

Intuitions about Disagreement Do Not Support the Normativity of Meaning
Derek BAKER*

This is the penultimate draft of a paper forthcoming in *dialectica*. Please quote and cite the published version.

ABSTRACT

Allan Gibbard (2012) argues that the term ‘meaning’ expresses a normative concept, primarily on the basis of arguments that parallel Moore’s famous Open Question Argument. This paper argues that Gibbard’s evidence for normativity rests on idiosyncrasies of the Open Question Argument, and that when we use related thought experiments designed to bring out unusual semantic intuitions associated with normative terms we fail to find such evidence. These thought experiments, moreover, strongly suggest there are basic requirements for a theory of meaning incompatible with Gibbard’s ultimate goal of providing an expressivist account of meaning-related concepts. The paper concludes by considering a possible way in which meaning could be normative, consistent with the intuitions about disagreement; but this form of normativism about meaning appears incompatible with Gibbard’s expressivism.

1. Introduction

In his (2012) book, *Meaning and Normativity*, Allan Gibbard argues that the term ‘meaning’ expresses a normative concept.¹ His primary argument is that the term seems to have the openness that G.E. Moore identified as characteristic of ‘good’. We have intuitions about cases of semantic disagreement very similar to the intuitions that surround normative disagreement; and so, on grounds of explanatory simplicity, we should conclude that MEANING is a normative concept.²

But there are more unusual semantic intuitions associated with normative terms: the intuitions about disagreement in the famous Missionary and Cannibals example (Hare, 1952) and the Moral Twin Earth case (Horgan and Timmons, 1991). Similar thought experiments applied to ‘meaning’ fail to elicit analogous intuitions. In other words, a more complete investigation of our semantic intuitions fails to reveal evidence that the concept MEANING is normative.

* Department of Philosophy, Lingnan University, 8 Castle Peak Road, Tuen Mun, N.T., Hong Kong SAR, email: derekbaker@ln.edu.hk.

¹ For other examples of the normativity of meaning thesis, see (Kripke, 1982; Boghossian, 1989; Brandom, 1994; and Whiting, 2007).

² This paper will follow the convention of naming concepts with small capital letters.

³ For other examples of the normativity of meaning thesis, see (Kripke, 1982; Boghossian, 1989; Brandom, 1994; and Whiting, 2007).
Dispositions* here should be read as a placeholder for a more systematic account, which characterizes

² This paper will follow the convention of naming concepts with small capital letters.

Intuitions about disagreement and associated Open Question-style considerations are Gibbard's primary argument in favor of MEANING's normativity. Consequently, the success of this paper's argument would be especially troubling for Gibbard's overall project—to offer an expressivist account of metasemantic theories—as it would undermine his primary motivation for pursuing the project in the first place.

Gibbard offers other arguments, but they are considerably less compelling. First, the normativity of MEANING would explain why naturalistic reductions of the notion have been so hard to arrive at (16-18). Second, it would explain why an agent ought (in some sense) not to believe that snow is white and nothing is white (15-16). But reductions are always difficult; and even if none are available, meanings could be non-natural entities without being normative entities, if they are Fregean senses, for example (Glüer and Wikforss, 2009). As for the second consideration, it could be explained by any number of other things: epistemic norms according to which evidence for p is evidence against $\sim p$, or coherence constraints on attitudes.

In the course of addressing Gibbard's primary argument, two general morals will emerge. First, plausible versions of the normativity of meaning thesis must appeal to normative notions that are generally understood to play an explanatory role: an example of a version that meets this restriction will be provided. Second, this restriction makes it very unlikely that normativism about meaning is true, *if* we assume Gibbard's brand of expressivism about the normative.

In short, the paper will argue that Gibbard fails to demonstrate the normativity of meaning. Meaning may still be normative, of course. But the paper will also present a significant objection to Gibbard's particular version of the normativity of meaning thesis. Unless this objection can be answered, Gibbard's expressivist interpretation of metasemantic theories is untenable.

1. Gibbard's Argument

Gibbard asks to imagine two philosophers. Jerry, an individualist about meaning, thinks meaning facts are determined by the psychological dispositions of the speaker; Tyler, a communitarian about meaning, by the dispositions of the speaker's community.³

Initially Gibbard stipulates that Jerry and Tyler share his belief that MEANING is normative (39-40). However, Gibbard (41) moves to explaining why we should interpret them as sharing this theoretical commitment, by calling our attention to the following problem. Meaning facts unquestionably supervene on non-meaning facts. But both philosophers could completely agree about all of the non-meaning facts (i.e., facts about psychology, sociology, causal relations with the environment) and still disagree about

³ "Dispositions" here should be read as a placeholder for a more systematic account, which characterizes the relevant dispositions in terms of causal regulation (Boyd, 1997), asymmetric dependence (Fodor, 1994b), function in an overall rationalizing interpretation (Lewis, 1974), etc.

meaning. Nonetheless, we don't hear them as having a merely verbal disagreement; rather, the disagreement is genuine (Gibbard 2012, 41-45).⁴

The analogy with the normative case *can* be striking. It is widely accepted that the normative supervenes on the non-normative. Nonetheless, two people could agree on all of the non-normative facts, but still find themselves in normative disagreement; and we do not hear this as a merely verbal dispute, but as genuine disagreement.

The Open Question Argument (Moore, 1903/1993) famously calls attention to the fact that for any *X*, 'Is *X* good?' seems like a legitimate question, which does not betray linguistic incompetence. Nonetheless, there must be *some X* on which goodness

⁴ As John MacFarlane notes, the question of what it is for two people to genuinely disagree "is surprisingly difficult to answer" (2007, 17-18), and Gibbard himself argues that an expressivist should take disagreement as her semantically primitive relation—rather than, for example, truth and truth-conditions (2003). I will thus rely on an intuitive sense of the distinction between genuine disagreements and merely verbal disagreement rather than any more precise theory. For a straightforward example of a merely verbal disagreement, we can imagine a US and a UK national arguing over whether there are any *chips* on the plate, or arguing over the rules of *football*. What we have here is a case where vocabulary creates the illusion that the speakers are contradicting each other, but in fact it is perfectly possible for the plate to contain what the British call 'chips' and yet fail to contain what Americans call 'chips'.

The most natural way of characterizing genuine disagreement is in terms of accepting claims which would be incompatible when jointly asserted. Two people genuinely disagree, just in case one accepts that *p* and the other accepts that *q*, and *p* and *q* could not both be true. MacFarlane notes problems with this definition, but maintains it is the best working account at hand. This characterization has the advantage of neutrality: for example, it can be easily modified to fit the thesis that in certain domains *faultless disagreement* is possible, or disagreement in which the disputants are correct relative to their own context of assessment, but incorrect relative to their rivals' contexts of assessment (MacFarlane, 2007). For the significance of truth-relativism to cases of *normative disagreement*, see (Schafer, 2014).

