression even though it is not part of what someone grasps in knowing what an expression means. As Dummett says, there must be some special means by which a speaker can grasp the meaning of an expression and, on Frege’s view, sense fulfills this role.

Consequences for Copp

The main point of this brief detour is not that in giving up the principle that meaning determines reference Copp would be committing himself to giving up Frege’s own conception of sense. Although we would be committing himself to giving this up, he would also be committing himself to giving up any account of meaning which, like Frege’s, ties the compositionality of meaning/sense to the compositionality of reference/semantic value. It is worth nothing that although Putnam himself takes the Twin Earth argument to show that we should jettison the Fregean idea that grasp of meaning is a psychological state, even Putnam (1975) retains the principle that meaning determines reference (or extension).115 So, in giving up the principle that meaning determines reference, Copp would be giving up a very broad set of accounts of meaning, a set that includes not only Frege’s own view of sense but also a rival view like that proposed in Putnam (1975) and obliquely endorsed by Copp himself. The main point in the text is that, prima facie, this is a very high price to pay for a response to the Moral Twin Earth argument.

115 Putnam claims that it would be “preferable to take a different route [to giving up the idea that meaning determines extension] [which] makes it trivially true that meaning determines extension . . . but totally abandons the idea that if there is a difference in meaning my Doppelgänger and I assign to a word, then there must be some difference . . . in our psychological state” (1975, p. 165). So, Putnam actually argues that meaning determines reference, but what's in the head doesn't determine reference, so meanings ain't in the head. Putnam accepts that meaning determines reference, but wants to deny that the grasp of meaning is a psychological state.
References

Publications.

Aristotelian Society*, 28, 63.

*Dialectica*, 70(1), 65-84.

OX, UK; New York, NY, USA: Blackwell.


Philosophy*, 98(392), 507-549.

GAP.5, Fifth International Congress of the Society for Analytical Philosophy*. Bielefeld:
Mentis.

Realism* (pp. 181-228). Ithaca, NY, USA: Cornell University Press.

NY: Cambridge University Press.

1101.

1110.

137.


Frege, Gottlob (1892). On Sinn and Bedeutung. In Michael Beaney (ed.), *The Frege Reader*.
Blackwell, 151-172.


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title and Details</th>
</tr>
</thead>
</table>

171


