Philosophers' Imprint

VOLUME 13, NO. 23
DECEMBER 2013

Disagreement and the Semantics of Normative and Evaluative Terms

David Plunkett
Dartmouth College

Tim Sundell
University of Kentucky

© 2013 David Plunkett & Tim Sundell
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.
<www.philosophersimprint.org/013023/>

Introduction.[1]

In *The Language of Morals*, R.M. Hare introduces what is now a famous case in metaethics. Hare asks us to imagine a missionary who arrives on a distant island inhabited by cannibals. Much like the missionary, these cannibals use the term 'good' as a general term of praise for people and their actions. However, the cannibals apply the term 'good' to different sorts of people and actions than the missionary does.[2] The cannibals are thus surprised by how the missionary applies the term. As Hare writes:

> [T]hey 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

1. Thanks to two anonymous reviewers for *Philosophers' Imprint*, Chris Barker, David Braddon-Mitchell, Josh Brown, Alexi Burgess, Sarah Buss, Fabrizio Cariani, Nate Charlow, Matthew Chrisman, Lenny Clapp, Terence Cuneo, Jorah Dannenberg, Stephen Darwall, Kenny Easwaran, David Ebrey, Andy Egan, David Enoch, Peter Galison, Jon Garthoff, Allan Gibbard, Thony Gillies, Sandy Goldberg, Bob Goodin, Mark Greenberg, Elizabeth Harman, Sally Haslanger, Frank Jackson, Chris Kennedy, John Ku, Brian Leiter, Adam Lerner, Dustin Locke, Don Loeb, Peter Ludlow, Kate Manne, Teresa Marques, Tristram McPherson, Eliot Michaelson, Kristie Miller, Daniel Nolan, Howard Nye, Alejandro Pérez-Carballo, Phillip Pettit, Peter Railton, Mark Richard, Mike Ridge, Paolo Santorio, Karl Schafer, Francois Schroeter, Laura Schroeter, Alex Silk, Sam Shpall, Michael Smith, Nicos Stavrapolous, Isidora Stojanovic, Sharon Street, Kevin Toh, Pekka Väyrynen, and Daniel Wodak. Versions of this paper were presented to audiences at the Australian National University, the 2011 Cerisy Meaning, Context, and Implicit Content Conference, University of Chicago, University of Houston, University of Kentucky, University of Leeds, Logos Research Group at the University of Barcelona, Monash University, University of Melbourne, Northwestern University, Princeton University, University of Sydney, University of Vermont, and Yale University. Thanks to everyone who participated in those sessions for their helpful questions and comments.

2. Throughout this work, we use italics to introduce terminology or for rhetorical stress, single quotes to mention linguistic expressions, small caps to name concepts, and double quotes for quoting other authors, "scare quoting", simultaneous use and mention, and other informal uses.

> accustomed to commend people who are bold and burly and collect more scalps than average.[3]

In introducing Hare's case, we do not endorse the pernicious stereotypes of native peoples that the case draws on. Rather, we introduce it because Hare's highly influential discussion of the case (originally published in 1952) exemplifies a common form of philosophical argument in metaethics — a form of argument that draws on intuitions of disagreement to reach semantic conclusions — that will be our focus in this paper.

One of the core, widely shared intuitions about Hare's case is that there is a disagreement between the missionary and the cannibals when they apply the term 'good' to different people and actions. To bring this out, imagine the following discussion between one of the cannibals and Hare's missionary:

(1) (a) It would be good for me to kill more people and collect more scalps. This is because, in general, it is good to collect as many scalps as possible.

(b) No. It is not good for you to kill more people in order to collect more scalps. Killing people is not good and collecting as many scalps as possible is not good either.

Most of us have the intuition that the exchange consisting of (1a) and (1b) reflects a genuine disagreement between the missionary and the cannibal and, moreover, that this disagreement makes good sense and is worth engaging in.

In light of this intuition, many philosophers, including Hare himself, have drawn the following conclusion: even though the missionary and the cannibal systematically apply the term 'good' to different things, the missionary and the cannibal must nonetheless mean the same thing by the word 'good'. This thought is exemplified in Hare's original discussion of the case. Hare asks us to imagine a theory according to

3. (Hare 1991, 148).

which the missionary and the cannibal did *not* mean the same thing by the word 'good'. (Perhaps because its meaning is determined by how the word is applied, or by what causes speakers to apply it.) Speaking about the latter type of theory, he writes:

> 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".[4]

Hare claims that there *is* a disagreement between the missionary and the cannibal when they apply the term 'good' to such different people. Hence, he argues, the missionary and the cannibal mean the same thing by the word 'good'.[5] Based on these considerations, he concludes in favor of a semantic version of non-cognitivism for the term 'good', a theory according to which the meaning of the term 'good' is understood in terms of the desire-like attitude that it is used to express.

Hare's reasoning — from the fact that the relevant exchange reflects a genuine disagreement to a conclusion in semantics — exemplifies the sort of argument that is our focus in this paper. This is the argument from (a) the premise that an exchange between two speakers expresses a genuine disagreement to (b) the thesis that those speakers mean the same things by the words they use in that exchange. This is a common type of argument in metaethical theorizing about the meanings of our moral or ethical terms. Indeed, it plays a central role quite generally in debates about the meanings of normative and evaluative

4. (Hare 1991, 148–149).

5. For a similar gloss on Hare's discussion, see (Horgan and Timmons 1993) and (Smith 1994).

terms — terms (moral, ethical, aesthetic, epistemological, etc.) whose meaning at least partly involves matters of what one *should* do, think, or feel, or, respectively, about what is *better* or *worse*. In other words, this is a key type of argument not just in metaethics but in metanormative theory more broadly.

We think that the reasoning exemplified in Hare's argument is flawed, and, moreover, that the conclusions in many arguments of this type are in fact false. 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. How is disagreement reflected in such a linguistic exchange? Via a largely tacit negotiation over how *best* to use the relevant words. Following Chris Barker, we use the term *metalinguistic usage* for cases where a linguistic expression is used (*not* mentioned) to communicate information about the appropriate usage of that very expression in context.[6] In the cases that Barker introduces (and which we discuss later in this paper) the metalinguistic usage of a term centers on descriptive matters of how a term is in fact used in the relevant context. However, by putting forward competing claims of this type, speakers can, *via* metalinguistic uses of their terms, debate how it is those terms *should* be employed. We call a dispute like this — one that employs competing metalinguistic usages of an expression, and that reflects a disagreement about the proper deployment of linguistic representations — a *metalinguistic negotiation*.

Two features serve to characterize the class of linguistic disputes that we are interested in — the class of disputes we label 'metalinguistic negotiations'. First, metalinguistic negotiations employ a distinctive communicative mechanism — metalinguistic usage. And, second, they concern a distinctive normative question — how best to use a word relative to a context. This general question — how should we use our words? — shows up in a range of forms. It could, on some particular occasion, be a matter of resolving ambiguity, prescisifying a vague term, setting a contextual parameter, or in any other way determining how some antecedently indeterminate matter of meaning should be settled. Supposing we take the meaning of a word to be the concept it expresses in the context, a dispute about any of these matters reflects in some sense a disagreement about which among some set of competing concepts should be used in the context at hand.

These normative questions about thought and talk — how should we use our words? which concepts should we use? how should we use them? — are questions in what we will call *conceptual ethics*.[7] Two important points about our use of the term 'ethics' here. First, we use the term 'ethics' in a broad sense, to designate questions about how to live and what to do. Second, we do not mean to indicate that conceptual ethics necessarily concerns practical norms as opposed to broadly theoretical ones.[8]

Disagreements about conceptual ethics need not be expressed via any special linguistic mechanism of course; they can be perfectly explicit. Indeed, many disagreements about which concepts to employ or about how best to use our words are straightforward exchanges wherein speakers *literally express* their competing claims about concept choice or word usage.

Nevertheless, we argue that many disagreements about conceptual ethics are not expressed explicitly. Metalinguistic negotiations are the most important instances of this implicit, or tacit, kind. Such disputes may not at first glance appear — either to the speakers themselves or to the theorist — to reflect disagreements about concept choice. But in fact they do reflect disagreements about concept choice. Many disputes that theorists have thought must be analyzed in terms of the *shared* literal content are thus best analyzed as speakers using their words in different ways, advocating (metalinguistically) for their preferred usage.

6. We do not assume that Barker would grant this characterization of metalinguistic usage, but it is in the spirit of his description in (Barker 2002). We go into greater detail below.

7. The term 'conceptual ethics' is drawn from (Burgess and Plunkett 2013a) and (Burgess and Plunkett 2013b).

8. For more on this point, see (Burgess and Plunkett 2013a).

A number of philosophers are actively involved in advancing normative arguments in conceptual ethics, a topic that will play a major role in our discussion.[9] And other philosophers have already noted the potential for metalinguistic analyses in responding to disagreement-based arguments in specific domains of normative and evaluative discourse.[10] Our paper is heavily indebted to these discussions. However, we have a more general aim. Our aim is to demonstrate that analyses of particular disputes as metalinguistic do not serve as exotic, last-resort maneuvers to avoid domain-specific disagreement-based arguments. Rather, the metalinguistic analysis follows from quite general and independently motivated observations about the linguistic mechanisms by which speakers routinely express their disagreements with one another.[11] When the relevant (and exceedingly slippery) notions — *meaning* and *disagreement* among them — are made sufficiently clear, the metalinguistic analysis is a natural and in some cases nearly inevitable consequence. Metalinguistic negotiation is, we argue, ubiquitous.[12]

Because of the generality of our aim in this paper, we do not argue (for present purposes) that the metalinguistic analysis is mandatory for any specific domain of normative or evaluative discourse. Rather, we aim to demonstrate that the metalinguistic analysis is a strong, if not inevitable, candidate in the explanation and analysis of *any* particular domain of normative or evaluative discourse. This by itself is enough to block quick arguments from disagreement to any particular conclusion about semantics. Just as importantly, the metalinguistic analysis provides theoretical resources to semantic views of normative or evaluative terms that are often criticized precisely for being unable to account adequately for disagreement. These views include, among others, contextualist views on which normative or evaluative claims have a determinate meaning only relative to some judge or group (*e.g.*, views where the term 'morally right' means something like "being approved of by me" or "being approved of by my community"). Our work in this paper shows that such views have a richer set of tools for explaining disagreement than is normally thought. Finally, as we argue at the end of the paper, the quite general plausibility of the metalinguistic analysis has important methodological consequences

9. See, for example, (Chalmers 2011), (Sider 2012), (Sundell 2011a), and (Haslanger 2012). For discussion of more examples of conceptual ethics, see (Burgess and Plunkett 2013a).

10. See, for example, (Sundell 2011b), (DeRose 2004), (Björnsson and Finlay 2010), and (Robinson 2009).

11. Perhaps alone among the philosophers working in this area, Peter Ludlow argues that linguistic communication is pervasively metalinguistic. He argues that nearly all conversations proceed *via* a negotiation over how to use the relevant linguistic expressions. See (Ludlow Manuscript). Part of what Ludlow is arguing for is the thesis that metalinguistic disputes are capable of expressing genuine substantive disagreements and that philosophers of language need to pay more attention to these disputes in their construction of theories of the meaning of terms (including normative and evaluative ones). In this respect, Ludlow is arguing for much the same thesis that we are. However, there is an important difference between Ludlow and us. Ludlow reaches his conclusion by arguing for a view of language — a view he calls the *dynamic lexicon* — that will strike many philosophers as quite radical. On Ludlow's picture, almost all linguistic communication involves conversation-specific settling of antecedently *highly* underspecified meanings. We are sympathetic to Ludlow's view. However, in this paper, we aim to demonstrate that the metalinguistic analysis of normative and evaluative disputes is consistent with, and indeed highly plausible on, entirely mainstream views of linguistic communication. In this respect, our argumentative strategy marks a significant departure from the one that Ludlow is currently pursuing.

12. Readers familiar with the work of Charles Stevenson might have already noticed that our proposal has certain resonances with his classic paper "Persuasive Definitions" (Stevenson 1938). Stevenson argues that many seemingly ordinary disputes in fact involve speakers putting pressure on each other to change the descriptive criteria that are paired with positive or negative emotive valence. Given what Stevenson here thinks goes into the meanings of terms (both a descriptive and a so-called "emotive" component), this process involves speakers trying to get others to change what they mean by their terms. We think that Stevenson was on to something important about many normative and evaluative disputes. However, there are crucial differences between our view and Stevenson's. For instance, in endorsing a version of emotivism, Stevenson makes significant assumptions about what goes in to the meaning of normative and evaluative terms that we do not make. Furthermore, he reaches quite specific conclusions about *which* aspects of meaning speakers try to change in normative and evaluative disputes. As will become clear in what follows, we think that there are many different aspects of meaning that speakers try to change via metalinguistic negotiation.

for philosophers working in metanormative theory. It means that philosophers cannot use intuitions about disagreement as a quick means for dismissing views that posit various forms of inter-personal, inter-contextual, or inter-theoretical variation in meaning — a move that is pervasive in much of the metanormative literature.

The paper is divided into six sections. In §1, we clarify the type of disagreement-based argument we are concerned with, and why it is so attractive. In §2, we propose a framework for thinking about the relevant notions — *meaning* and *disagreement* among them — that matter in the sort of disagreement-based argument we are concerned with. We argue that, when the relevant notions are made precise and held constant, it becomes quite easy to see that at least some disputes involving differences in meaning can nonetheless express genuine disagreements. In §3, we make this case with respect to disputes in general. In §4, we make this case with respect to normative or evaluative disputes — that is, disputes involving at least one normative and evaluative term — in particular.[13] In §5, we conclude our main argument by explaining the general methodological upshots of the metalinguistic analysis for those who work on the semantics of normative or evaluative terms.