This account may be thought to rule out semantic expressivism, since it characterizes contents in terms of truth-evaluable propositions rather than in terms of non-cognitive attitudes. This probably overstates the worry, given the quasi-realist ambitions of most modern expressivists; though see (Richard, 2008) for cases where the worry may be apt. In any case, it is generally accepted that expressivists need to offer some account of how expressed attitudes are incompatible in a way that mimics semantic inconsistency among propositions, if they are to solve the Frege-Geach problem. For a variety of accounts of this incompatibility relation and how it might mimic semantic inconsistency, see (Gibbard, 2003; M. Schroeder, 2008; Ridge, 2014; and Baker and Woods, 2015).

Alternately, David Chalmers (2011, 522) offers the following working account: "A dispute over *S* is (broadly) verbal when, for some expression *T* in *S*, the parties disagree about the meaning of *T*, and the dispute over *S* arises wholly in virtue of this disagreement regarding *T*." This has the advantage of allowing us to identify disagreements as genuine or verbal without reference to truth. On the other hand, it may produce false-positives. Two parties may drop their dispute over *S* upon acquiring more accurate beliefs about the meanings of terms, simply because they no longer care about the matter. (Maybe you and I really do have incompatible beliefs about whether Napoleon crossed the alps, but upon learning that 'Napoleon' was not the name of a famous Hollywood actor, drop the discussion in order to make more time for talking about Jared Leto.) In any case, Chalmers writes of his account: "We can ... see the characterization as pointing us toward a salient and familiar phenomenon, rather than delineating its contours precisely" (525).

Thanks to an anonymous referee for asking me to clarify the distinction between genuine and merely verbal disagreement.

supervenies. Gibbard's disagreement scenario is supposed to show that 'meaning' has the same kind of *openness* as 'good' (2012, 7-9). In other words, 'meaning' is a term which is (1) known by competent speakers to pick out supervenient properties, but (2) for which genuine (and not merely verbal) disagreements about correct application can persist despite agreement about all the relevant subvening facts.

Gibbard concludes from this Moorean openness that the term 'meaning' expresses a normative concept. The argument is abductive. We can explain why 'meaning' displays (1) and (2) by concluding that it is a special case of the family of terms most associated with such openness—normative terms.

Moreover, this argument has implications for the *nature* of normativity. Utilizers of the Open Question Argument have typically presented the conjunction of (1) and (2) as a mystery in need of explanation; they then reconcile the pair by proposing either that 'good' refers to an irreducibly normative property necessarily but not analytically connected with the supervenience base, or that 'good' is not a normal referring expression. Gibbard puts his thought experiment to the same work. He ultimately favors the latter *expressivist* account for MEANING; but he allows that his arguments could also be used to support a *non-naturalist* account of meaning instead (Gibbard 2012, 18-21).

So Gibbard uses Moorean openness, first, as evidence that semantic terms express normative concepts, and second, as evidence that they either refer to non-natural properties or do not refer at all. But Moorean openness seems to be the norm for most of the terms in our language, normative or not. So it cannot be evidence of normativity. What's more, there are several well-known debunking explanations of the apparent 'openness' of normative terms and concepts.⁵ So it cannot serve as evidence of non-naturalism or expressivism about meaning, either.

With regard to the first point, philosophers could know all the relevant facts about a human organism's conception, birth, first word, first sentence, first self-reflective thought, first long-term plan, and so on, and still disagree about when the *person* started to exist. Similar things could be said about the persistence of any mereologically complex object. Jamin Asay (2013, §6.1) shows how debates about truth often make appeal to Open Question Arguments as a reason to reject one theory or another. It is even an open question whether an unmarried male is a bachelor, if that male is the Pope, for example. One disputant could think our reluctance to call the Pope a bachelor pragmatic; the other could think he falls outside the extension of the concept. But the terms 'identity', 'truth', and 'bachelor' are not normative, or at least we would need more argument before we concluded that they were.

We can also sketch out two general explanations for the Moorean openness of natural language terms that make no appeal to expressivism or non-natural supervening properties. First, the meaning of a term may be fixed to a greater or lesser degree by facts *external* to a speaker's psychology, and hence inaccessible to her. Most famously, 'water'

⁵ For a survey of these problems see (Miller 2003, 15-8; and McPherson, 2013).

means H₂O; but because this rests on fairly arcane empirical facts, not all competent speakers could know it. If meanings are complex brain-states or complex dispositional properties of communities, it is unsurprising that competent speakers can find themselves in disagreement about the term's correct application.

Gibbard explicitly embraces a more internalist semantics, one with a strong distinction between *sense* and *reference* and a clear identification of meanings with senses (26ff.). He would simply deny that 'water' means H₂O, at least in the sense of 'meaning' that interests him. 'Water' presumably means something like WET AND CLEAR AND TASTELESS AND POTABLE STUFF AROUND HERE.

Yet even if meanings are psychologically "internal" concepts or senses, concepts may not be transparent to their possessors (Jackson 1998, Chapters 2-3; and 2004, 272ff).⁶ That is, *deploying* a concept in accurate judgments does not obviously depend on accurate judgments *about* that concept. This is effectively the lesson of the paradox of analysis. This paradox draws on the fact that an analysis is an account of the meaning of some term with which we are familiar. The sort of familiarity involved in all the classic cases of analysis is that of knowing what the word means. (It is presupposed that we know what 'knows' means, for example, when we try to give an analysis; the analyzer isn't presuming to *teach* us a new word.) But if we know what the word means, an account of its meaning, *an analysis*, could not be informative. Yet it seems that analyses can be informative—we have all encountered seemingly informative analyses.

How can this be? A simple resolution of the paradox identifies it as relying on the fallacy of equivocation. All of us can successfully follow the grammatical rules of our native language and identify sentences that violate those rules, even if we are unable to say what those rules are (Jackson 1998, Chapter 3; and Schroeter and Schroeter, 2009). Similarly, we can correctly apply concepts and terms without being able to *explain how* we identify instances. Informative analyses, then, are informative because they make *implicit* knowledge *explicit*—a distinction that the paradox ignores (Miller 2003, 16-17).

This solution to the paradox has intuitive appeal. My three-year-old is competent with the term 'water', but I doubt she could list the term's *a priori* criteria. I am not even confident in my own description of the sense of 'water'—it strikes me (and was intended as) an approximation of the term's meaning. In any case, Moorean openness could simply be a reflection of the fact that two speakers could express the same concept, but have different theories on how the concept should be analyzed, just as they could speak the same language while having different theories about its grammatical rules.⁷

⁶ Also see (Schroeter and Schroeter, 2009) for a helpful summary of recent internalist semantics and their application to specifically normative terms, along with criticism.