Finally, in §6, we consider and respond to three important objections to our main argument. The first objection concerns externalism (both content-externalism and context-externalism). The second objection concerns the distinction between explanatorily basic and explanatorily non-basic normative or evaluative terms. By 'basic', we mean those terms (including, perhaps, 'ought' or 'good') whose meaning is not explained in terms of the meaning of any other normative

13. In what follows, we will take this to be definitional of "a normative or evaluative dispute". That is: when we claim that a dispute is "a normative or evaluative one", what we are claiming is just that this dispute is one in which a speaker uses at least one normative or evaluative term. It should also be noted that, in this paper, we will not take a stand on exactly what it takes to be a normative or evaluative term. Instead, we will take it as given that there is a broad class of terms here that metanormative theorists have been concerned with, and that such terms, however their nature is best understood, are the ones that we are talking about in this paper.

or evaluative terms. The third objection concerns the ability of the metalinguistic analysis to account for certain linguistic data about the acceptability of metalinguistic *negation* in various contexts. We take it that addressing these three objections is crucial to fulfilling one of our main goals in this paper: to provide a more thorough and systematic defense of the metalinguistic analysis of normative and evaluative disputes than has hitherto been on offer. However, we put our discussion of these important issues at the end of the paper to streamline the central line of argument. Moreover, we recognize that individual readers will be more concerned with some objections than others. §6 can thus be sampled or skipped per the reader's own interest and inclination.

§1 Disagreement-Based Arguments for Semantic Conclusions in Metanormative Theory.

To get a better sense of the type of disagreement-based argument that we are concerned with, consider the following three examples:

> *Terrance Horgan and Mark Timmons on the case of Moral Twin Earth*: Horgan and Timmons imagine the following case: there are no differences between the planets Earth and Moral Twin Earth except for the fact that people on Earth apply moral terms such as 'good' and 'right' in a way that is causally regulated by natural property N1 whereas people on Moral Twin Earth apply those terms in a way that is causally regulated by a different natural property N2. In figuring out how to respond to the possibility of Moral Twin Earth, Horgan and Timmons claim that we have two possibilities: On the one hand, we could claim that Earthlings and Twin Earthlings mean or refer to different things by their moral terms. On the other hand, as Horgan and Timmons write, "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."[14]

Michael Smith on 'Right': Smith argues that metaethical theorists "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."[15]

Peter Lasersohn on Taste Predicates: Lasersohn argues that the contextualist analysis of taste predicates "cannot be right […] because of the relation between contents and contradiction […]. If I say 'Roller coasters are fun,' and you say 'No, roller coasters are not fun,' on this analysis, you are not contradicting me, because the negated sentence doesn't express the same content for you as it does for me. In effect, my utterance means roller coasters are fun for me, and your utterance means roller coasters aren't fun for you, and there is no conflict or contradiction between those at all — indeed, there is no reason to think we disagree in any way, on this analysis."[16]

14. (Horgan and Timmons 1993, 130).

15. (Smith 1994, 35).

16. (Lasersohn 2005, 649).

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.[17] 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.[18]

Why think that there is this sort of tight link between facts of disagreement and facts of meaning? The core line of thought can be put as follows. Let's use the term *dispute* to refer to any linguistic exchange that *appears* to evince or express a genuine disagreement. We as theorists then have some intuition — an intuition that we want to preserve — that certain disputes *in fact* evince genuine disagreement. The standard way to non-debunkingly explain our intuition of disagreement — to give an analysis that *vindicates* our intuition about the dispute — is to posit that the 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.

Call any dispute that centers on the truth or correctness of the content literally expressed by the speakers a *canonical dispute*. If it is true

17. Although we have glossed just three examples of this disagreement-based form of argument in this section, there are *many* instances of the argument in the contemporary metanormative literature. A quick sampling: (MacFarlane In Progress), in discussing the semantics of aesthetic terms; (Egan 2012), in discussing the semantics of value terms in general; and (Dworkin 2011), also in discussing the semantics of value terms in general. For readers interested in our views on Dworkin's use of this argument, see (Plunkett and Sundell, Forthcoming-b). For an overview of the way in which expressivists make use of this argumentative strategy, see (Darwall, Gibbard, and Railton 1997). See (Gibbard 1990) for one of the clearest examples of this.

18. At this stage in our discussion, we paint in broad brush strokes with our use of the term 'meaning'. For instance, we are not yet distinguishing between character and content. We turn to this distinction in §2.

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.[19] 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.

The above line of reasoning is solid as far as it goes. If a dispute expresses a disagreement over the truth or correctness of literally expressed content, then it *is* likely that speakers mean the same things by their words. However, we take issue with the truth of the *premise* in that line of reasoning.[20] 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 reflect 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.[21]

If this kind of analysis is right, then it undermines a philosophical argument that is truly pervasive. And not just in contemporary work. As our opening example of Hare's case of the missionary and the cannibals suggests, the argument has a *long* history in discussions of expressivism and other forms of non-cognitivism. In the context of mobilizing the argument in support of his own form of expressivism, Allan Gibbard emphasizes its close connection to Moore's highly influential "open question argument".[22] And one of the main reasons Stevenson advocates a version of what we now call "expressivism" over subjectivism is that "we must be able sensibly to *disagree* about whether something is 'good'"[23] and that this is a requirement subjectivism can never satisfy.[24]

Finally, the use of the disagreement-based form of argument that we consider in this paper is hardly parochial to metanormative theory.

19. There are exceptions to this that depend on the specific ways in which words can vary in their meaning. Suppose that, by employing the relative gradable adjective 'tall' in uttering the sentence 'John is not tall', I express the proposition that John is *not* tall for a philosopher, while you respond with a usage of 'tall' that expresses the proposition that John *is* tall for basketball player. In the relevant sense, we do not mean the same thing by 'tall'. Nevertheless, we do express inconsistent propositions: If John is not tall for a philosopher, there's no way he's tall for a basketball player. Our goal here is to critique the argumentative move from the presence of disagreement to sameness of content, and we take cases like this one to advance that general point. However, it is orthogonal to our more general argument, so we set it aside for now.

20. One might describe the disagreement-based argument we are concerned with as based on an inference that is invalid. Yet here we identify a false premise. But there is no deep distinction between these diagnoses. If a particular disagreement-based inference does not make use of a premise to the effect that the dispute in question is canonical, then it is invalid: the premises about disagreement do not entail the conclusion about sameness of content. If the inference *does* make (perhaps enthymematic) use of a premise to the effect that the dispute is canonical, then it still of course requires that that premise be true. We argue that often it is not.

21. It is worth emphasizing here that, given the way we are setting things up, classifying a dispute as a normative or evaluative one means *only* that it is a dispute in which a speaker uses at least one normative or evaluative term. It does not mean, for instance, that the dispute is in fact about such-and-such topic. And it certainly does not mean that it is a *canonical* dispute about such-and-such normative or evaluative subject matter.

22. (Gibbard 2003). For Moore's original discussion, see (Moore 1993).

23. (Stevenson 1937, 16).

24. Consider also the case of A.J. Ayer. He observes, in language strikingly similar to that of contemporary statements of the argument, that "if a man said that thrift was a virtue and another replied that it was a vice, they would not, on this theory, be disputing with one another. One would be saying that he approved of thrift, and the other that *he* didn't: and there is no reason why both these statements should not be true". (Ayer 1971, 113–114). (Ayer is here explicitly drawing on (Moore 1970).) Unlike Hare, Stevenson, Gibbard, and the others, however, Ayer does not take this observation to refute the theories — his own emotivism *or* subjectivism — against which it is directed. But that is *not* because he rejects the central inference from disagreement to shared meanings of terms. Rather, Ayer simply bites the bullet, conceding the admittedly "paradoxical assertion" that "one really never does dispute about questions of value" (Ayer 1971, 114).

On the contrary, it appears throughout philosophy as a whole, in areas ranging from meta-metaphysics to epistemic modality.[25] Our focus in this paper is on the use of this disagreement-based form of argument within metanormative theory, but we take our discussion to have implications for the argument wherever it appears.

§2 Moving Parts.

As we stated in the introduction, the view that certain normative and evaluative disputes are both metalinguistic *and* that they express genuine disagreement flows naturally from a plausible and quite general picture of language and communication. In order to make this case, we put forward here a way of thinking about the relevant notions. In this section, we make four proposals about how to carve up the terrain. The first proposal concerns the term 'meaning'. The second concerns the nature of canonical disputes. The third concerns the individuation of words. And the fourth concerns the notion of disagreement itself. When these four proposals are put together, we will be in a good position to explain the general appeal of the metalinguistic analysis.

§2.1 Meaning.

Tyler Burge writes that "the term 'meaning' has always been vague, multi-purposed, and to some extent adaptive to the viewpoint of different theories".[26] We agree. So it's crucial to emphasize that when we talk about "meaning" in this paper, we mean something narrower than *the full communicative upshot* of what is said. It is now widely accepted by linguists and philosophers of language that there is some distinction to be made, within the category of information we communicate when we speak, between information that is part of the linguistically encoded content of the words we use (henceforth, the domain of *semantics*) and information that is communicated in virtue of other features of our use of those words (henceforth, the domain of *pragmatics*). By the "meaning" of our normative and evaluative terms we mean to indicate that we are talking about the semantics of those terms.

We do not have new or deep things to say about the semantics/pragmatics distinction in this paper. But we take it to be part of the rules of the game that any analysis of some linguistic expression will involve a package deal of semantics and pragmatics. We understand part of the aim of metanormative theory to be the construction of semantic analyses of our normative and evaluative terms. Thus, those analyses must, like any semantic analysis, be judged in this way, as a package deal of semantic and pragmatic theory. With this in mind, one of our core theses in this paper can be glossed as follows: if one aims to vindicate the intuition that a dispute evinces a genuine disagreement, more work can be done on the pragmatics side of this package deal than is often assumed by those making disagreement-based arguments for semantic conclusions.

Even having made this clarification between semantics and pragmatics, there still remains an ambiguity in the term 'meaning'. After all, even if we set aside pragmatic phenomena like presupposition and implicature, an expression, as it is used in a context, still has "meaning" in two senses: its linguistically encoded, contextually invariant meaning, and its reference or denotation, relative to that context. This, roughly, is the division Kaplan aims to capture with his distinction between an expression's *character* — its contextually invariant meaning — and its *content* — what it picks out, relative to the context.[27]

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.[28] Rather, we use his terminology to mark the relevant distinctions where necessary, however those distinctions are ultimately to be understood. For example, consider a contextualist view of 'tall' according to which speakers use the term to pick out different

25. Here are some illustrative examples of the role this kind of argument plays in these other areas: for meta-metaphysics, see (Manley 2009) and in epistemic modality, see (Egan, Hawthorne, and Weatherson 2005).

26. (Burge 1979, 398). Thanks to Nate Charlow for directing us to this passage.

27. See (Kaplan 1989).

28. See (Lewis 1979) and (Stalnaker 1999).

height-properties relative to a context. If such a view is right, then speakers in different contexts "mean" different things by the word 'tall' in one sense 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 threshold*.[29] Using this Kaplanian terminology, we keep these distinctions clear where necessary to avoid confusion. Where confusion seems unlikely, or where our points apply both to character *and* content, we will use 'meaning' with precisely this ambiguity in mind.

§2.2 Canonical Disputes.

How one should understand the notion of *canonical dispute* is closely connected to our proposal for understanding the relevant notion of "meaning". Many philosophers think that the linguistically encoded content of statements should be modeled in terms of *propositions*. But as the example of Hare's own brand of metaethical non-cognitivism demonstrates, one might reject such a view for how the linguistically encoded content of some (or perhaps all) statements should be modeled. For our purposes, it makes no difference whether one understands literally expressed content in terms of propositions or not. This has important consequences for our definition of a canonical dispute. Consider, first, how canonical disputes would be characterized if literally expressed content were understood exclusively in terms of propositions.

> *First-Pass Definition of a Canonical Dispute*: A dispute consisting in Speaker A's utterance of *e* and Speaker B's utterance of *f* is canonical just in case there are two propositions p and q such that Speaker A's utterance of *e* literally expresses p and Speaker B's utterance of *f* literally expresses q, and q entails not-p.

29. We borrow this gloss of 'tall' from (Barker 2002), which we discuss in §3.

This characterization of canonicalness obviously does not make room for theories that explain content in terms other than propositions, such as some versions of expressivism. Yet expressivists are prominent among those who offer disagreement-based arguments, which themselves depend on some notion corresponding to what we have called 'canonicalness'.

What matters for the disagreement-based argument is not the nature of content as such, but simply the question of whether the incompatible contents must be literally expressed. Therefore, it would be a mistake to rule out such expressivist analyses *as* canonical in the relevant sense. Fortunately, the characterization of canonicalness is easily generalized.

> *Second-Pass Definition of a Canonical Dispute*: A dispute consisting in Speaker A's utterance of *e* and Speaker B's utterance of *f* is canonical just in case there are two objects p and q (propositions, plans, etc.) such that Speaker A's utterance of *e* literally expresses p and Speaker B's utterance of *f* literally expresses q, and q is fundamentally in conflict with p in the manner appropriate to objects of that type. (By p entailing not-q in the case of propositions; by the satisfaction of p precluding the satisfaction of q in the case of desires; by p's implementation precluding q's implementation in the case of plans, etc.)

In what follows, this second-pass definition of a canonical dispute is what we have in mind. Nevertheless, for ease of presentation, it will often help to assume that literally expressed content is best understood in terms of propositions.

§2.3 The Individuation of Words.

In describing the disputes we consider below, and especially in arguing that some of them are best analyzed as *metalinguistic negotiations*, we talk as if the question is whether speakers "mean the same thing"

or "express the same concepts" by their words. On this way of talking, we use the term 'concepts' to refer to *constituent components of thoughts* (roughly, the mental equivalent to words). And we are taking it that (a) the meaning of a word is given by the concept it is used to express and (b) the same word can be used to express different concepts. But does the type of situation we have in mind really involve speakers who *mean different things by their words*? Or do they, rather, use *different* words altogether — words that are homophonous, but nevertheless distinct in virtue of differences in meaning? Does the word 'bank' mean two different things, or are there two words, 'bank_1' and 'bank_2' that happen to be pronounced the same?

For our purposes, nothing substantive hangs on the distinction between these two ways of carving things up. Disputes that we analyze as metalinguistic negotiations — in which speakers negotiate how a word should be used, or which concept it should be used to express — can, with no important theoretical changes, be re-described as disputes in which speakers negotiate which of two competing, homophonous words should be used. In fact, in some discussions of concepts and word meaning, the latter form of description — which of the competing homophonous words should we choose? — is more apt. However, the former form of description — how should this word be used in these circumstances? — has certain expository advantages, and we stick to that way of talking for the duration of the paper.

§2.4. Disagreement.

Disagreement is a tricky notion, and theorists who discuss "disagreement" do not always have the same thing in mind.[30] It is thus crucial that we identify what we have in mind when using the term 'disagreement'. In approaching this topic, we need to identify a core notion of disagreement that both (a) matters for metanormative theory and (b) meaningfully connects to the sorts of disagreement-based arguments for semantic conclusions that are our focus in this paper. We thus need a notion that (a) picks out something that connects meaningfully to phenomena worth explaining in metanormative theory, and (b) is broad enough to make room for a range of theoretical perspectives in metanormative theory, and hence does not beg important questions about how disagreement must be understood.[31]

In that spirit, here are three more specific desiderata. First, we want a notion of disagreement that can explain what is common between those disagreements that are evinced in linguistic exchanges (such as the disagreement between the missionary and the cannibal) and those disagreements that are not (for example, the disagreement between David Lewis and Hegel about philosophical methodology). We have already set things up in a way that is responsive to this desideratum: by using the term 'dispute' to refer to any linguistic exchange that appears to express a genuine disagreement, we have implicitly acknowledged that some disagreements aren't expressed at all. One upshot of this desideratum is that disagreement is a *state* that obtains, rather than an activity that people engage in.[32]

30. For a helpful discussion of some of the different important senses of "disagreement" that matter in contemporary discussions in philosophy of language and beyond, see (MacFarlane In Progress). See also chapter 6 of (Ridge In Progress).

31. This last point is especially important given that metanormative theorists often take it as a *platitude* that speakers disagree with each other in certain central cases. For instance, in the quote from Michael Smith in *The Moral Problem* that we gave in §1, he states that it is a "platitude that if A says 'x is right' and B says 'x is not right' then A and B disagree" (Smith 1994, 35). If one wants to take it as a platitude that speakers disagree in certain central cases — or, more generally, if one simply wants to take it as a data point that a wide range of theorists can agree on — then one shouldn't build too much into one's account of disagreement. The more one builds in, the less clear it will be that it is really *disagreement* in such-and-such central cases that is going on here. And the status of purported platitudes such as Smith's could then be severely undercut, and, with it, the viability of using facts of the presence of *disagreement* as such in the premises of disagreement-based arguments of the form that we are considering in this paper.

32. In making the distinction between "state" vs. "activity" understandings of disagreement, we here draw on (MacFarlane In Progress), who in turn draws on Cappelen and Hawthorne (Cappelen and Hawthorne 2009, 60–61).

Second, we want a notion of disagreement that allows for disagreement with one's past or future self, such as the disagreement between President Obama and his past self on gay marriage.

Finally, given that we want to make room for expressivist accounts of canonical disputes, we want a notion of disagreement that allows for there to be disagreement based on a conflict in attitudes such as plans, desires, or preferences. We want to accommodate, for instance, the intuitive idea that if Damien used to desire to become an astronaut when he grew up, and now desires to be a professional philosopher, then there is some sense in which Damien disagrees with his past self.

There are other issues on which we aim to remain neutral. But these desiderata allow us to be precise enough in stating a notion of disagreement relevant to the arguments we consider here. That notion is this: disagreement essentially involves some incompatibility (of the relevant kind) between contents (whatever they turn out to be) accepted (in the relevant sense) by different people (who may or may not be in conversation with one another). This conception is captured in following principle:

> *Disagreement Requires Conflict in Content* (DRCC): 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.)[33]

We think that DRCC has plausibility as an analysis of, or at least as a theoretically helpful claim about, disagreement.[34] However, precisely because DRCC covers disagreements that aren't voiced in any sort of linguistic exchange (like Lewis/Hegel, or past/present Obama), it is not a principle that concerns semantics as such. This should not be surprising: if DRCC is on the right track, then disagreement must be understood as something involving the conflicting *attitudes* of the subjects involved (*e.g.*, their beliefs, desires, intentions, etc.) rather than anything specifically about the sentences that they utter. *Accepting*, whatever it is, is a mental attitude, not a linguistic act.

The upshot of these points is that while the presence of genuine disagreement may be a data point that a semantic theory should take into account, it isn't one that is fundamentally about what is going on at the level of language. This, we think, opens up the door for exploring different linguistic mechanisms — semantic and pragmatic — by which speakers might represent or express their conflicting mental contents. In other words, this way of thinking about disagreement, especially when coupled with the proposals we made in §2.1–§2.3, lends itself naturally to the possibility of non-canonical disputes that express genuine disagreements. That is the possibility that we now explore.[35]

§3. Non-Canonical Disputes.

Recall that the argument type we critique begins with the premise that a given dispute expresses genuine disagreement. From there, one

33. This principle is modeled broadly on Gibbard's treatment of disagreement in (Gibbard 2003).

34. There are further wrinkles that must be addressed in a finished analysis of disagreement itself. For example, if Person A in world w believes that p, while person B in world w′ believes that not-p, and if p is true at w and false at w′, then in an obvious sense A and B do not disagree with each other despite accepting incompatible contents. Issues of this kind are ably addressed in (MacFarlane In Progress). But this type of case does not pose problems for the view that we advocate in this paper. This is a case where people intuitively do *not* disagree, while the primary worry for the metalinguistic analysis is capturing all of the cases where people intuitively *do* disagree. Thus, cases like the one above do not impact our arguments here, and we set them to the side for the purposes of this paper.

35. We recognize that some readers might well reject DRCC. If so, the question then is this: is there an alternative non-question-begging claim about (or definition of) disagreement that both (a) meets the relevant desiderata laid out in this section and (b) will vindicate the sort of disagreement-based arguments that we are attacking? We think not. Focusing on DRCC helps to make this point clear. However, for the reason we just stated, readers who are skeptical of DRCC shouldn't let the general point we are making be obscured by the formulation of DRCC itself.

argues (perhaps based 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.)

In §1, we granted for the sake of argument the second step in this line of reasoning. In other words, we granted that if a dispute in fact involves the literal expression of incompatible contents — if it is *canonical* — then the speakers involved are very likely to mean the same things, in the relevant sense, by their terms. In this section, we argue that the first step is flawed. Many disputes expressing genuine disagreement are in fact non-canonical. In other words, many genuine disagreements are expressed *via* disputes in which the speakers literally express *compatible* contents.

§3.1 Non-Canonical Disputes and Genuine Disagreement.
The point that genuine disagreements can be expressed in disputes in which the speakers literally express *compatible* contents, while it may at first seem bizarre, can be seen via entirely un-exotic types of cases. Consider, for example, disagreements centered on information that is communicated via implicature.

(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.[36]

In this case, there is disagreement between the speakers, despite the fact that the propositions expressed by the two speakers are logically consistent. Indeed they are both true.[37] 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.

However, while it is true that the speakers communicate these incompatible contents, it is not the case that they literally express them. If the familiar Grician story about number words is correct, then the speaker of (2a) literally says that there is *at least* one proton in the nucleus of a helium atom, while the speaker of (2b) literally says that there are at least two protons in the nucleus of a helium atom. In this, they are both correct, since there are exactly two protons in the nucleus of a helium atom.

The speakers of (2a) and (2b) really do disagree with each other, both intuitively and by the lights of the principle DRCC. There is nothing non-genuine about the disagreement, and indeed it is a disagreement very much worth having. The speaker of (2a) is mistaken in virtue of believing, and communicating, that there is exactly one proton in the nucleus of a helium atom, when in fact there are exactly two. None of this is made the least bit problematic by the fact that according to

36. This example is introduced in a related context in (Sundell 2011a).

37. We assume here the familiar scalar-implicature analysis of number words. That analysis is hardly uncontroversial. (See (Chierchia 2004) among others.) But alternative examples are easy to come by. Even if one prefers a semantic account of scalar implicature quite generally, relevance implicatures will do the trick. Suppose that Sally is extremely bright but has a short attention span. Speaker A says "Sally was able to solve the last problem on the test", literally expressing the proposition that Sally had the relevant ability and communicating, *via* relevance implicature, that in fact she solved it. (That implicature need not be present, but it is easy to imagine contexts where it is.) Speaker B objects and responds that "she chose to quit before she got there". There is no incompatibility between the literal contents of the expressions uttered. But there *is* an incompatibility between further content that is conveyed pragmatically, namely that Sally in fact solved the last problem on the test. It is in virtue of that incompatibility that such a dispute expresses a genuine disagreement.

the semantic theory under consideration, the relevant propositions are communicated pragmatically and not semantically.

The example in (2) does not provide a case of speakers who genuinely disagree and yet mean different things by their words. We'll come to that type of case shortly. But it does make vivid two other points. First, it demonstrates the danger of assuming, based on the existence of genuine 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.[38]

The second point demonstrated by (2) is that theorists take a wrong turn as soon as they conflate the question of whether a disagreement is *genuine* with the question of how the information on which a disagreement centers happens to be communicated. Any principle like DRCC 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. Thus, from the fact that a given dispute reflects a genuine disagreement, we can already see that caution is required before any specifically semantic conclusions can be drawn.

38. It also clearly demonstrates the falsity of a principle that is sometimes used to support this inference: that if two speakers genuinely disagree with each other, then at least one of them says something false. Disputes such as that in (2) clearly demonstrate that this type of principle is much too strong.

§3.2 Metalinguistic Disputes.

Disputes centered on implicated information make vivid the independence of questions about substance from questions about communicative mechanism. However, as noted, there is little reason to think that the parties to the dispute in (2) mean different things by their words. Even if we allow that some genuine disagreements are expressed in non-canonical disputes, we might remain skeptical that a disagreement could be both genuine and expressed in a non-canonical dispute that is non-canonical *in virtue of a difference in word meaning*. However, such skepticism is unwarranted.

The fact that certain disputes are non-canonical in virtue of a difference in word-meaning is easiest to see in the case of context-sensitive expressions, so we begin with those. Consider, first, linguistic expressions that are context-sensitive in virtue of being gradable. This includes expressions like 'tall', 'big', 'cold', etc., (*relative gradable adjectives*, in the sense of Kennedy (2007)) that denote a specific property only once some parameter — a threshold along a scale of height, size, or temperature, say — has been settled by the conversational context or by the parties to the discourse in which the expression is used.[39] 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.

If, in a context, we know how that parameter is set — 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*. This latter kind of usage is described in Barker

39. See (Kennedy 2007).

(2002), who calls it a *sharpening* or *metalinguistic* usage of a term. Consider the following case, described by Barker:

> 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 truth 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.[40]

Why would you bother asking about "merely linguistic" facts like how to use the word 'tall'? Because how we use words matters. Knowing how people in your new surroundings use the word 'tall' will allow you to communicate more smoothly with them of course. But beyond that, knowing how they use the word 'tall' reveals potentially important information about the range of heights you can expect to see locally.

40. (Barker 2002, 1–2).

In his (2002), Barker does not consider cases where speakers disagree about the information communicated by this type of usage.[41] 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*.

Within this category of metalinguistic disputes, there is a further distinction to be drawn. In some metalinguistic uses of an expression, it is quite natural to think that there are antecedently settled facts about the linguistically relevant features of the conversational context, facts which are at least partially independent of the intentions — or at least the very local intentions — of the parties to the conversation. The dispute over the appropriateness of calling Feynman 'tall' is plausibly of this sort. In such cases, it is natural to think of the speakers as exchanging information that is in some (perhaps quite loose) sense, *objective* — information about what the context is actually like. If a disagreement should arise over that information, as it does in our extension of Barker's case, then the disagreement is a factual one about which of two or more competing characterizations of the shared conversational context is most accurate. However, not all cases of metalinguistic usage fit this profile.

Suppose that Oscar and Callie are cooking together. In fact, they are cooking food for the party where Barker's dialogue takes place. Oscar and Callie have each just tasted the chili, and Oscar utters (4a) while Callie utters (4b).

41. Though see (Barker 2013). (Sundell 2011a) extends Barker's case to include disagreements involving metalinguistic usages.

(4) (a) That chili is spicy!

(b) No, it's not spicy at all.

In this case, it is much less natural to think that there is some antecedently settled, objective fact of the matter about the contextually salient threshold for "spiciness". Rather than advancing competing factual claims about some independently determined threshold, it seems most natural to think of Oscar and Callie as negotiating what that threshold shall be.

Why would Oscar and Callie consider it worth their time to engage in such a disagreement, when they already agree on what the chili actually tastes like? Why engage in a dispute over how to use a word? The answer is the same as before: it is worth engaging in such a dispute because how we use words matters. 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"?

We use the term *metalinguistic negotiation* to refer to this second type of metalinguistic dispute — those disputes wherein the speakers' metalinguistic use of a term does not simply involve exchanging factual information about language, but rather negotiating its appropriate use. We think that metalinguistic disputes of this latter 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.[42]

Why are such exchanges perceived as disputes, when the speakers fail to assert inconsistent propositions? Because in addition to asserting those 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 "spiciness" threshold is preferable to some competing "spiciness" threshold is very much the kind of thing over which two speakers can disagree. To see that, we can just imagine them having a *canonical dispute* about the very same topic. The view we are proposing is that Oscar accepts the content that *we should use 'spicy' in such a way that it applies to the chili* and Callie accepts the content that *we should not use 'spicy' in such a way that it applies to the chili*. Those contents are rationally incompatible; this fact would be obvious if the two of them were engaged in a canonical dispute where Oscar said "we should use

42. Metalinguistic uses themselves — for example, the use of 'tall' in Barker's original case — may be analyzed as pragmatic or as a special kind of semantic usage, depending on the details of one's account of the mechanism involved. For example, the changes brought about in the listener's beliefs about the extension of 'tall' might be thought of as resulting from a kind of Lewisian accommodation, in which case metalinguistic usage is a kind of pragmatic mechanism. By contrast, one might focus on the fact that if the literal content of the speaker's assertion is accepted by the listener, this straightforwardly eliminates from the context set worlds where certain (linguistic) facts obtain — worlds, for example, where the word 'tall', relative to the context, fails to apply to Feynman. Such an analysis might suggest that metalinguistic usage is a semantic mode of communication, if an unusual sort. Or, it might be analyzed, as it is in Barker's work, within the framework of dynamic semantics, in which the distinction between semantics and pragmatics is blurry and yet more complex. None of this will affect our argument here, since in any case, the mode of communication is not semantic in any sense that would require that speakers who disagree with one another assert and deny inconsistent contents. Moreover, as we emphasize below, our focus is specifically on metalinguistic *negotiations*, in which ways of using linguistic expressions are not just described but advocated. That advocacy is unlikely to be part of the literal content of the expressions on any of these analyses, and thus our thesis that a good deal of the action is happening pragmatically holds, whatever one's specific views about the mechanism underlying metalinguistic usage itself.