⁷ Gibbard acknowledges that his own argument risks exploiting the paradox of analysis (41). But he does not explicitly consider the possibility just presented—the possibility of genuine disputes about the correct analysis. He notes that his theory would tell us "what is at issue" between Jerry and Tyler: they disagree about which sentences a speaker ought to accept (44). But the alternate hypothesis also tells us what is at issue between the two—the correct analysis of the shared term 'means'.

It should be noted that whatever we think of either explanation of Moorean openness, Gibbard's argument is fallacious. First, as the above examples showed, Moorean openness is a property of terms not plausibly thought of as normative; thus it is not evidence of normativity. Second, consider again the paradox of analysis. Gibbard concludes from the fact that a term is open that it admits of no analysis (a fact that can then be given a non-naturalist or expressivist gloss). But if analyses can be informative, this inference is straightforwardly fallacious. On the other hand, if Gibbard denies that analyses can be informative, he undermines the plausibility of his own positive view. Competent speakers can disagree about the truth of expressivism about semantic terms (or non-naturalism about the referent of those terms), after all.

In the next section we will see that, when we look at other thought experiments designed to test intuitions of disagreement, we find independent reason to conclude that Gibbard's test exploits idiosyncrasies of the open question argument instead of indicating unusual features of the term 'means'. At the very least, 'means' lacks some unusual features of normative terms. It makes most sense, then, to think that in this case the Moorean test reveals a property shared by normative and non-normative terms alike, rather than any evidence of normativity.

2. Successors to the Open Question Argument

In the century since Moore put forward the Open Question Arguments, substantially better arguments from disagreement have been offered in support of non-naturalism and expressivism.⁸ These identify intuitions that depart from standard responses to non-normative referring expressions, and this divergence cannot be easily explained by invoking the possibility of either externalism or informative semantic theses about familiar terms.⁹ I will consider both thought experiments that have served to illustrate the unusual intuitions about disagreement surrounding moral terms; then I will construct analogous thought-experiments regarding semantic terms, in order to test for evidence that these terms are normative. 'Meaning', we will see, fails both tests.

⁸ Michael Ridge (2015) makes the plausible claim that these latter thought experiments, the Missionary and Cannibals and Moral Twin Earth, are descendants of the Open Question Argument, designed to get at the same basic point that Moore was reaching at. See (Ridge 2014, 71-76 and 89ff.) for discussion of the significance of this point in metaethical debate.

⁹ But see (Foot, 1959; and Dowell, *forthcoming*) for extensive criticism of these thought experiments; also see (Plunkett and Sundell 2013) for what is effectively a debunking account of these intuitions. I will not address these criticisms in this paper, for the simple reason that if they are successful, Gibbard's case is even more hopeless. His argument for the normativity of meaning, remember, is an abductive argument that proceeds in three steps: (1) there are unusual intuitions associated with semantic terms; (2) the same sorts of unusual intuitions regarding disagreement are standard for normative terms; and (3) the best explanation of (1) and (2) is that semantic terms are a species of normative terms. If (2) simply turns out to be false or to be a case of misleading evidence, Gibbard's argument is defeated, albeit for different reasons than those presented here (see also footnotes 16 and 25).

In R.M. Hare's example of the Missionary and the Cannibals, a missionary visits an island of cannibals and begins to translate their language (1952, 148-149). By remarkable coincidence they have words that sound exactly like 'good' and 'bad', and they seem to use them in evaluative contexts. But all of the things the missionary calls 'good', being helpful, humble, and forgiving his enemies, the cannibals call 'bad'. On the other hand, everything he tells the cannibals is bad, enslaving one's rivals, lording it over others, and collecting many skulls, they call 'good'. The missionary notices, however, that in another respect the cannibals use the sounds 'good' and 'bad' much as he does. 'Good' is associated with recommendation, choice, and rewards; 'bad' with the opposite.

The question is whether it would be appropriate for the missionary to translate their word 'good' into the English 'good', or whether it actually means bad. If the cannibals call some act of successful treachery 'good', and the missionary calls it 'bad', are they disagreeing with each other, or is the missionary in fact in agreement with them, a fact which he has disguised by slipping into his own language? It is natural to hear the disagreement as genuine, and to think that the cannibal word 'good' means GOOD. The cannibals just have radically different views on which things are good.

But if the cannibals called in their language 'red' everything that appears green, and 'green' everything that appears red, it would be more natural to simply think that these words, in their language, had different meanings from ours. A general rule in translation seems to be, all else being equal, one should posit that two words have the same meaning only if their applications by competent speakers are roughly coextensive.

This principle is supported, and arguably predicted, by both of the broad accounts of meaning—externalist and internalist—considered in the previous section. If meaning just is reference, then the meaning of a predicate will be some property. If two speakers application of a term is not roughly coextensive, then either one of them is using the term to pick out a different property (and hence with a different meaning), or at least one of the speakers is systematically and pervasively misapplying the word, which will generally mean the speaker is not competent. In the specific example, we are imagining that all members of the community apply 'green' to red objects. Even an extremely minimal version of the Principle of Charity (cf. Davidson, 1970) would leave us with a strong presumption against treating the *entire linguistic community* as incompetent with its own terms.¹⁰

The more internalist picture predicts the same. If MALE is part of the sense of 'bachelor' as spoken by me, then anyone who typically applies the term to females is using the term to express a different sense, all else being equal. 'Good' and 'bad' seem strange, then, in that this general rule of translation, supported by both general models of

¹⁰ A similar point is made in (Wedgwood 2007, 165-166).

meaning, does not seem to apply to them.¹¹ Both of these models allow that there can be informative semantic disputes, as we saw in the last section; and so Moorean Openness could not be used by itself as evidence of normativity. But *radical* disputes seem precluded. And yet our intuitions seem to allow for the possibility of radical disputes in the case of normative terms. Hare took this to support his expressivism: our intuitions differ with these terms because they are not genuine referring expressions at all.

So imagine that Jerry and Tyler travel to the island of the cannibal metasemanticists. Technology is fortunately very advanced there. There is a giant machine showing all the non-normative and non-meaning facts of the world, all the facts which potentially form part of meaning's supervenience base. They use it to look at Jenny. Jenny uses the term 'green', among others.

Jenny is disposed to apply 'green' to all and only green objects, but people in her community are disposed to apply the term 'green' to green objects before *t* and blue objects after.¹² No one is disposed to apply the term to red objects.

Jerry says, "'Green' uttered by Jenny means GREEN."

Tyler says, "'Green' uttered by Jenny means GRUE."¹³

But the cannibal says, "'Green' uttered by Jenny means RED."