'spicy' in such a way that it applies to the chili" and Callie responded by saying "no, we should not use 'spicy' in such a way that it applies to the chili".

The contents accepted by Oscar and Callie do not suddenly become rationally compatible when they express those contents pragmatically rather semantically. Thus, by the lights of DRCC, the disagreements that are reflected in metalinguistic negotiations are entirely genuine. Moreover, given the right context — for example, a context where we must coordinate our chili seasoning, or our thermostat adjusting, or our basketball picks, or our progressive taxation brackets — such disagreements can be very much worth having, a point to which we return in §4. And to emphasize our earlier point, the question of whether such disagreements are worth having is entirely independent of whether the competing claims are advanced via semantic or pragmatic mechanisms.

In the cases of metalinguistic negotiation we have considered so far, the negotiations have concerned how to fix parameter settings for bits of context-sensitive terminology. In other words, they have concerned "meaning" in the sense of *content*, but not of *character*. The contextualist grants, after all, that a term like 'tall' does not vary in its character, but only in its content. But metalinguistic negotiation is not confined to gradable adjectives or other context-sensitive expressions. It can even concern words that are seemingly quite fixed in their meaning.

In order to illustrate this point, we turn to a case from Peter Ludlow. In "Cheap Contextualism", Ludlow describes a heated debate he heard on sports radio.[43] The debate concerned a list of the greatest athletes of the 20th century, and the question of whether that list should include the racehorse Secretariat. Simplifying a bit, we can imagine the following exchange as part of that debate:

(5) (a) Secretariat is an athlete.

(b) No, Secretariat is not an athlete.

43. (Ludlow 2008).

Unlike the cases of metalinguistic sharpening involving gradable adjectives, there is little reason to think that the relevant linguistic expression here — 'athlete' — is semantically context-sensitive.[44] But as in those cases, there is also little reason to think that the dispute in (5) concerns straightforward factual matters about the topic at hand. The speakers of (5a) and (5b) mutually know all of the facts about Secretariat's speed, strength, etc., and what races, awards, metals he won, etc., just as Oscar and Callie mutually know the facts about the chemical hotness of the chili. The question on the table, then, is this: suppose we want to preserve the intuition that there is a genuine disagreement that is expressed in the exchange that Ludlow overheard. How then should we understand what is going on?

Regardless of one's precise views in semantics, it should be uncontroversial that at least *one* crucial type of data for figuring out what a speaker means by a term T are facts about the speaker's usage of T — patterns of usage that reflect her disposition to apply that term one way or another, more generally. Many philosophers, of course, claim that there are other factors that matter too, and we come back to this when we discuss externalist objections to our argument in §6. For now, though, we emphasize simply that — uncontroversially — one crucial piece of the puzzle in figuring out the meanings of terms are facts about speakers' patterns of usage for that term.

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. This, at the very least, provides *prima facie* reason for thinking that the speakers mean different things by the word 'athlete'. Thus, it is not

44. In this we set aside theories, such as Ludlow's *dynamic lexicon*, on which all or nearly all words are in some sense context-sensitive. In fact we are sympathetic to Ludlow's view, but we aim here to show that metalinguistic analyses are consistent even with less radical views.

unnatural to conclude — as Ludlow does — that the speakers do indeed mean different things, and that the dispute in (5) reflects a disagreement about which of two competing concepts, C1 or C2, is more appropriate to the 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'.[45]

On this understanding of the dispute, each speaker literally expresses a true proposition given the concept they in fact express with their term. But beyond that, the speakers pragmatically advocate for the concept that they are using and in virtue of which they assert those propositions. Thus, their metalinguistic dispute reflects a genuine disagreement about how to use the word 'athlete'. In particular, it is a debate in conceptual ethics about which among a range of competing concepts, and in particular, which of C1 or C2, is most appropriate to the conversation and should be expressed by the term 'athlete'.

Ludlow's Secretariat case is particularly vivid, but there is little reason to think that this type of negotiation is at all uncommon. 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 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 "spiciness" of their chili, metalinguistic negotiations influence and advance more general processes of collective decision-making and action.

Before moving on, we want to emphasize an important point about what is and is not involved in seeing a given dispute as a metalinguistic negotiation. Suppose that our proposed analysis of the Secretariat case is right. 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. 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 the "merely linguistic" issue about how to use the word 'athlete') not *really* what their disagreement is about?

To answer this question, we need to distinguish two things: (a) the background normative views held by the speakers that motivate them to go in for a dispute vs. (b) the conflicting mental states that are actually being expressed by the speakers. On the assumption that a given exchange is a metalinguistic negotiation, the immediate topic of disagreement in that exchange is one in conceptual ethics. But it is perfectly consistent to think both (a) that such-and-such dispute is one about conceptual ethics and (b) that a crucial reason why the speakers engage in this dispute is because of normative views they have about some topic *other* than conceptual ethics.

Suppose that Martha and George are car shopping and get in an argument about whether Subarus are good cars. It is entirely natural to affirm both of the two following theses: (1) Martha and George are having an argument about whether or not Subarus are good cars — the immediate topic of their dispute — and (2) the reason why Martha and George are having this argument is because they need to figure out what car they should buy together — the background issue that motivates them to have the argument about the merits of Subarus. The dispute between Martha and George is about whether Subarus are good cars, but they enter into that dispute because they differ on what car to

45. Of course, further factors — external environment, speech community, or the like — might intervene to override the *prima facie* evidence for this conclusion. But, as we emphasize in our reply to externalist objections in §6, an argument for an externalist alternative to our analysis of the Secretariat case must (a) be strong enough to override the primary, countervailing evidence of usage, and (b) not fall back on the need to explain the possibility of disagreement.

buy. Similarly, one can hold the following: the dispute between Ludlow's speakers is about which concept is best picked out by 'athlete', but they enter into that dispute because they differ on whether horses are aptly afforded certain kinds of praise (or because of some other normative issue that they disagree about). 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.

§4. Normative and Evaluative Disputes.

In the last section (§3), we described several cases of metalinguistic disputes that express genuine disagreements. These disputes have the following 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.

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.

If a theorist hopes to show that a particular dispute really does license the conclusion that speakers mean the same things by their words, she must provide independent support for the thesis that the dispute not only reflects a genuine disagreement, but also that it happens to be canonical. But of course, in demonstrating that the dispute is canonical, she will have done, by other means, most of the work necessary to show that the speakers mean the same things by their words. The disagreement-based argument itself is thus either unsound, unsupported, or unnecessary.

The question we turn to in this section is this: how plausible is it that at least some normative and evaluative disputes of the kind that interest philosophers are metalinguistic? One of the disputes that we have already canvassed — namely the one about Secretariat — involves a term, 'athlete', that is plausibly understood as either normative or evaluative. Calling someone an 'athlete', at least in the context we were concerned with, is plausibly in some way to praise her or him. 'Athlete' is thus (arguably) an example of a so-called *thick* normative (or evaluative) term: roughly, a term whose meaning sets application-conditions that involve both descriptive criteria and normative or evaluative criteria.[46]

Given this understanding of 'athlete' as a thick normative term, we have already seen an example of the relevant type. Nonetheless, one might worry that this is not an analysis that can extend to many of the normative and evaluative disputes that have interested metanormative theorists. This, however, would be a mistake. Absent further argument about what is special about some specific normative or

46. For a helpful recent discussion of thick normative terms, some of which is in the same spirit as our discussion, see (Väyrynen 2013). Note that, on Väyrynen's view, only descriptive application-conditions are built into the *semantics* of thick normative terms.

evaluative dispute, we see no reason why the metalinguistic analysis shouldn't be a viable option in a wide variety of cases. The Secretariat case provides a plausible model for accounting for a range of other normative or evaluative disputes that have interested metanormative theorists. On such an analysis, those normative or evaluative disputes would share two features with the Secretariat case: (a) they would be taken to involve metalinguistic usage (a distinctive linguistic mechanism) and (b) they would be analyzed as centering on a question of conceptual ethics (a distinctive normative topic).

In this section, we discuss two different cases to demonstrate the plausibility of this type of analysis. We start with a case involving the use of the term 'torture', which, like 'athlete', is plausibly read as a thick normative term. We then discuss one of the cases that we introduced at the start of this paper, the case of Moral Twin Earth.[47]

Let's start with the case of 'torture'. Suppose that, in the context of a policy debate, two speakers disagree about the status of waterboarding and utter, in turn, (6a) and (6b).

(6) (a) Waterboarding is torture.

(b) Waterboarding is not torture.[48]

Let us suppose further that the speaker of (6a) follows the United Nations in defining torture as any act inflicting severe suffering, physical or mental, in order to obtain information or to punish[49], while 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.[50]

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.

Now consider again Horgan and Timmons's case of Moral Twin Earth. To recall, in Horgan and Timmons's thought experiment, there are two planets that are distant from each other — Earth and Moral Twin Earth. These planets are exactly the same except for the fact that people on Earth apply moral terms in a way that is causally regulated by natural property N1 whereas people on Twin Earth apply those terms in a way that is causally regulated by a different natural property N2.[51] Horgan and Timmons think that in order to explain the possibility of genuine disagreement between Earthlings and Twin Earthlings, one needs to understand them as meaning and referring to the same

47. In §6, we look at a kind of normative or evaluative dispute that, for reasons discussed there, appears perhaps *most* resistant to the metalinguistic analysis. These will be cases of disputes involving *basic* normative or evaluative terms, *i.e.*, those terms whose meaning — such as 'right', 'warrant', or 'good' — can perhaps be used to explain the meanings of other normative and evaluative terms. We argue that, absent considerable further argument, the disagreement-based argument is unwarranted even in those cases.

48. The example of torture is mentioned in (Chalmers 2011) and discussed in some detail in (Sundell 2011a). Both make points similar to the one we make here.

49. See (United Nations 1984, 85).

50. See (U.S. Department of Justice 2002, 340A).

51. We are presenting the Moral Twin Earth case in the main way that Horgan and Timmons discuss it: namely, as a case about two different planets that are distant from each other. This presentation parallels Putnam's original version of his twin earth case about the use of the term 'water'. There are, however, also versions of both Horgan and Timmon's case and Putnam's case that one can run that involve planets located in different possible worlds. Our arguments here, suitably modified, carry over to such cases. (We address worries about cross-context disagreement in this section; those points apply across planets, or across worlds.) We therefore stick with the main Horgan and Timmons case (*i.e.*, the one about different planets in the same possible world) in what follows.

things by their moral terms.[52] But this move is too quick: at least some of the disputes involving moral terms between Earthlings and Twin Earthlings may be metalinguistic.

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". Now imagine that some people from Earth and Moral Twin Earth end up meeting. As part of this meeting, imagine that there is the following exchange between Bob the Earthling and Chris the Twin Earthling:

(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.

Can we view the exchange between Bob the Earthling and Chris the Twin Earthling as a genuine disagreement even if we hold fixed our assumption that analytic utilitarianism is true of Bob's term 'morally right' and analytic Kantianism is true of Chris's term 'morally right'? Recall our discussion of the Secretariat case. Just as in that 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.

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 social-historical setting — a setting in which certain words (largely independent of which specific concept 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.[53] 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.[54]

52. As our focus on *moral* terms here indicates, it is important to emphasize that we are here taking Horgan and Timmons to be talking about *moral* twin earth, and not *all-things-considered normative* twin earth. Many of the issues that Horgan and Timmons are discussing become more complicated when one is talking about *all-things-considered normative* twin earth. The reasons why will become clear in §6.2, when we discuss the difficulties of applying our style of metalinguistic analysis to explanatorily basic normative or evaluative terms (such as, arguably 'good' and 'right'), as opposed to non-basic ones such as 'athlete' or '*morally* good'.

53. The story we suggest here resonates with stories offered in (Copp 2000) and (Merli 2002). Thus, as we emphasize in the introduction of this paper, one way of seeing what we are doing here is to show why this general sort of response to the Moral Twin Earth case isn't a one-off response to a specific argument, but rather flows naturally from a general picture of communication.

54. We would not be worried if Bob and Chris themselves had intuitions about their dispute that ran counter to our analysis; ordinary speakers have all kind of mistaken intuitions about language. In fact, we think it is a mistake to think that ordinary speakers in fact have intuitions *one way or the other* on distinctions as subtle as this. Recall that on either analysis, the dispute between Bob and Chris ultimately turns on substantive normative questions about what to do. Whether that disagreement is reflected via inconsistent contents expressed with shared concepts, or via competing concepts vying to play a single functional role, is more than any ordinary speaker should be expected to have intuitions about. Moreover, even if they did have such intuitions, and even if our analysis contradicted these intuitions, we would not take this to be devastating to our proposed analysis. First, folk intuitions about how thought and talk works can often be mistaken. Second, as we explain later in

But one rather serious worry remains.[55] Most of us have the intuition that, even if Bob the Earthling and Chris the Twin Earthling never talked with each other — even if they didn't know the other existed, much less how they used their words — there would still be a disagreement between them. So how could our proposed analysis have a hope of being correct? More generally, how can a metalinguistic account of this type of dispute explain *any* case of disagreement across contexts — disagreement in which there is no shared context to be arguing about, or no shared language to be negotiating?