They assign meanings to the rest of Jenny's terms. The general pattern continues. Jerry and Tyler actually agree a lot (Jenny has internalized the dispositions of her community, naturally). The cannibal never agrees. Jerry and Tyler ask him if he realizes that on his theory of meaning, neither individuals nor communities ever use their own concepts correctly.

"Yes," says the cannibal, "we agree about the subvening facts, we just disagree about what meanings supervene upon them."

Jerry and Tyler are plausibly having a disagreement. The cannibal is talking past everyone else.¹⁴ By 'meaning' he means something other than MEANING.

3. *Explanation versus Prescription*

Even if two speakers apply the term 'good' in radically different ways, we are still able to hear disagreements between them as genuine. But in the case of 'meaning', there is a much more significant limit to how radical disagreement can be before we hear it as merely verbal. Jerry and Tyler, for instance, obviously disagree. But notice that their hypotheses make the same predictions in a wide range of cases. An individual's use of a

¹¹ This may be the Principle of Charity at work as well. Possibly 'good' is unique because we find it more charitable to attribute false evaluative beliefs than radically counter-evaluative motives. But this still raises the question of how strikingly inaccurate evaluative beliefs are interpretable as such in the first place.

¹² This dispute is based on the example provided by Gibbard (42-46).

¹³ *Grue* is the notorious property of being observed before time *t* and being green, or else not being observed before time *t* and being blue, introduced by Nelson Goodman (1955/1983).

¹⁴ That is to say, he is offering a merely verbal disagreement.

term will almost always converge with her community's use of a term. Thus, Jerry and Tyler's predictions about meaning in ordinary, familiar, earth-bound cases will be in *almost* complete agreement. Jenny's case—the case where they disagree—is unusual.

The problem with the cannibal's theory is that it offers almost no such overlap in predictions. A non-negotiable desideratum on a theory of meaning is that it explain linguistic behavior: why people speak and judge as they do.¹⁵ It is true that Tyler's communitarian theory may fail to explain certain instances of usage—it does not explain, in the case imagined, why Jenny applies 'green' to green objects after *t*. But no theory of meaning will explain every single instance of usage, not without auxiliary hypotheses at least. Speakers will misapply a term thanks to ignorance or irrationality; some applications may be literally incorrect but acceptable thanks to pragmatics; none of these are elements of a theory of meaning proper. Presumably it will be part of Tyler's larger theory of human linguistic behavior that people sometimes fail to correctly internalize the meaning of a term in their language. But for Tyler that story about internalization and its misfires will be an auxiliary hypothesis.

Nevertheless, a theory of meaning must explain some suitably large portion of human beings' actual linguistic behavior. That behavior includes how individuals actually speak. It will also include, Tyler will want to remind us, how we defer to the correction of other speakers in our community and try to bring our use into conformity with theirs.

Notice that Tyler is in a position to explain the last—our tendency to deference and conformity—straightforwardly: people do these things because of what the words they are using mean. This also provides an explanation of most of a given individual's uses of terms: since she has these dispositions to defer and conform in her usage, a communitarian theory of meaning will be generally predictive of her use. There will be failures, however, for which Tyler will owe us an alternate explanation.

Jerry's individualist theory, on the other hand, will be more predictive of individual usage in the rare cases when it departs from communal use. At the same time, he must invoke auxiliary hypotheses (such as our desire to use language to communicate) to explain deference and conformity. Furthermore, he must invoke auxiliary hypotheses to explain why Jenny, if she starts to use 'green' as her community does under the pressures of criticism, is more likely to regard herself as learning to use the term correctly

¹⁵ Boghossian notes that this is a potential stumbling block for non-naturalist accounts of meaning and something that a normative account of meaning must finesse (1989, 549). More extensive discussion of this point is also found in (Horwich 1998, 47ff.). Full reconstruction of Horwich's argument is beyond the scope of this paper, but the main points are: (1) it is pretheoretically intuitive that people use the words they do because of what those words mean; (2) if meaning were not explanatory of something in principle observable (such as behavior), it is unclear how we could have any empirical evidence for what words as used by others mean; and (3) that translation dictionaries are useful is completely mysterious if meaning is not explanatory. Note that one can accept these points without a commitment either way on Horwich's *meaning is use* thesis.

than as changing her idiolect to one in which communication is easier. Finally, because individuals tend over time to use words as the members of their linguistic community do, Jerry's assignments of meaning, while they are initially more predictive of individual behavior, *may* be less temporally and counterfactually robust than Tyler's assignments.

So there are some things for which Jerry has the simpler explanation, and some things for which Tyler does. This is what we would expect of competing research programs. We would also expect competing research programs to disagree to some extent on which of the phenomena are central and which peripheral. There will even be, at the margins, disagreement about which phenomena should be simply dismissed as misleading. Nonetheless, there is a class of phenomena which intuitively make up the data for a theory of meaning to explain, and which both sides agree their theories must *by and large* explain, with some small remainder to be explained with auxiliary hypotheses, and an even smaller remainder to be at the very least explained away.

This requirement on explanatoriness is the source of the divergence from a term such as 'good'. The cannibal can prescribe that people act in ways radically different from the way they do, and so can the missionary. But that is intelligible, in part, because there is no requirement that a theory of goodness will explain actual behavior. Explanatoriness is a requirement on a theory of meaning, however, and that puts serious limits on intelligible disagreement.

It should be made clear that the explanatoriness constraint is consistent with the normativity of MEANING. It is simply introduced here to show that the case against radical disagreement is not merely intuitive. The intuition is rationally defensible; sacrificing it would require that we give up on much more.¹⁶

Gibbard, or a defender, could note a difference in the thought experiments. The original cannibals gave other signs they were using 'good' to encourage behavior: the rewards, smiles, and their own choices. Do we hear disagreement if we imagine that the cannibal metasemanticist tries to encourage Jenny to apply 'green' to red objects?

If the cannibal begins to prescribe, there will probably be disagreement about something, but the cannibal's behavior only helps us to discern a disagreement about MEANING if we understand the cannibal as engaged in a special, meaning-related form of prescription. Jerry and Tyler could come to judge, for all sorts of reasons, that Jenny ought to apply the term 'green' to red objects—if an evil demon threatened mass torture unless she did so, for example (Gibbard 2012, 14-15). But this prescription is not

¹⁶ What if the normative is itself explanatory? I will return to this point later (in section 5). For now it is enough to note that this position is at odds with Gibbard's argumentative strategy. If normative properties can be explanatory we should not expect to find radical disagreement surrounding normative terms, true. But then Gibbard's strategy of using to unusual intuitions about metasemantic disagreement to support the normativity of meaning is misguided, because unusual intuitions about disagreement should not be a characteristic feature of normative terms in the first place; or if there are such intuitions, they must be misleading (see also footnotes 9 and 25).

relevantly meaning-related. It is not based on the fact that this would be a correct application of the term; nor is it a prescription that *determines* the correct application.^{17, 18}

The cannibal's use of exhortations must be based on a sense of 'ought' which is either based on correct application of the term, or else determines correct application of the term. Otherwise his prescriptions have nothing to do with meaning, and so are irrelevant to the debate at hand.

But this shows us that invoking the cannibal's prescriptions to support the claim that there is genuine disagreement about meaning is doubly question-begging. First, a philosopher who denies that meaning is normative must also deny that there is a specially meaning-related sense of 'ought'. So we cannot insist that the cannibal is using 'ought' in this sense to establish that meaning is normative.

Second, if we cannot make sense of the claim that 'green' as uttered by Jenny means RED, we cannot make sense of the claim that the cannibal's prescription "Jenny ought to apply 'green' to red objects" uses the meaning-related sense of 'ought'. The meaning-related sense of 'ought' must either be explained by, or explain, what the meaning of some term is. But if we cannot make sense of some claim, being told that this claim is either the *explanans* or *explanandum* of some further ought-claim does not by itself make the initial claim clearer.

Until something is done to make the claim that 'green' as uttered by Jenny means RED intelligible, we must interpret any prescriptions of the cannibal as using a sense of 'ought' that is not meaning-related. The cannibal is trying to get Jenny to speak a different language, perhaps; perhaps he wants her to fool or appease the evil demon.

Maybe we could add additional behavioral evidence. The cannibal could also use 'green' to refer to red objects when speaking with Jenny, and he could act in ways that show that he interprets her assertions involving 'green' as asserting redness of something. This might help give us some sense that he is in genuine dispute with Jerry and Tyler.

The problem here is that we have already stipulated that the cannibal knows all of Jenny's dispositions, and so he knows that if he wants to helpfully to speak her language, telling her that some red stones are 'green' will not be helpful at all, and actually lead to substantial confusion. Insofar as he wants to help, he will assert that stones are 'green' only when he believes they are green, because of how *she* is disposed to respond to his utterances. It is similarly unclear what is involved in his taking her assertions about 'green' as predicating redness of something. Insofar as he is responsive to evidence, he will treat her utterances using 'green' as evidence for green objects, because of when she is disposed to make such utterances. In other words, in the absence of serious

¹⁷ Note that for Gibbard the meaning-related use of 'ought' is the Right Kind of Reasons use of 'ought' (2012, 12-16). This complication doesn't seem to affect the argument here.

¹⁸ The normativity of meaning thesis can either be the thesis that meaning is normativity-*engendering*—that is, that meaning explains certain normative facts—or normativity-*determined*—that is, that normative facts explain the meaning of a term (Glüer and Wikforss, 2009; and 2010). Gibbard does not specify which version of the thesis he holds.

irrationality, the cannibal's behavior regarding Jenny's term 'green' will be indistinguishable from the behavior of someone with a more plausible theory of meaning who also shares his aims and all of his non-semantic beliefs. His personal behavior cannot reveal any difference in semantic-commitments, then, insofar as he is rational.¹⁹

'Means' displays Moorean openness, but not openness to radical disagreement. In other words, it displays the kind of openness characteristic of normative and non-normative terms alike, but not the openness that seems peculiar to normative terms.

4. *Moral Twin Earth*

Moral Twin Earth also generates unusual intuitions about disagreements, and can thus be used as a model for another test of the normativity of terms.²⁰ In Putnam's (1975) original Twin Earth example we are asked to imagine a planet that is qualitatively identical to the actual earth, except that all H₂O is some alien chemical compound, XYZ. This substance tastes like water, falls from clouds in the skies, fills the Twin Atlantic, etc. Now, we are to imagine that before the development of modern chemistry on both planets, an earthling uses the term 'water', and her twin uses the same. The standard intuition is that these terms have different meanings: 'water' as used by the earthling means H₂O and 'water' as used by the Twin Earthling means XYZ. Should an astronaut travel from earth to its twin and say of the contents of some river there, "That is not water," there is no genuine disagreement with the locals who continue to call the stuff "water"—or so intuitions typically go.

Terence Horgan and Mark Timmons (1991) point out that our intuitions are different when confronted with thought experiments involving normative terms.²¹ Imagine we discover Moral Twin Earth. Moral Twin Earth is much like earth, with a Moral Twin EU and a Moral Twin Demilitarized Zone between Moral Twin North and South Korea. The Moral Twin "Anglophones" of the planet also use words that sound

¹⁹ One might worry that Jerry and Tyler's behavior regarding Jenny's use of 'green' will be indistinguishable as well, insofar as they are rational, given that they know all of her dispositions. This is correct. But it only shows that we cannot understand them as entertaining rival theories of meaning *on the assumption* that a theory of meaning should be action-guiding in the way that ought-judgments are normally thought to be. If it seems that they do still disagree, one should take this as evidence against the thesis that meaning-claims express any sort of action-guiding commitments. In this context, it is worth considering the criticism that meaning is only instrumentally normative—that one ought to apply words in accordance with their meaning *if one wants to communicate successfully, accept true sentences, etc.*; see (Horwich, 1998; Hattigandi, 2006; and Glüer and Wikforss, 2010); but also see (Whiting, 2007) for reply.

²⁰ A referee wonders if the Twin Earth cases illustrate any difficulties with the normativity of meaning thesis in addition to those illustrated by the Missionary and Cannibals cases. I suspect that they do not, which is why considerably less time is dedicated to drawing out the lessons of these thought experiments. Twin Earth cases are addressed here for the sake of thoroughness, to see if these thought experiments generate different intuitions from the cannibal cases.

²¹ For an excellent discussion of the Moral Twin Earth debate, see (McPherson, 2013).

remarkably like ‘right’ and ‘wrong’ or ‘good’ and ‘bad’. These terms seem to be connected with behavior in much the same way as their homonyms in English. As Horgan and Timmons put it (459):

Moral Twin Earthlings ... normally take considerations about what is 'good' and 'right' to be especially important, even of overriding importance in most cases, in deciding what to do....