In addressing this worry, recall, first, that metalinguistic negotiations are distinguished from canonical disputes in two ways. There is the distinctive linguistic mechanism — metalinguistic usage — and a distinctive normative topic — a topic within what we have called "conceptual ethics". On our proposal, Bob and Chris's disagreement concerns which concept to express with the term 'morally right'. Our proposal is that they each advocate a view about which concept is best suited to play a certain functional role in thought and practice, a role that includes matters of how to treat others, what to hold each other responsible for doing, and how to live more generally. One reason that Bob and Chris want their preferred concept to be the one expressed by the term 'morally right' is because they each believe (correctly) that whatever is called 'morally right' will likely play this role, and that people who think that 'morally right' should mean something different have a different view about what concept should play this functional role. As in the case of 'athlete', this may have nothing to do with what is analytic about the term. It might simply have to do with social, historical, and psychological facts about what is standardly associated with the use of that term.

Given this sort of disagreement in conceptual ethics, it is entirely sensible to suppose that *before* they engage in any linguistic exchange — before they are even aware of one another's existence, much less one another's language — Bob and Chris have views about which concept should play this important functional role in thought and practice. They would have those views entirely independently of any awareness of the other person, or of the existence of another language. Bob's view is that the analytically utilitarian concept C1 should play this functional role in organizing our lives and Chris's view is that the analytically Kantian concept C2 should play that role. These views are incompatible, irrespective of whether Bob and Chris engage in a conversation where they express these views. Thus, by the lights of DRCC, they disagree, irrespective of whether they engage in conversation. In turn, if they *do* start talking to each other it is entirely sensible that they have a disagreement not just about which concept to deploy, but which concept to deploy using the expression 'morally right'. Thus, the normative issue becomes not only which concept to use, but also which concept will be paired with this particular word 'morally right'. That linguistic exchange, once it happens, can proceed via metalinguistic negotiation.

Our goal here is not to advocate, once and for all, for this particular analysis in this particular domain. The important point is that an analysis of the disagreement between Bob and Chris as concerning conceptual ethics before they meet, and concerning conceptual ethics and expressed *via* metalinguistic negotiation once they do — is more than sufficiently plausible to undercut quick inference to the best explanation arguments from intuitions about their disagreement to semantic conclusions about what Earthlings and Moral Twin Earthlings mean by the relevant moral terms. It may not be able to undercut them all-things-considered, but it is certainly enough to block the quick way in which these inference to best explanation arguments in this case are often deployed.

We have just suggested that the dispute between Bob the Earthling and Chris the Twin Earthling can plausibly be read as a metalinguistic one about which concept deployment and articulation patterns to adopt. (fn 56 here). A wide range of semantic theories of 'morally right'

this paper, there is good reason to suspect that people will often resist viewing one of their disputes as a metalinguistic one, even when it in fact is. See (Plunkett and Sundell Forthcoming-a) for more detailed discussion.

55. Thanks to Mike Ridge for emphasizing this worry.

can make use of this kind of metalinguistic analysis of this kind of dispute. However, the availability of this kind of analysis will be most helpful to those views that are often accused precisely of being unable to explain the possibility of genuine disagreement. These include contextualist views on which normative or evaluative claims have a determinate meaning only relative to some judge or group — for example views where the term 'morally right' means something like "being approved of by me" or "being approved of by my community"[56]. They also include views that take a substantive view from within moral theory to be built into the meaning of a term such as 'morally right' — for example analytic utilitarianism, which takes the meaning of 'morally right' to be something like "maximizes expected utility". Such views have the promise of explaining disagreements by drawing on the full range of non-canonical disputes, and especially metalinguistic ones. How viable such views are will depend on a more detailed exploration of the specific cases, some of which are given only sketchy descriptions when they appear in disagreement-based arguments. Crucially, however, once the metalinguistic analysis is available to them, these views cannot be ruled out simply in virtue of the presence of genuine disagreement.

§4.1 Do Metalinguistic Negotiations involve Disagreements Worth Having?. We have argued that the metalinguistic analysis of a given normative or evaluative dispute can't be ruled out by the intuition that the dispute expresses a genuine disagreement, even granting the aim of vindicating that intuition. Perhaps, though, metalinguistic analyses of such disputes can be ruled out by their inability to vindicate other intuitions that we have about them — intuitions that are tied to the fact that there is a disagreement *of a certain sort* that is expressed in the dispute. In this section, we consider what we take to be perhaps the most important issue on this front: namely, the question of whether or not metalinguistic negotiations can reflect *disagreements worth having*.

In thinking about our choices in how to use our terms, it is crucial to understand that that matters of word usage are not limited to mere definition or stipulation, but can themselves be answerable to substantive adjudication. Indeed, in some cases questions of words and usage can be answerable even to purely objective, descriptive considerations in the natural sciences. To see this, suppose that, in the context of a biology classroom, two speakers disagree about the status of tomatoes and utter, in turn, (8a) and (8b).

(8) (a) Tomato is a fruit.

(b) No, tomato is not a fruit.

Let us suppose further suppose that the speaker of (8a) is a plant biologist, and the speaker of (8b) is a chef. Just as in the 'torture' case, even if we suppose that the speakers mean different things by the relevant expression, it is clear that we have not exhausted the evaluative work to be done. While the chef's definition of 'fruit' has advantages relative to our culinary practices and gustatory tradition, the biologist's definition — one according to which tomatoes really are fruit — is better suited to the biology classroom. It is better suited to this scientific context not as a matter of convention or stipulation, but objectively better: the objects in the extension of the biologist's term go together more metaphysically naturally than the objects in the extension of the chef's definition.[57] And it is precisely those

56. For helpful statements of this sort of contextualist view, see (Dreier 1990) and (Harman 2000). For criticism of this sort of contextualist view, including on the grounds that it doesn't adequately account for disagreement, see (Gibbard 1990), (Smith 1994), (Boghossian 2006), and (Egan 2012).

57. For discussion advocating the idea that some concepts are better suited to describe the world given its objective nature, see (Sider 2012). The 'tomato' case and a number of related examples are discussed in more detail in (Sundell 2011a). It should be noted that, although we have put our argument here in terms of a type of metaphysical realism, our point here is a general one. If one starts out as a metaphysical anti-realist, nothing about the metalinguistic analysis of this dispute would force one to see it as adjudicable with reference to metaphysically real natural kinds. The important (and more surprising) point is that the metalinguistic analysis is entirely *consistent* with a fully robust metaphysical realism about natural kinds. Moreover, it is entirely consistent with thinking that there are metaphysically fundamental normative

metaphysically natural categories that we aim to identity when we are in the biology classroom.

Thus, even if we assume that the dispute in (8) is like the disagreement about Secretariat, and that the parties really do mean different things by their words, there remain very real issues to be resolved and there exist all the usual resources — difficult as it may be to sort them out — on the basis of which we can take one party or another (or both!) to be mistaken. This despite the fact that, owing to differences in meaning, both parties to each of the disputes literally express true propositions. As we have been emphasizing throughout, disagreements in conceptual ethics can be closely tied to substantive, parallel issues outside of conceptual ethics. They can be tied to (even if not *directly* about, for reasons discussed at the end of §3.2) coordinating decision-making and action in our chili-preparation, questions about the praise-worthiness of Secretariat, and even entirely objective questions about how best to carve up the natural world into biological categories. Thus, we think metalinguistic negotiations are entirely capable of expressing genuine disagreements of just the type that we think we're having in the respective cases, substantive disagreements that are well worth engaging in.

One might nonetheless worry that even if we are right that metalinguistic negotiations are capable of expressing the relevant sort of substantive disagreement worth having, there is still something fishy going on here; that there is at least *some sense* in which these disagreements remain less substantive than disagreements that are expressed in canonical disputes. One way to try to make good on this intuition is to appeal to facts about what speakers themselves take to be worth spending their time arguing about. We can ask: would the parties to metalinguistic negotiations really see themselves as expressing disagreements worth having if they knew what was going on? Perhaps they would just restate their view in a shared vocabulary, and the disagreement would dissolve. David Chalmers, among others, has seen paraphrasing as a way of distinguishing "merely verbal" disputes (that aren't worth having) from substantive ones (that are worth having) in roughly the following way: substantive disputes are ones that survive paraphrase, while merely verbal ones do not.[58] Thus, it is worth asking: do metalinguistic negotiations of the sort we have been discussing survive paraphrase?

Given that we have already shown that metalinguistic negotiations can center on information that *in fact* can be worth debating, the question here amounts to this: are speakers involved in a metalinguistic negotiation likely to recognize that their dispute is one worth having once it has been paraphrased into a fuller theoretical description of what is going on? Or, instead, once they recognize their dispute to be a metalinguistic negotiation, will they think that their debate is "merely verbal"?

In order to get a handle on this question, first consider that when speakers involved in a dispute *take themselves to be at odds* there is little reason to think that they have fine-grained intuitions about the modes of communication with which they express their disagreements, or about subtle theoretical distinctions amongst equally substantive kinds of disagreement. Ordinary speakers do not have intuitions of the kind that could make subtle distinctions about the semantics/pragmatics distinction. How could they, when theorists are in many cases divided about which aspects of meaning belong on which side of the divide? And neither should we think that ordinary speakers have intuitions of the kind that would make subtle distinctions between first-order issues and closely corresponding issues in conceptual ethics. Questions about modes of communication and first-order vs. (corresponding) conceptual issues are *theoretical questions*. While ordinary intuitions of acceptability, entailment relations, and the like are here, as elsewhere, a crucial source of data in answering these questions, the answers themselves are not the kind of thing we have intuitions about. (Recall that the intuitions we *do* have — that there is a disagreement;

facts — including metaphysically fundamental normative facts about what concepts one should use.

58. (Chalmers 2011).

that the disagreement may be worth having; that the disagreement is ultimately about how to season the food, treat prisoners, categorize the biological world, etc. — are *vindicated* by the metalinguistic analysis.) Thus, the metalinguistic analysis is not really the kind of thing that speakers could, in the relevant sense, *resist*.

Suppose, though, that a metalinguistic analysis of some dispute *were* incompatible with the speakers' own rough understanding of what the dispute is about — that the speakers, for example, firmly believed that they were having a canonical dispute directly about *the real nature of torture*, rather than a metalinguistic negotiation about how the term 'torture' should be used in the context at hand. In such a case, whether or not the speakers will see the debate as one worth continuing to engage in will depend on (a) whether there is something substantive at stake in how the relevant terms are used in the context (which, as we have argued, there often will be) and (b) whether the speakers recognize this fact. If there is something at stake, and the speakers are in a position to recognize this, there is no reason to think that the speakers will dismiss their dispute as one not worth having.

Before moving on, there is one further point about paraphrase to mention. This is that, in some metalinguistic negotiations, there can be an important reason why the participants in the dispute resist the paraphrase of their discussion *even if* both parties in the dispute would still continue to think that their dispute was worth having were they presented with the paraphrase. Consider that, as we have discussed, many metalinguistic negotiations worth having are ones in which the participants are involved in a (largely tacit) disagreement about which concept should play a certain functional role in the thought and practice of a group of people. In social contexts wherein the concept associated with a given word (*e.g.*, 'torture', 'athlete', 'democracy', 'freedom', 'happiness', or 'equality') plays some important functional role, partly *in virtue of* it being associated with that given word, it makes sense that the people care about their preferred concept being paired with the word in question. Thus, for example, given that whatever counts as an "athlete" is eligible to be held in a certain esteem (and perhaps showered with certain monetary rewards) — a feature of the term (let's suppose) that has nothing to do with what is analytic about 'athlete', but rather with sociological facts about its functional role — the participants in Ludlow's dispute have very good reason to resist their interlocutor's usage, even if they recognize that, given the difference in meaning, it may be literally true. Allowing for such paraphrase would limit the ability of the participants to effectively advocate for their preferred concept.

To emphasize this last point, consider that in a context such as a debate on sports radio – and, indeed, in many contexts of more public communication (such as political rallies, protests, and arguably large parts of political *philosophy* itself) – one is not just trying to communicate with one interlocutor, but with a whole range of people who have some rough associations with the relevant words such as 'athlete' or 'freedom'. In such a context, it will be very hard to advocate for the sorts of concepts that one thinks should play a certain functional role without drawing on rough attitudes that people already have to the relevant word itself. Paraphrasing in such a context will thus be a tactically less attractive move than it will be, say, in a philosophy seminar room where the participants' main goal is to get a grip on exactly what is at issue, rather than advocating some political or moral cause.[59]

§5. Methodological Implications.

We have argued that metalinguistic negotiation is a pervasive means of expressing disagreement and that this phenomenon extends to disputes involving normative or evaluative terms. We have argued further that this view is perfectly compatible with the view that such disputes evince genuine disagreement. If this argument has been successful, it means that philosophers are wrong to reason from the fact that a particular exchange involves a genuine disagreement to the thesis that the speakers involved in that exchange mean the same things by the words they use. Furthermore, it opens up an important tool in

59. Thanks to Bob Goodin for helpful discussion on these last few points.

accounting for the possibility of genuine disagreements on views that frequently are thought to run into trouble on precisely this front, including various forms of contextualism.

In concluding our main argument, we want to briefly canvass one final upshot of these conclusions. Many philosophers working on the semantics of normative and evaluative terms understand their task as figuring out how people with very different substantive views about the extension of a term such as 'moral obligation' or 'right' could nonetheless be genuinely disagreeing with each other when they have a dispute couched partly in those terms. In many cases, the task is taken to amount to this: figuring out a meaning of the term that could be shared by all parties to those disputes, a meaning that, despite the systematic variation in usage, allows for the conceptual coherence of their respective views of what falls into the extension of the term. Call this *The Shared Meaning Task* for a given term.

This is a difficult task: the theorist must thread a very fine needle, constructing an analysis which is substantive enough to differentiate the expression under analysis from other, nearby expressions, while simultaneously being so devoid of specific content as to plausibly be the meaning shared by speakers who apply the term in systematically differing ways. A great many theorists appear to understand their task in largely these terms, and to take the satisfaction of these two competing considerations to provide a framework for metanormative theory.[60]

But if our work is on the right track, then this methodology is seriously undermined. If we could be assured, ahead of time, that some subset of normative or evaluative disputes were canonical — that they expressed genuine disagreement *via* the literal expression of conflicting contents — then, perhaps, the Shared Meaning Task would accurately characterize the subsequent theoretical undertaking. But, as we have argued, many normative and evaluative disputes are simply not canonical. Hence, the widespread understanding of metanormative theory as centered on the Shared Meaning Task rests on a mistake.

Metanormative theorists should *not* assume that a given normative or evaluative dispute is canonical until there is independent reason to think it is. Instead, assuming that the dispute really does express a disagreement, theorists should hold simply that there is some rational conflict in the mental contents accepted by the speakers involved, a conflict which might or might not be reflected in the semantic content of the speakers' utterances. The theorist should look to the contents that this disagreement might be about (the surface of what it *seems* to be about might be misleading) as well as to the different mechanisms by which that conflict in attitudes might be expressed. Finally, once we have different proposals on the table for how to account for what is going on in a given exchange (proposals that will include both semantic and pragmatic elements), we can then evaluate the proposals for how well they do in accounting for the various factors that all sides can agree need to be accounted for — including facts about the feeling of conflict between the speakers, the linguistic features of the discourse, and the substantiveness, such as it is, of the debate.

If this is right, then it is not at all clear that philosophers working on the semantics of normative and evaluative terms will face the Shared Meaning Task for the term(s) in question. Indeed, it might *often* be that the speakers in question mean different things by their normative or evaluative words but that, nevertheless, their disagreement is entirely genuine. The metalinguistic analysis is not, after all, a theory that debunks our intuition of disagreement in the relevant cases; it *vindicates* our intuition that the disputes in question express genuine, worthwhile disagreements. While not every normative or evaluative dispute will turn out, on final analysis, to be a metalinguistic negotiation, we go seriously wrong in allowing the intuition of disagreement by itself to rule out those theories positing inter-personal, inter-contextual, or inter-theoretical variation in meaning. Thus, the Shared Meaning Task — unavoidable in cases where, antecedently, we know a dispute

60. For representative examples of this general line of thinking in metanormative theory, see (Gibbard 1990), (Gibbard 1992), (Smith 1994), and (Darwall 2002).

to be canonical — cannot play the debate-structuring role it often has, on pain of leaving our theoretical toolkit seriously impoverished.

§6. Objections and Replies.

§5 concludes our main line of argument in this paper. Recall, however, that one of our goals is this: to provide a more thorough and systematic defense of metalinguistic negotiation than has been offered in the metanormative literature. In order to make good on this, we consider and respond to three important objections to the thesis that many normative and evaluative disagreements are expressed via disputes that are best analyzed as metalinguistic negotiations. The first objection concerns externalism about contents or contexts. The second objection concerns explanatorily basic normative or evaluative terms. The third objection concerns the ability of the metalinguistic analysis to make accurate predictions regarding a seemingly related phenomenon — metalinguistic negation.

§6.1. Objection 1: Externalism.

In our presentation of the cases meant most clearly to exemplify metalinguistic negotiations — the *athlete* case and the *spicy* case — we have assumed that the relevant linguistic items in fact express different contents in the mouths of the respective speakers. That is, we assumed that in the mouth of the speaker of (5a), 'athlete' was truly predicated of horses, while in the mouth of the speaker of (5b), 'athlete' was not truly predicated of horses. Correspondingly, we assumed that, in Oscar's utterance of (4a), 'spicy' was, relative to the context, truly predicated of the chili, while in Callie's utterance of (4b), 'spicy' was, relative to the context, *not* truly predicated of the chili.

At this stage, those inclined to externalism either about linguistic content or about contexts may wish to object. After all, according to the externalist, what 'athlete' means does not depend on facts about the speaker alone, either her mental states or her actual patterns of usage. According to the externalist about content, while the mental states and facts about usage play some role in the determination of linguistic content, a major role is also played by, say, the causal-historical facts about the word 'athlete', or its meaning for the relevant speech community, or the contours of the corresponding natural kind. Correspondingly, according to an externalist about context, the contextually-salient threshold for being spicy — the setting of the contextual parameter that determines the meaning of 'spicy' relative to a context — is not merely a function of the intentions of the individual speakers, but rather is determined in part by objective features of the context they share. Content externalism and context externalism need not go together.[61] But a defender of either type of externalism may object to our presentation of the corresponding type of putatively-metalinguistic dispute. Given the role of factors outside the control or even acquaintance of the speakers, the relevant linguistic facts are, for the externalist, stabilized between the parties to the disagreements.

To this type of objection, we have two responses, one a dialectical clarification and the other a substantive point. First, the dialectical clarification. Though we have presented these cases as if the speakers in fact express different contents with their words, we do not ask the reader to assume that this is the correct analysis; that the speakers express different contents is not a *premise* in our argument. Rather, our presentation of the cases has the force of a supposition: even if we *suppose* that the speakers mean different things by their words, we can still explain how those disputes have the properties that they do. Perhaps the externalist can also explain how these disputes manage to express genuine disagreements. That's fine. But the metalinguistic analysis has its own account of those facts. There are independent reasons to be an externalist. (Whether they are decisive is another question.) But the considerations raised here show that those reasons do not include an inability of competing theories to explain what is going on in disputes like those above. An analysis making no mention of such meaning-stabilizing factors can, in the sense identified earlier in

61. For example, (Stalnaker 1999) advocates externalism about content, but not about context.

this paper, do an entirely successful job of explaining the possibility of disagreement in these cases.

Our more substantive response is this: any form of externalism — content or context — strong enough to address *any* plausible case of metalinguistic negotiation will be massively too strong. Thus, anyone, externalist or not, will require the resources to describe *some* disputes as metalinguistic negotiations in any case. The question is simply how far the analysis should extend. It *can* extend all the way to cases like the ones we have considered in §3.2 (*e.g.*, the cases involving 'spicy' and 'athlete'). We submit that to draw the line somewhere in the middle prior to investigating the details of the relevant cases will prove *ad hoc* at best.

To see that even the externalist requires the notion of metalinguistic negotiation, first consider the case of externalism about content. In the case about Secretariat, we considered two speakers who, aside from their usage of 'athlete', were linguistically very similar. But similar cases can be imagined between speakers of distinct, mutually comprehensible dialects. Consider a debate between the speakers of (9a) and (9b). They are describing a restaurant they've both visited. However, the speaker of (9a) is British, while the speaker of (9b) is American.

(9) (a) Burgers come with chips.

(b) No, burgers do not come with chips.

The dialogue in (9) is a clear case of what we can call a *merely verbal dispute*. The speaker of (9a) uses 'chips' to mean french fries, while the speaker of (9b) uses 'chips' to mean potato chips (or 'crisps', for the speaker of (9a)). Their dispute is premised on a linguistic misunderstanding and is for that reason pointless — it fails to reflect a disagreement, in our sense, at all. But there are other contexts in which such a dialogue might take place, contexts in which it is not at all pointless.

Imagine for instance that the dispute occurs between two speakers who know perfectly well that they are not from identical language communities, but who have found themselves working for the same catering company. The two employees know that burgers come with french fries, though they would describe that fact in different ways. They engage in the dispute in (9) in the context of an argument about how the menu is to be written up and how the dishes are to be described to their clients. In this case, there is a great deal hanging on whether 'chips' is used to refer to french fries or to crisps.

In this dispute, just as in the Secretariat case, there is mutual knowledge of all the relevant non-linguistic facts about the objects the speakers describe. And just as in that case, the speakers nevertheless persist in their dispute, and in a way that is in no obvious sense pointless or confused. What is most natural in this case is to say, just as we said in the Secretariat case, that the speakers disagree about how they should use the relevant term. Note, however, that for an externalist theory of linguistic content to reanalyze *this* case as a canonical dispute over the truth of literally expressed content would be to collapse the differences between two entirely distinct dialects of English. It is one thing to say that external factors conspire to keep meaning constant in cases where there are small differences or gradual changes in usage. It is another to say the same where there are profound, systematic, and community-wide differences in usage. The 'athlete' case and the 'chips' case have precisely the same structure. But no plausible form of externalism is strong enough to absorb cases of metalinguistic negotiation between speakers of distinct, mutually comprehensible dialects. Therefore, any theorist, externalist or not, must have the resources to describe metalinguistic negotiations in their toolkit.

Although the cases are somewhat different, similar considerations apply in the case of the context-externalist. Here, in order for the context externalist response to have traction, it must first be agreed that the relevant terminology is in fact context-sensitive — otherwise the dependence of context-sensitive linguistic items on objective features of the conversational context would be moot. Having noted that, the response to the challenge of context-externalism is entirely parallel to the response to the content-externalist. Just as metalinguistic negotiations about character can occur across differences in dialect, so too can

metalinguistic disputes about parameter settings (and thus contents) occur across contexts.

For example, if we suppose that contextualism about 'knows' is true, we must be able to offer some account of cross-context disagreement about knowledge attributions. The metalinguistic analysis is ready-made for this type of case: the speaker in the high-standards context faults the speaker in the low-standards context, not for literally expressing a false proposition, but for employing overly lax epistemic standards.[62] And it is one thing to suppose, in cases where speakers have slightly different beliefs about their shared conversational setting, that objective features of the context conspire to keep contextual parameter settings constant. It is another to say the same in cases where speakers are participants in entirely distinct conversational contexts. Given the need to separate out distinct conversational contexts — a requirement for defending contextualism in the first place — we must be able to describe certain disputes involving context-sensitive terminology as metalinguistic. The question, as before, is simply how far the phenomenon extends.

Of course, one may not buy into this analysis of disagreements about knowledge attributions in particular. Contextualism about 'knows' is hardly a consensus view. Indeed, one might reject epistemic contextualism precisely on the basis of cross-context disagreements about knowledge attributions. But recall that the objection under consideration here is levied by the *context*-externalist. That objection only works by first granting that the expressions in question are context-sensitive, and then submitting that objective features of the conversational context conspire to hold fixed the relevant parameter settings. Thus, for the advocate of the context-externalist objection, this move — the rejection of context sensitivity altogether — is not available. In fact, quite generally, disagreement-based arguments of the kind we consider in this paper are deployed *against* contextualisms of various stripes.[63] Thus, for a defender of those arguments to make use of the context-externalist objection is in many cases for them already to have given up the game.

Both the content- and the context-externalist will require the metalinguistic analysis for certain cases of disagreement — certain disagreements between speakers of mutually comprehensible dialects, or certain disagreements across different conversational contexts. The question then is simply how many disagreements that, *prima facie* are susceptible to this type of analysis are in fact best analyzed as metalinguistic. Once metalinguistic negotiation is in the picture, we find it quite natural to think that it extends to the type of cases considered above. In any case, to draw the line somewhere in between, and to have that line not appear disqualifyingly arbitrary, would require serious argumentation of a kind that does not appear in the disagreement-based arguments we have considered.

§6.2. Objection 2: Basic Normative or Evaluative Terms.

We argue in this paper for the viability of the metalinguistic analysis of a range of normative and evaluative disputes. Such disputes are, in our sense, non-canonical. A remaining issue concerns a difference among *types* of normative and evaluative terms. In some cases, there may be reason to think that the dispute in question is canonical after all based on the explanatory role that certain normative and evaluative terms have in accounting for the meaning of other normative and evaluative terms.

In some accounts of normative and evaluative semantics, the meanings of some normative and evaluative terms are analyzed in terms of the meanings of other normative and evaluative terms. Call those terms that are so analyzed *non-basic*. For instance, on a buck-passing account about the meaning of 'goodness', 'goodness' would be a non-basic evaluative term. In contrast, some normative or evaluative expressions will *not* be analyzed in terms of other normative or

62. (Chrisman 2007) makes a relevantly similar point about the same issue.

63. See, for instance, (Richard 2004), (MacFarlane 2007), (Smith 1994), (Boghossian 2006), (Egan 2006), and (Gibbard 1990).

evaluative terms. Call those expressions the *basic* normative or evaluative terms.[64] For instance, on some semantic accounts, 'rationality' will be analyzed in terms of 'goodness', and, in turn, 'goodness' will not be analyzed in terms of any other normative or evaluative term. On such accounts, 'goodness' is a basic term, while 'rationality' is non-basic.[65] For other accounts, however, it is precisely the reverse.[66]

The metalinguistic analysis has its clearest and most striking implications for work on the semantics of *non-basic* normative and evaluative terms, terms like 'torture', 'athlete', and 'moral obligation'. In contrast, the situation with basic normative or evaluative terms is more complex. To see why, consider that the metalinguistic negotiations we have discussed involve disagreements about what concepts we *should* be using or how we *ought* to be using words. Now suppose that we are asking these questions not about terms such as 'moral obligation' or 'coercion' or 'cruel' but rather such putatively basic normative terms as 'should' or 'ought'. If so, then the question we are asking arguably becomes self-reflexive: "how should we use the concept SHOULD?"[67] In those cases, a statement of the non-canonical forms of dispute we have canvassed will itself require the expression of the very concept that is under debate. Because of this, it seems that in order to make sense of the disagreement, one must take the dispute to be canonical — the alternative simply cannot be coherently stated, because the terms necessary to state it are not stable ground. So perhaps genuine disagreements involving basic normative or evaluative terms — whatever they turn out to be — must be canonical, and thus the meanings of those terms in particular must be shared for a disagreement to be genuine.

This sort of argument might give some force to Hare's original cannibals and missionary example, provided we think of it in terms of "goodness" *simpliciter* as opposed to something more specific such as "moral goodness".[68] The same might be true for certain of the moral terms used by Earthlings and Twin Earthlings. Moreover, we think this sort of restricted use of a disagreement-based argument is what lends much of the appeal to disagreement-based arguments for various forms of non-cognitivism about basic normative or evaluative terms, such as the sorts of disagreement-based arguments for expressivism that are at the center of some of Allan Gibbard's core arguments for expressivism in *Wise Choices, Apt Feelings* and *Thinking How to Live*.[69] We think this more restricted (and more sophisticated) version of the disagreement-based argument is worth exploring at greater length. However, there are a number of important points that must be underscored about this possibility.