But social scientists discover that Moral Twin Earthling use of ‘right’ is determined by a deontological property. On Earth, their counterparts discover that a consequentialist property determines people’s use of ‘right’.²² Both properties play a role in explaining the functioning of society—they seem like genuine social-scientific kinds. Finally, ‘right’ on both worlds is connected with choice, recommendation, reward, and the regulation of behavior; ‘not right’ with abstention, anger, proscription, and punishment. Earthlings and Moral Twin Earthlings meet. There is a dilemma: one must be killed to save five. The earthling says, “Killing the one is right.” The Moral Twin Earthling says, “Killing the one is not right.” This disagreement seems genuine. We hear normative disagreements as genuine in cases where non-normative disagreements would sound merely verbal.²³

‘Meaning’ does not generate similar intuitions about disagreement. Gibbard is completely correct that we do not hear the debate between Jerry and Tyler as merely verbal. But we could imagine *discovering* that it is a merely verbal dispute, or that it is based on the false presupposition that ‘meaning’ in these contexts is unambiguous—rather than being used to express both PUBLIC MEANING and PRIVATE MEANING. Both the properties public and private meaning could do unique and important work explaining linguistic usage and cognition.²⁴

There are reasons to assume at present that this is not the case. As we saw above, meaning individualists can offer plausible accounts of how the social aspects of language are explained by the linguistic and cognitive dispositions of individuals; likewise,

²² “Determines” is, like “dispositions,” a placeholder for a more systematic account (see footnote 3). In the original paper, the earthlings and twin earthlings have their use of ‘right’ *causally regulated* by the consequentialist and deontological property, respectively. But again, one could appeal to asymmetric-dependence, or possibly a number of other accounts, to specify the determination relation.

²³ The paper will remain agnostic on whether Moral Twin Earth is good evidence of non-naturalism or expressivism about the normative, though its inventors argue that it is. It certainly seems like *better* evidence than Moorean openness. The latter can be explained away by semantic externalism or the assumption that our knowledge of meanings is implicit. Neither thesis explains, however, why a referring term would fail to work like a natural kind term even when it is stipulated that speakers standardly apply it to a natural kind property.

²⁴ Something like this is suggested in (Horwich 1998, 86-87), where it is argued that Chomskian theories of I-languages are not really in competition with communitarian theories of meaning, but have different theoretical aims.

meaning communitarians can argue that individual usage is parasitic on collective practice. That is, each theory has grounds to argue that either public meaning or private meaning is a derivative property, dependent on what the theory identifies as meaning proper. But it could turn out that a theory positing distinct, explanatorily independent properties of public meaning and private meaning actually provides a more elegant and powerful account of the data.

It is much more difficult to imagine discovering that ‘right’ *as used in moral contexts* is similarly ambiguous. (It is of course ambiguous if we include non-moral uses such as “what’s the right answer?”) Even if both the consequentialist property and the deontological property play roles in explaining how society functions, and even if ordinary speakers apply ‘right’ to both properties with nearly equal frequency, we are not inclined to think ‘right’ ambiguous between CONSEQUENTIALIST RIGHT and DEONTOLOGICAL RIGHT. It’s more natural to think that people make false claims about what it’s right to do in about fifty-percent of cases.

The evidence for this is the Moral Twin Earth case. Even when we imagine linguistic communities on *two different planets*, each using ‘right’ to express one of the proposed precisifications of the term, we still hear disagreement.²⁵

Do our intuitions differ because in the case of ‘meaning’ we have not filled in the story in enough detail? What if we imagine that Semantic Twin Earthlings who treat their use of ‘meaning’ as action-guiding or inference-guiding in some respect? The problem with this suggestion is that, even if ‘meaning’ is a normative term, meaning-judgments are only plausibly action- and inference-guiding to a very weak extent, overruled by other kinds of considerations (cf. Whiting, 2007). Here on earth, judgments about ‘meaning’ are not taken to be “especially important, even of overriding importance in most cases, in deciding what to do”—in contrast to judgments about ‘good’ and ‘right’. Jerry may be a meaning individualist, but if his attitudes are anything like those of a normal person’s, he will advise Jenny to apply ‘green’ gruesomely when conversing with other members of her community (because for her to do otherwise is imprudent). Tyler may be a meaning communitarian, but once he knows Jenny’s unusual dispositions he will use and interpret her use of ‘green’ incorrectly (by his lights) when communicating with her.

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

²⁵ If we do not hear disagreement, this is no help to Gibbard. My argument has been that when we look at some of standard cases, disagreements involving semantic terms do not elicit the intuitions we associate with normative terms. My argument would be defeated, of course, if it rested on a false presupposition: but Gibbard’s entire argumentative strategy rests on the same presupposition. If our intuitions about normative terms are not unusual, disagreement cases could not serve as the type of evidence he needs (see also footnotes 9 and 16). In any case, the extensive literature on Moral Twin Earth suggests that the intuitions of disagreement are fairly widespread; again, see (McPherson 2013).

departures in behavior on Moral Twin Earth. Even if MEANING is normative, we cannot expect similar departures on a Semantic Twin Earth.

Again, our intuitions about disagreement fail to reveal any way in which ‘meaning’ is not like a standard referring expression. What is more, even if meaning is normative, most speakers seem to treat its normative force as overruled with enough ease that it is unclear how we could develop thought experiments that might elicit more non-standard intuitions.

5. Conclusion: Explanation Versus Prescription Again

Later thought experiments have shown that there is a genuinely unusual feature of moral terms (or at least our intuitions about those terms). ‘Meaning’ does not seem to share this unusual feature.

Jerry and Tyler offer theories of meaning with substantial overlap about standard cases. The theories come apart in highly arcane thought experiments. We have agreement in the overwhelming majority of familiar cases and disagreement on the margins. Rather than hearing radically diverging uses as expressing radical disagreement we hear them as talking past the target audience. These are the typical symptoms of informative debate involving a normal referring expression.

Of course, showing that Gibbard’s argument fails does not show that his thesis is incorrect. Nonetheless, accepting the normativity of MEANING fits poorly with an expressivist or non-naturalist understanding of what normativity is. If someone were to tell us that no one ever did what they had most reason to do, that no one ever did what they ought, or what was right, or pursued the good, there would be a lot to worry about with such a theory, but we would understand them. The normative seems like it need not, at least as a conceptual matter, explain people’s actual behavior. The missionary’s theory of the good need not explain what the cannibals actually do.

This point seems to favor the non-naturalist or expressivist, and they offer theories especially suited to account for it. For the traditional non-naturalist, normative properties are not and do not reduce to natural properties—properties that can explain events in the natural world. For the traditional expressivist, our concepts OUGHT and GOOD are concepts that do not present their objects as *explanans* of natural phenomena, but rather play special non-theoretical roles in directing choice and feeling.

This is reasonable, but, at least in its most straightforward versions, it cannot be combined with the view that MEANING itself is a normative concept, for the simple reason that a theory of meaning must explain people’s actual behavior and psychology (to some degree), otherwise we don’t understand how it is a theory of meaning. The cannibal’s theory of meaning had better explain what Jenny actually does with words and thoughts. But then, if one is to maintain that MEANING is normative, one must first try to explain away the intuition that a normative concept does not present its object as doing

explanatory work (because clearly MEANING presents its object as doing explanatory work). *That strategy*, however, is simply incompatible with the claim that normative properties are not explanatory or normative terms not even standard referring expressions.