The first point is simply that this more sophisticated version of the disagreement-based argument, even if it could be made to work, is much more limited in scope than the disagreement-based arguments typically on offer in the literature. In practice, proponents of the disagreement-based form of argument have freely applied versions of this argument to non-basic normative and evaluative terms. For instance, take the case of Gibbard. Gibbard not only uses a disagreement-based argument for conclusions about the semantics of basic normative and

64. The distinction between basic and non-basic normative (or evaluative) terms carries over to basic and non-basic normative (or evaluative) *concepts*. However, given our focus on semantics, and not on issues about the contents of thoughts, we put the distinction here in terms of words rather than concepts. One might o also use the basic/non-basic distinction to mark the same type of distinction among other things. For instance, one might use it to mark a distinction among normative (or evaluative) properties or among normative (or evaluative) facts.

65. See, for instance, (Railton 2003)

66. See, for instance, (Gibbard 1990).

67. In this paper, we follow one standard convention and designate concepts with smallcaps. On this convention, see (Margolis and Laurence 1999).

68. Hare himself is emphatic that his choice to focus on the "simplest, most typical, and most general value-words" is purely for "simplicity of exposition", and thus would be unlikely to go this route. (Hare 1991, 80).

69. See the opening chapters of both (Gibbard 1990) and (Gibbard 2003). Gibbard endorses the distinction that we've called "basic" vs. "non-basic". However, he shows no sign of thinking that the distinction is relevant to the applicability of disagreement-based arguments, as he employs such arguments indiscriminately to terms that are, for him, basic, and those that are not. Nevertheless, an intuitive grasp of the special force of disagreement-based arguments as they apply to basic terms may be why his arguments about those terms seem so compelling.

evaluative terms (the basis for his expressivism), but also uses the same form of disagreement-based argument for terms that he himself takes to be *non*-basic (such as 'morality' and 'moral right and wrong').[70] Or consider the case of Lasersohn that we introduced at the start of §2 — a case in which Lasersohn uses a disagreement-based argument for conclusions about the semantics of 'fun'. Lasersohn hardly means to suggest that 'fun' is an explanatorily basic term, a claim that would be questionable in it's own right. Thus, even if the more sophisticated form of disagreement-based argument we are considering here is on the right track — a form of argument that restricts the class of terms to explanatorily basic terms — then this does nothing to support *many* applications of the form of disagreement-based argument that we target.

To drive this point home, consider here Gibbard's account of non-basic normative and evaluative terms, such as 'jerk' or 'moral wrongness', which exemplifies a common pattern of explanation of such terms among expressivists. The story that Gibbard gives here — one that stays basically the same from *Wise Choices, Apt Feelings* to *Thinking How to Live* — is roughly this: the meaning of explanatorily non-basic normative and evaluative terms ultimately consists in a combination of the most explanatorily basic normative term (which, for Gibbard, is something along the lines of 'the thing to do') with further *such-and-such* descriptive application-conditions.[71] It is the first component that is the central normative kernel for Gibbard — it is having this kernel that makes a term normative at all. That kernel is what gets the expressivist analysis, and which is the main concern of most of Gibbard's work. Now suppose that Gibbard is able to run a more sophisticated version of the disagreement-based argument for whatever he thinks is the explanatorily basic normative term that ultimately matters here (*e.g.*, 'rational', 'the thing to do', or 'should'). Even if this argument works, that only tells you about the meaning of *one* normative or evaluative term. For all other terms, assuming that Gibbard's basic pattern of explaining their meaning is on the right track, you will only know something about one component of their meaning (namely, the normative kernel), but not about the other part (the such-and-such descriptive application-conditions). If one is going to explain explanatorily non-basic terms in the way that Gibbard does, this *other part* is crucial for explaining the meaning of explanatorily non-basic terms.

Gibbard uses disagreement-based arguments — of essentially the same form that he uses to settle the first, normative part — to settle on this other part of the meaning of explanatorily non-basic normative and evaluative terms.[72] But, if our argument in this paper is right, then his arguments here are in serious trouble. And saving the disagreement-based argument for the most explanatorily basic normative terms won't change that, even if Gibbard is right that *one component* of the meaning of explanatorily non-basic normative and evaluative terms is a central normative kernel (which is one single explanatorily basic normative term). So even if the more restricted version of the disagreement-based argument is onto something, it leaves untouched a very large part of normative and evaluative semantics.

The second point is that this more sophisticated version of the argument would require substantially more setup and argumentation than is typically on offer. After all, to *effectively* make use of it to establish the meaning of a given normative or evaluative term (rather than simply the explanatorily basic normative and evaluative terms, *whatever they are*), one must first establish which normative and evaluative terms are in fact basic, a matter which remains entirely unresolved. This is in part a reflection of the wide range of views that metaethical theorists have about which *concepts* are most explanatorily basic. For instance, some accounts (such as Gibbard's in *Wise Choices, Apt Feelings*) seek to understand the meaning of all normative and evaluative concepts in terms of a single basic normative concept, such as the concept RATIONAL or the concept WARRANT.[73] On such a view, the concept

70. See, for instance, (Gibbard 1992).

71. See (Gibbard 1990), (Gibbard 1992), and (Gibbard 2003).

72. See, for instance, (Gibbard 1990) and (Gibbard 1992).

73. See (Gibbard 1990) and (Gibbard 2003).

GOODNESS is analyzed in terms of RATIONAL. In contrast, other accounts (such as Peter Railton's) seek to understand the evaluative notion of VALUE as prior to that of normative notions such as WARRANT.[74] Other accounts still (such as Ralph Wedgwood's in *The Nature of Normativity*) hold that there are a number of basic normative and evaluative concepts that inter-define each other — or at least none of which can be seen as any more basic than the others.[75] This fact of the diversity of views here suggests that it cannot be taken for granted *which* normative or evaluative terms are basic ones. Figuring out which terms are basic in this sense is obviously no easy task. And, moreover, it is precisely the sort of task that we think cannot be pursued in isolation from thinking about the possibility that disputes involving these terms are non-canonical. This thus blocks a quick justification for concluding that speakers mean the same things by a specific normative or evaluative term (such as 'good' or 'right') that is *seemingly* a basic one. And, if that's right, then disagreement-based arguments even for *solely* the most explanatorily basic normative and evaluative terms — such as Gibbard's disagreement-based arguments for expressivism — are in fact much more complicated to make than it is normally thought, and can't be made effectively for *specific* normative and evaluative terms (*e.g.*, 'should', 'good', 'obligation' etc.) without seriously considering the metalinguistic option that we have put forward here. Which is to say: even if the restricted disagreement-based argument is helpful in establishing something about the explanatorily basic normative and evaluative terms (whatever they are), it won't help us much for actually doing the semantics of actual specified terms such as 'should', 'good', or 'obligation'.[76]

74. See (Railton 2003)

75. See (Wedgwood 2007).

76. Two things to note here. First, this upshot might not cause that much trouble for those who want to run a disagreement-based argument for conclusions about shared *concepts*, and who aren't worried about what (if anything) this means for the semantics of any particular terms. The core arguments of (Gibbard 2003) are arguably best read in this way. However, as we have made clear throughout, our focus in this paper is on disagreement-based arguments for *semantic* conclusions, arguments which Gibbard also gives, especially in his earlier work. The question of disagreement-based arguments for conclusions about shared *concepts* is a different topic, which we plan to explore in further work. Second, there is perhaps another option for how to account for disputes involving basic normative and evaluative terms. This is to grant that these metalinguistic disputes are about some antecedently unsettled question, but insist that we have identified the wrong question. On this front, one might insist that these disputes aren't about a normative topic in conceptual ethics about what concept *should* be used, but rather about a slightly different topic: namely, which concept *to* use. Is this really a viable alternative? Is it even an alternative at all? Part of the answer will depend on what the correct metanormative story is about how *what to do* questions are related to normative thought about what one *should* do. Perhaps, for instance, (Gibbard 2003) is right that the latter should be analyzed in terms of the former. Or, less strongly, perhaps one can't seriously advance a view about which concept to use (at least in the relevant contexts that matter for this discussion) without having some sort of normative or evaluative view about the merits of using that concept for the purposes at hand. In any case, we leave this idea to the side, given that we do not need to draw on it in order to advance our main line of argument. More generally, it should be emphasized that we fully support those wishing to give a more fine-grained taxonomy of metalinguistic disputes than we have done in this paper. We believe that there is much more work to be done on this front.

§6.3. Objection 3: Metalinguistic Negation.

Setting aside philosophical issues about the nature of disagreement, can a metalinguistic analysis make the right empirical predictions about linguistic phenomena? Can it predict that denial is licensed in the disputes in question? Can it predict that *the right kind* of denial is licensed? One thing that's clear is that ordinary canonical disputes license a particular conversational move: linguistic denial. In English, the speech act of denial can be marked with expressions such as 'nuh uh', 'nope', and variations of 'no it isn't'. For many linguists and philosophers, whether or not linguistic denial is felicitous is a crucial factor in discussions of disagreement. Indeed, for some, it is *the* crucial thing. For instance, Tamina Stephenson writes

> I should clarify that when I say that disagreement is possible in a certain dialogue, I mean very narrowly that expressions like *no (it isn't)* and *nuh-uh* are allowed. I don't mean that we have an intuition that the speakers disagree

> about something, which may be a broader phenomenon. I also don't mean that the disagreement is necessarily a rational or sensible one to engage in. The dialogues [about taste] are just the kind of arguments that are often pointed out to be futile, given that people's tastes simply differ. That is not my concern. The only fact that matters for my purposes is that such dialogues can and do occur — often enough, in fact, to give us ample opportunity to perceive their futility.[77]

As Stephenson points out, disputes about taste clearly license linguistic denial. So do all of the other normative and evaluative disputes we have considered here. Canonical disputes obviously allow for the use of the expressions Stephenson mentions. But we have proposed a different analysis of at least some such disputes. Can our analysis predict that denial is licensed in the relevant cases?

At first, the problem would not appear to be serious. If one is persuaded that examples like the dispute about the chili or the dispute about Secretariat are metalinguistic disputes, then it must simply be acknowledged as a datum that metalinguistic disputes (including metalinguistic negotiations) do license denial. If a metalinguistic analysis is proposed for some type of dispute, then it posits that that dispute is *like the chili case* or *like the Secretariat case*. Since in those cases denial is licensed, then an analysis claiming that some other dispute is relevantly similar is perfectly capable of predicting the felicity of denial in those cases.

That being said, one might wonder whether there is an even more-fine grained test in the vicinity. Consider that, in metalinguistic disputes, speakers disagree with content communicated by their interlocutors, but *not* the literally expressed content. Given this feature of the disputes, one might wonder why they do not license *metalinguistic negation* in the sense that Horn discusses in *A Natural History*

77. (Stephenson 2007, 493).

of Negation.[78] Metalinguistic negation is a special linguistic move — on Horn's theory, the exploitation of a pragmatic ambiguity in the negation operator — that allows speakers to object to some feature of an utterance *other* than its literal semantic content. For example, in response to (10a), an interlocutor might reply with (10b) if she felt that the proposition expressed in (10a) was false. But she might just as easily reply with (10c).

(10) (a) John saw some of his students at the party.

(b) Nuh uh, his students were all at a different party.

(c) John didn't see *some* of his students at the party; he saw *all* of them!

78. (Horn 1989). Not everyone agrees that metalinguistic negation is a genuinely distinct phenomenon (see (van der Sandt and Meier 2003)). If that were right, our case would be easier to make, so we do not stack the deck in our favor by granting Horn's analysis in this regard. In an important series of works, Robyn Carston, for example, argues that while there are distinctive uses of language to be observed in metalinguistic negations, they do not line up with the category Horn picks out and are best analyzed in other ways. (See (Carston 1996), (Carston and Noh 1996), and (Carston 2002) among others.) In particular, Carston (1) disputes Horn's claim that the negation operator is in any sense ambiguous, and (2) argues that what is crucial to metalinguistic negations is a more general phenomenon, namely the "echoic" use of an expression, in the sense of (Sperber and Wilson 1995). A full treatment of Carston's view (much less the full literature on this topic) is beyond the scope of this paper. But we note that without a strong further assumption — namely, that any denial that is in any sense "metalinguistic" should be predicted to also be echoic — these expressions (on Carston's analysis) present no particular problem for our analysis. Metalinguistic negotiations do not seem to involve echoic uses. Only when combined with that additional assumption would this suggest that they are therefore not metalinguistic after all. We see no reason to think this further assumption is true, and no reason to attribute such an assumption to Carston, who notes that "quotations, echoes, and other representations employed for purposes other than referring to, or describing, aspects of situations in the world are very common elements of verbal communication generally", and that "…this non-descriptive use of a representation may or may not be overtly signaled" (Carston 2002, 299). Thus, in addressing Horn's analysis specifically, we address the potential worry in its most worrisome form.

In (10c), metalinguistic negation provides the speaker with a way to indicate disagreement, not with the proposition asserted, but rather with a proposition implicated, namely that John did not see all of his students at the party. But the targets of metalinguistic negation are not limited to implicature. Metalinguistic negation can target information communicated along almost any non-literal dimension, including presuppositions ("He doesn't *regret* going to the party; he never went!"), connotations or other features of word choice ("He's not a smart *canuck*, he's a smart *Canadian*!") or even pronunciation ("It doesn't target [pro-*noun*-ciation], it targets [pro-*nun*-ciation]!").

Metalinguistic negations have, according to Horn, certain telltale markers. They tend to come in sequences consisting of a denial component ("She's not *smart...*") and a correction component ("she's a *genius*!"). The relevant expressions are marked off with focus intonation. (The distinctive emphasis placed on 'smart' and 'genius'.) They do not tolerate morphological embedding of the negation in the usual way:

(11) He's not *happy*; he's *ecstatic*!

(12) #He's *unhappy*; he's *ecstatic*![79]

And finally, the negations themselves do not license negative polarity items in the usual way.

(13) #John didn't see *any* of his students at the bar; he saw *all* of them there!

Each of these features of metalinguistic negations serves, on Horn's account, to differentiate them as a class from ordinary descriptive negations of the kind that appear in typical disputes over the truth of literally expressed propositional content. Thus, since metalinguistic disputes do not concern the truth of literally expressed content, perhaps we should expect them to license metalinguistic negation.

79. We follow standard convention within linguistics in using the '#' sign for marking infelicity.

However, no such difference obtains between the denials in canonical disputes and those that are licensed in metalinguistic ones. Think back to the denials that are licensed in the disputes we have called metalinguistic. ("Nuh uh, Secretariat is not an athlete.") Those denials show none of the signs of metalinguistic negation. Denials in these settings show no pressure towards appearing in the form of a denial-correction sequence; they do not require special intonation; they happily tolerate morphological incorporation of the negation ("Nuh uh, Secretariat is a non-athlete!"), and if there are polarity items in the sentence, they behave in all the usual ways. (For example, in the typical licensing of the NPI 'ever' in (14b).)

(14) (a) Secretariat has always been an athlete.

(b) Nope, Secretariat has not ever been an athlete; 'athlete' only applies to people!

The fact that this dispute does not license metalinguistic negation shows why appeal to metalinguistic negation cannot serve as a quick test for separating out canonical disputes from metalinguistic ones. However, it also creates a potential problem for our analysis.

Disputes over information of almost any other non-literal type of content — implicature, presupposition, connotation, even manner — all license metalinguistic negation. That pattern might naturally lead one to think that the metalinguistic analysis of chili- and Secretariat-type disputes is committed to the false prediction that those disputes *should* license metalinguistic negation, when in fact the denials they license seem for all the world to employ perfectly ordinary descriptive negations. So perhaps the failure of these disputes to license metalinguistic negation is a reason to think that the disputes themselves are not metalinguistic after all!

However, a more careful look at the details of metalinguistic negation reveals that in fact no such prediction is generated. To see why, consider a case of relevance implicature:

(15) Sally was able to solve the problem.

In (15), the proposition literally expressed is that Sally had the ability to solve the problem. What's implicated of course is that Sally in fact solved it. But suppose a listener objected to (15), not because she disagrees with the literal content. (She *agrees* that Sally had the ability to solve the problem.) Rather our interlocutor disagrees with the implicated content, that Sally actually solved the problem. This should be an ideal case for metalinguistic negation. Even in this position, however, our interlocutor cannot respond using metalinguistic negation:

(16) #Sally wasn't *able* to solve the problem; she didn't solve it.

Whatever it does mean, (16) cannot be read as saying that Sally *was* able to solve the problem but refrained from solving it for other reasons. The best our interlocutor can do in objecting to (15) is employ ordinary descriptive negation *as if* she were responding to the literal content.

Why should relevance implicatures fail to license metalinguistic negation in situations precisely parallel to those where implicatures of other kinds do license metalinguistic negation? Horn argues that the reason concerns the logical relationship between the two components of typical metalinguistic negations. That relationship is inconsistency. Consider again the metalinguistic negation in (10c).

(10) (c) John didn't see *some* of his students at the party; he saw *all* of them!

If the denial component and the correction component of this sequence were each interpreted descriptively, then the two would be logically inconsistent. If John did not see some of his students at the party, then he could not have seen all of them there. Likewise, if he isn't happy, then he can't be ecstatic, and if Sally is not smart, she can't be a genius.

Horn's suggestion is that it is this feature of metalinguistic negation that allows for it to occur. When a listener attempts to parse the sequence descriptively, she immediately realizes that if it were read that way, then the speaker will have said something contradictory. This forces the listener to go back and reparse the denial component in such a way that the negation attaches to some feature of the utterance other than its literal content. However, where there is no inconsistency between the denial and correction components of the sequence even on a descriptive reading, no reparsing is necessary, and metalinguistic negation will be impossible. The reason metalinguistic negation is not licensed in denials of relevance implicatures is that there is no inconsistency between an ordinary descriptive denial of the expression uttered and the correction of the implicated content: Even on a perfectly ordinary reading of the negation, there is no inconsistency in saying that Sally wasn't able to solve the problem and saying that she did not solve the problem.

For us, what is crucial about this explanation is that it applies equally well to denials in the context of metalinguistic disputes. Consider again a dispute over relevance implicature:

(17) (a) Sally was able to solve the problem.

(b) Sally wasn't able to solve the problem; she didn't solve it.

(17b) cannot be interpreted as employing metalinguistic negation because the denial component and the correction component are consistent, even on a descriptive reading of the negation. This is exactly the same structure as the denials in metalinguistic disputes:

(18) (a) Secretariat is an athlete.

(b) Secretariat is not an athlete; 'athlete' can only apply to humans.

Whatever their actual truth, the propositions that Secretariat is not an athlete and that 'athlete' can only be predicated of humans are entirely consistent. Thus, even if the speaker of (18b) is correctly interpreted as objecting to some feature of (18a) other than its literal semantic content, she cannot express that objection via the use of

metalinguistic negation. Metalinguistic negation is *never* licensed in denials with this structure.

The denials that appear in metalinguistic disputes pattern logically with denials of relevance implicatures rather than with denials of other types of implicature or other non-semantic features of an utterance. Thus an analysis of these disputes *as* metalinguistic is in no way committed to the false prediction that they should license metalinguistic negation. A metalinguistic analysis of chili- and Secretariat-type disputes predicts that they should license perfectly ordinary descriptive denials. And that is precisely the right prediction to make.

This argument about metalinguistic negation underscores a general theme of this paper. We have argued that metalinguistic negotiations about normative or evaluative terms are capable of instantiating *many* of the properties that theorists tend to associate with canonical disputes involving normative or evaluative terms. For instance: they are capable of expressing genuine disagreements; they can be very much worth having (in multiple senses of "worth having"); the speakers involved in such negotiations take themselves to be at odds in perfectly ordinary ways; and they license ordinary descriptive — rather than metalinguistic — negation. One might therefore ask: how, then, do we tell whether a dispute is a metalinguistic negotiation or a canonical dispute? There is no simple answer to this question. Importantly, there is no way to answer it in abstraction from a more general view of how to do semantics — which will include, crucially, issues about how one understands the semantics/pragmatics divide.

Different theorists with different commitments on this front will have different answers for how to distinguish canonical disputes from metalinguistic negotiations. We have not offered any quick or easy answers to that, nor could anyone without delving into the details of a particular case and offering up their own package deal of semantics and pragmatics. Rather, if our work in this paper has been on the right track, we hope to have convinced those working on the semantics of normative and evaluative terms that it is a question that must be asked.

Works Cited

Ayer, A.J. 1971. *Language, Truth, and Logic*. New York: Penguin Books.

Barker, Chris. 2002. The dynamics of vagueness. Linguistics and Philosophy, 25, 1–36. *Linguistics and Philosophy* 25:1–36.

———. 2013. Negotiating Taste. *Inquiry* 56 (2–3):240–257.

Björnsson, Gunnar, and Stephen Finlay. 2010. Metaethical Contextualism Defended. *Ethics* 121 (1):7–36.

Boghossian, Paul Artin. 2006. *Fear of Knowledge : Against Relativism and Constructivism*. Oxford: Oxford University Press.

Burge, Tyler. 1979. Sinning Against Frege. *The Philosophical Review* 88:398–432.

Burgess, Alexis, and David Plunkett. 2013a. Conceptual Ethics I. *Philosophy Compass* 8 (12), 1091-1101.

———. 2013b. Conceptual Ethics II. *Philosophy Compass* 8 (12), 1102-1110.

Cappelen, Herman, and John Hawthorne. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Carston, Robyn. 1996. Metalinguistic negation and echoic use. *Journal of Pragmatics* 25:309–30.

———. 2002. *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell.

Carston, Robyn, and Eun-Ju Noh. 1996. Metalinguistic negation is truth-functional negation, with evidence from Korean. *Language Sciences* 18:485–504.

Chalmers, David J. 2011. Verbal Disputes. *Philosophical Review* 120 (4):515–566.

Chierchia, Gennaro. 2004. Polarity Phenomena and the Syntax/Pragmatics Interface. In *Structures and Beyond*, edited by A. Belletti. Oxford: Oxford University Press.

Chrisman, Matthew. 2007. From Epistemic Contextualism to Epistemic Expressivism. *Philosophical Studies* 135 (2):225–254.

Copp, David. 2000. Milk, Honey, and the Good Life on Moral Twin Earth. *Synthese* 124 (1):113–137.

Darwall, Stephen L. 2002. *Welfare and rational care*. Princeton, N.J.: Princeton University Press.

Darwall, Stephen L., Allan Gibbard, and Peter Albert Railton. 1997. Toward Fin de siecle Ethics: Some Trends. In *Moral Discourse and Practice : Some Philosophical Approaches*. New York: Oxford University Press.

DeRose, Keith. 2004. Single Scoreboard Semantics. *Philosophical Studies* 119 (1/2):1–21.

Dreier, James. 1990. Internalism and Speaker Relativism. *Ethics: An International Journal of Social, Political, and Legal Philosophy* 101 (1):6–25.

Dworkin, Ronald. 2011. *Justice for Hedgehogs*. Cambridge: Harvard University Press.

Egan, Andy. 2006. Epistemic Modals, Relativism, and Assertion. *Philosophical Studies* 133 (1):1-–22.

———. 2012. Relativist Dispositional Theories of Value. *The Southern Journal of Philosophy* 50 (4):557–582.

Egan, Andy, John Hawthorne, and Brian Weatherson. 2005. Epistemic Modals in Context. In *Contextualism in Philosophy*, edited by G. Preyer and G. Peter. Oxford: Oxford University Press.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings : A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press.

———. 1992. Moral Concepts: Substance and Sentiment. *Nous Supplement: Philosophical Perspectives* 6:199–221.

———. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.

Hare, R.M. 1991. *The Language of Morals*. Oxford: Oxford University Press.

Harman, Gilbert. 2000. *Explaining Value: And Other Essays in Moral Philosophy*. Oxford: Oxford University Press.

Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.

Horgan, Terry, and Mark Timmons. 1993. New Wave Moral Realism Meets Moral Twin Earth. In *Rationality, Morality, and Self-Interest*, edited by J. Heil. Lanham, Md.: Rowan and Littlefield Publishers.

Horn, Laurence R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.

Kaplan, David. 1989. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives. In *Themes From Kaplan*, ed. by J. Almog and J. Perry. Oxford: Oxford University Press.

Kennedy, Chris. 2007. Vagueness and grammar: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy* 30 (1):1–45.

Lasersohn, Peter. 2005. Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy* 28 (6):643–686.

Lewis, David. 1979. Scorekeeping in a Language Game. *Journal of Philosophical Logic* 8:339–59.

Ludlow, Peter. 2008. Cheap contextualism. *Philosophical Issues* 18:104–129.

———. Manuscript. *The Dynamic Lexicon*.

MacFarlane, John. 2007. Relativism and Disagreement. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 132 (1):17–31.

———. In Progress. *Assessment Sensitivity: Relative Truth and its Applications*.

Manley, David. 2009. Introduction: A Guided of Metametaphysics. In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by D. J. Chalmers, D. Manley and R. Wasserman. Oxford: Oxford University Press.

Margolis, Eric, and Stephen Laurence. 1999. *Concepts : core readings*. Cambridge, Mass.: MIT Press.

Merli, David. 2002. Return to Moral Twin Earth. *Canadian Journal of Philosophy* 32 (2):207–240.

Moore, G.E. 1970. The Nature of Moral Philosophy. In *Philosophical Studies*. London: Routledge and Kegan Paul Ltd.

———. 1993. *Principia Ethica*. Cambridge: Cambridge University Press.

Plunkett, David, and Tim Sundell. Forthcoming-a. Antipositivist Arguments from Legal Thought and Talk: The Metalinguistic Response.

In *Pragmatism, Law, and Language*, edited by G. Hubb and D. Lind: Routledge.

———. Forthcoming-b. Dworkin's Interpretivism and the Pragmatics of Legal Disputes. *Legal Theory*.

Railton, Peter. 2003. *Facts, Values, and Norms : Essays Toward a Morality of Consequence*. New York: Cambridge University Press.

Richard, Mark. 2004. Contextualism and Relativism. *Philosophical Studies* 119 (1):215–242.

Ridge, Michael. In Progress. Impassioned Belief.

Robinson, Denis. 2009. "Moral Functionalism, Ethical Quasi-Relativism, and the Canberra Plan". In *Conceptual Analysis and Philosophical Naturalism*, edited by D. Braddon-Mitchell and R. Nola. Cambridge, MA: MIT Press/Bradford, pp. 315–348.

Sider, Theodore. 2012. *Writing the Book of the World*. Oxford: Oxford University Press.

Smith, Michael. 1994. *The Moral Problem*. Cambridge: Blackwell.

Sperber, Dan, and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Cambridge: Harvard University Press.

Stalnaker, Robert. 1999. *Context and Content : Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stephenson, Tamina. 2007. Judge Dependence, Epistemic Modals, and Predicates of Personal Taste. *Linguistics and Philosophy* 30 (4):487–525.

Stevenson, Charles L. 1937. The Emotive Meaning of Ethical Terms. *Mind* 46 (181):14–31.

———. 1938. Persuasive Definitions. *Mind* 47 (187):331–350.

Sundell, Tim. 2011a. Disagreement, Error, and an Alternative to Reference Magnetism. *Australasian Journal of Philosophy* 90 (4):743–759.

———. 2011b. Disagreements about taste. *Philosophical Studies* 155 (2):267–288.

U.S. Department of Justice, Office of the Assistant Attorney General. 2002. 'Memorandum for Alberto R. Gonzales, Counsel to the President, Re: Standards of Conduct for Interrogation under 18 U.S.C. xx 2340–2'

United Nations, General Assembly. 1984. 'Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment.' United Nations, Treaty Series 1465.

van der Sandt, Rob, and E Meier. 2003. Denials in discourse. Paper read at Michigan linguistics and philosophy workshop (ms., University of Nijmegen).

Väyrynen, Pekka. 2013. *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics*. Oxford: Oxford University Press.

Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.