To illustrate this further, consider an extant account of the normativity of meaning according to which the proposed normative property is explanatory. This is a theory on which terms are assigned those meanings which would make most sense of the speaker's use of the term and dispositions to use the term, in combination with her wider attitudes, behaviors, and dispositions—that is, the meaning of the term is that possible meaning on which the speaker's use and wider behavior is most rational (Davidson, 1970; and 1985; Lewis, 1974; Wedgwood, 2007).²⁶ There are disputes about whether rationality is genuinely normative.²⁷ I won't address them here, other than to say that many philosophers have thought that it is, and to note that calling some conclusion or decision *irrational* seems like criticism. Rationality is also plausibly explanatory. By and large, we reach the conclusions we do because those conclusions are rational, given our evidence and prior beliefs; we make the decisions we do by and large because those decisions advance our aims.²⁸ This isn't to claim that humans are always or perfectly rational—but *ceteris paribus* explanations are, we've already noted, acceptable.

It is beyond the purposes of this paper to defend or argue for such a view: I will simply show how such a view could make sense of Jerry and Tyler's debate as an implicitly normative one. It will turn out that this is the sort of normative debate that Gibbard's expressivism handles poorly.

No human agent is perfectly rational, and this is good, because no theory of meaning should try to vindicate all of an agent's applications of a term: some must be mistaken. On this metasemantic view, however, the meaning of a term is that which *minimizes* an agent's irrationality. It is not simply a matter of assigning to them the smallest number of actual mistakes, either. There are their dispositions to make mistakes under various scenarios, and the relative severity of mistakes must be taken into account. A mistake made under highly unfavorable epistemic circumstances plausibly could count against the rationality of an agent less than those made in better conditions, for example. There are also issues of the types of terms agents use: it may be that to deploy terms and concepts picking out natural kind properties or other meaningful, potentially explanatory

²⁶ As both Lewis and Wedgwood emphasize, this view can be decoupled from Davidson's interpretivism about the mental; also see (Zangwill 1998, 185-186).

²⁷ For examples of criticism or skepticism, see (T. Schroeder, 2003; and Kolodny, 2005). A summary of the debate is found in (Way, 2010). It should be noted, however, that Schroeder is not primarily concerned to deny the normativity of rationality, but rather to argue that whatever normativity rationality may possess is irrelevant to the role it plays in a Davidsonian theory of mind and meaning.

²⁸ Different versions of this thesis are endorsed in, for example, (Dennett, 1971; Fodor, 1994b; and Wedgwood, 2007). But also see (Glüer and Wikforss, 2009) for argument that the explanatory role of rationality undermines its normative role.

properties that can figure in interesting generalizations is more rational than deploying terms and concepts that pick out Cambridge properties, for example.

This allows a normative interpretation of the debate between the communitarian and the individualist: it is a debate about the relevant severity of different kinds of rational failings. Jerry holds that assigning a meaning to ‘green’ that makes Jenny disposed to systematic and pervasive misapplication of the term in what should be epistemically favorable circumstances is to make her unacceptably irrational. Tyler, on the other hand, takes her higher-order dispositions to defer and to accept correction as bearing more rational weight: such deference would be highly irrational if ‘green’ in the larger community were simply a different word from Jenny’s word ‘green’.²⁹

The debate won’t stop here, but its continuation is beyond our purposes. It has already illustrated how normativity could determine meaning: disagreements about the relative irrationality of various failings lead to a disagreement about the correct assignments of meaning. What is more, this kind of debate does not seem amenable to radical disagreement. Jerry and Tyler might disagree about the *precise significance* of various failings, but there will be general agreement about what would count as failings: misapplications, and dispositions to misapply under the pressure of other people’s idiolects, in this case.

This, however, is exactly the kind of normative disagreement that Gibbard’s expressivism is poorly designed to handle. For Gibbard, normative judgments express desire-like or intention-like endorsements. But whereas a theory of rationality must be constrained to some extent by an agent’s actual dispositions to choose and infer—both because it is in part a theory of what would make sense, *given* her existing attitudes, and because it is explanatory—and whereas a theory of meaning must be constrained to some extent by actual use for very similar reasons, I can endorse truly extensive departures from the actual. I can fervently desire that everyone start speaking in the made up words of Lewis Carroll and Dr. Seuss poems. But this means that for Gibbard’s expressivism to be plausibly married to a normativity of meaning thesis, we would first need a non-*ad hoc* justification for restricting the range of possible endorsements to those that don’t depart too far from the actual.³⁰

²⁹ See (Schroeter and Schroeter, 2009) for a detailed version of a view similar to Tyler’s, albeit without any endorsement of the normativity of rationality or the consequent normativity of meaning.

³⁰ A referee notes that Gibbard’s original (1990) expressivist theory was primarily an expressivist treatment of ‘rational’. Does this argument suggest that Gibbard’s entire earlier project is a failure, since *rationality* must also meet the explanatoriness constraint? This would be premature. Gibbard is explicit that his use of “rational” is supposed to be synonymous with “the thing to do” or “makes sense.” In other words, (Gibbard, 1990) uses the term to pick out an unqualified prescription, whereas more recent work in metaethics uses ‘rational’ to pick the option that would make most sense given such qualifications as one’s beliefs, aims, and epistemic limitations; see (Way, 2010) for summary and discussion. (Gibbard, 1990) clearly intends to offer an expressivist theory of *unqualified prescription*. So the argument here is in merely verbal disagreement with this aspect of Gibbard’s earlier work.

In general, the same objection will apply to non-naturalism. Non-naturalists generally hold that normative properties are not part of the causal-explanatory order, and while they must supervene on natural facts that are part of that order, the *particular* supervenience relation is not analytic. But in that case we again stand in need of a non-*ad hoc* justification for restricting the intelligible range of disagreements about rationality and consequently meaning—since the claim that the particular relation is non-analytic entails that one could be radically mistaken about which particular facts the normative supervenes on *without* semantic incompetence. There may be exceptions to all of this, however. Ralph Wedgwood (2007), for example, argues for normative non-naturalism, while also arguing that the normative can be causally explanatory.³¹ My objections do not apply to that brand of non-naturalism.

To conclude, semantic terms may display Moorean Openness, but that is just to say they are like most terms of English—it isn't evidence for anything else. There is another form of openness which seems unique to moral and some other normative terms, but our intuitions about semantic disagreement do not provide evidence of that. There is, moreover, good reasons to think that they shouldn't, as semantic properties need to explain a more or less agreed upon set of linguistic phenomena. This seriously constrains the kinds of normative properties in terms of which semantic properties might be understood: namely to those which themselves meet explanatoriness constraints. But given Gibbard's own commitments regarding the normative, it seems that he lacks any principled way of accepting this restriction. More generally, the thesis that meaning is normative fits poorly with an expressivist or non-naturalist metaethic, at least in their more orthodox forms.*

REFERENCES

- ASAY, J. 2013, *The Primitivist Theory of Truth*, New York: Cambridge University Press
- BAKER, D. AND WOODS, J. 2015, "How Expressivists Can and Should Explain Inconsistency", *Ethics* 125, pp. 391-424
- BOGHOSSIAN, P. 1989, "The Rule-Following Considerations", *Mind* 98, pp. 507-549
- BOYD, R. 1997, "How to be a Moral Realist", in: S. DARWALL, A. GIBBARD, and P. RAILTON, eds., *Moral Discourse and Practice*, New York: Oxford University Press, pp. 105-136
- BRANDOM, R. 1994, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, MA: Harvard University Press
- CHALMERS, D. 2011, "Verbal Disputes", *The Philosophical Review* 120, pp. 515-566

³¹ Also see (Zangwill, 1998).

* Research appearing in this article was partially funded by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (LU342612). Thanks to Jamin Asay, David Faraci, Tristram McPherson, Jack Woods, the audience at the American Philosophical Association's 2014 Pacific Division Conference in San Diego, California, and the referees and editors at *dialectica* for critical feedback and helpful suggestions.

- DAVIDSON, D. 1970, "Mental Events", in: L. FOSTER and J.W. SWANSON, eds., *Experience and Theory*, University of Massachusetts Press, pp. 79-101; reprinted in D. DAVIDSON, *Essays on Actions and Events*, 1980/2001, New York: Oxford University Press, pp. 207-228
- DAVIDSON, D. 1985, "Incoherence and Irrationality", *dialectica* 39, pp. 345-354
- DENNETT, D.C. 1971, "Intentional Systems", *The Journal of Philosophy* 68, pp. 87-106
- DOWELL, J.L., forthcoming, "The Metaethical Insignificance of Moral Twin Earth", in SHAFER-LANDAU, ed., *Oxford Studies in Metaethics*, New York: Oxford University Press
- FODOR, J. 1994a, "Fodor's Guide to Mental Representation", in (Stich and Warfield 1994, 9-33)
- FODOR, J. 1994b, "A Theory of Content II: The Theory", in (Stich and Warfield 1994: 180-222)
- FOOT, P. 1959, "Moral Beliefs", *Proceedings of the Aristotelian Society* 59, pp. 83-104
- GIBBARD, A. 1990, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, MA: Harvard University Press
- GIBBARD, A. 2003, *Thinking How to Live*, Cambridge, MA: Harvard University Press
- GIBBARD, A. 2012, *Meaning and Normativity*, New York: Oxford University Press
- GOODMAN, N. 1955/1983, *Fact, Fiction, Forecast, Fourth Edition*, Cambridge, MA: Harvard University Press
- GLÜER, K. and WIKFORSS, Å. 2009, "Against Content Normativity", *Mind* 118, pp. 31-70
- GLÜER, K. and WIKFORSS, Å. 2010, "The Normativity of Meaning and Content", E. ZALTA, ed., *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/win2010/entries/meaning-normativity/>
- HARE, R.M. 1952, *The Language of Morals*, New York: Oxford Clarendon Press
- HATTIANGADI, A. 2006, "Is Meaning Normative?", *Mind and Language* 21, pp. 220-240
- HORGAN, T. and TIMMONS, M. 1991, "New Wave Moral Realism Meets Moral Twin Earth", *The Journal of Philosophical Research* 16, pp. 447-465
- HORWICH, P. 1998, *Meaning*, New York: Oxford University Press
- JACKSON, F. 1998, *From Metaphysics to Ethics: a Defense of Conceptual Analysis*, New York: Oxford University Press
- JACKSON, F. 2004, "Why We Need A-Intensions", *Philosophical Studies* 118, pp. 257-277
- KOLODNY, N. 2005, "Why Be Rational?", *Mind* 114, pp. 509-563
- KRIPKE, S.A. 1982, *Wittgenstein on Rules and Private Language*, Cambridge, MA: Harvard University Press
- LEWIS, D. 1974, "Radical Interpretation", *Synthese* 27, pp. 331-344
- MCPHERSON, T. 2013, "Semantic Challenges to Normative Realism", *Philosophy Compass* 8, pp. 126-136
- MACFARLANE, J. 2007, "Relativism and Disagreement", *Philosophical Studies* 132, pp. 17-31

- MILLER, A. 2003, *An Introduction to Contemporary Metaethics*, Malden, MA: Polity Press
- MOORE, G.E. 1903/1993, *Principia Ethica*, Revised Edition, T. BALDWIN, ed., New York: Cambridge University Press
- PLUNKETT, D. and SUNDELL, T. 2013, "Disagreement and the Semantics of Normative and Evaluative Terms", *Philosophers' Imprint* 23, pp. 1-37
- PUTNAM, H. 1975, "The Meaning of 'Meaning'", in: K. GUNDERSON, ed., *Language, Mind, and Knowledge: Minnesota Studies in Philosophy of Science, vol. 7*, University of Minnesota Press, pp. 131-193
- RICHARD, M. 2008, *When Truth Gives Out*, New York: Oxford University Press
- RIDGE, M. 2014, *Impassioned Beliefs*, New York: Oxford University Press
- RIDGE, M. 2015, "Internalism: *Cui Bono?*", in: G. BJÖRNSSON, C. STRANDBERG, R. FRANCÉN OLINDER, J. ERIKSSON, and F. BJÖRKLUND, eds., *Motivational Internalism*, Oxford: Oxford University Press, pp. 135-149
- SCHAFER, K. 2014, "Constructivism and Three Forms of Perspective-Dependence in Metaethics", *Philosophy and Phenomenological Research* 89, pp. 68-101
- SCHROEDER, M. 2008, "How Expressivists Can and Should Solve Their Problem with Negation", *Nous* 42, pp. 573-599
- SCHROEDER, T. 2003, "Davidson's Theory of Mind Is Non-normative", *Philosophers' Imprint* 3, pp. 1-14
- SCHROETER, F. and SCHROETER, L. 2009, "A Third Way in Metaethics", *Nous* 43, pp. 1-30
- STICH, S. and WARFIELD, T.A., eds., 1994, *Mental Representation: A Reader*, Cambridge, MA: Blackwell Publishers
- WAY, J. 2010, "The Normativity of Rationality", *Philosophy Compass* 5, pp. 1057-1068
- WEDGWOOD, R. 2007, *The Nature of Normativity*, New York: Oxford Clarendon Press
- WHITING, D. 2007, "The Normativity of Meaning Defended", *Analysis* 67, pp. 133-40
- ZANGWILL, N. 1998, "Direction of Fit and Normative Functionalism", *Philosophical Studies* 91, pp. 173-203