Disagreement Lost and Found

Stephen Finlay

ABSTRACT: According to contextualist and other content-relativist views in metaethics, different speakers use the same moral and normative sentences to say different things. These views face a classic problem of Lost Disagreement, which they attempt to solve by identifying pragmatic, non-content-based kinds of disagreement. This paper critically compares two broad strategies of this kind, (1) *quasi-expressivist* views that analyze disagreement over whether S ought to do A in terms of conflicting attitudes toward S doing A, and (2) *metalinguistic* views that analyze such disagreement in terms of conflicting attitudes toward how to talk about S's doing A. While the main objection to quasi-expressivist views (concerning the felicity of semantic negation markers like 'wrong', 'incorrect', and 'false') fails, objections to metalinguistic views are argued to be decisive. Content-relativists should be quasi-expressivists about fundamental normative disagreement.

Consider the following, familiar metaethical view:

Content-Relativism: Different speakers use the same moral sentences, e.g. of the form 'S ought to do A', to say different things.[1]

For example, we might suppose that if Immanuel were to assert, 'One ought to tell the truth even to axe-wielding murderers,' he would be saying that telling the truth even to axe-wielding murderers is necessary for acting on a universalizable maxim, while if Jeremy were to assert the same sentence he would be saying that telling the truth even to axe-wielding murderers is necessary for maximizing happiness. Content-relativist views have appealed to many philosophers, at least since Westermarck (1906-08, 1932), for a variety of reasons including: their promise to identify what in the world moral claims could be about (solve the "location problem"), to explain the diversity in moral opinion, and (more recently) to fit into a unifying semantic treatment of modal terms like 'ought' now orthodox in linguistics. For almost as

---

[1] Content-relativism is a broad category, which includes (many) *indexical* views that liken moral words to indexicals such as 'I' and 'now', (many) more broadly *contextualist* views that assign moral words implicit argument-places taking different values on different occasions of use, and views on which different speakers simply use moral words with different meanings. However, it excludes the kind of view most commonly labelled 'relativism' today, which locates the relativity in the truth value rather than the content of moral sentences and utterances. Note that there are also content-absolutist forms of contextualism, as advocated in Dowell ms.

long—at least since Moore 1922 (pp. 333-4)—they have been the target of one central objection that has widely been considered fatal.[2] If two speakers use the same sentence to say different things, then when one asserts the sentence and the other asserts its negation, the contents of their utterances will not (ipso facto) be inconsistent or incompatible. However, intuitively when one speaker asserts 'S ought to do A' and another asserts 'S ought not to do A'—and both intend to make moral, rather than prudential or legal (etc.) claims—they do thereby have a moral disagreement.[3] Hence, we have the problem of

Lost Disagreement: Content-relativist views seem unable to accommodate the existence of some intuitive cases of normative disagreement.

This problem is a central motivator of views that embrace the contrary thesis,

Content-Absolutism: Different speakers always use the same moral sentences to say the same things.

Absolutism comes in both cognitivist and noncognitivist varieties, with cognitivist absolutism holding that a moral sentence always has the same descriptive content (in virtue of its normative terms, at least, setting aside other contextually variable words like 'you' and 'now'), and noncognitivist absolutism holding that a moral sentence always has the same attitudinal content. Parallel problems familiarly arise for many other words of philosophical interest, including aesthetic predicates ('fun', 'tasty') and epistemic modals ('might'), and there are rapidly expanding and roughly parallel literatures addressed to these analogous problems. While this paper focuses on the metaethical case, I expect its points will also apply, in general and mutatis mutandis, to content-relativism in these other debates.

Content-relativists have typically responded to the Problem of Lost Disagreement (since at least Harman 1996: Ch. 3)[4] by challenging the assumption that disagreement requires utterances

---

[2] See the list of references in Khoo & Knobe 2016: 4n, for example.
[3] Strictly the negated form is 'It is not the case that S ought to do A', but for convenience I'll assume that this is entailed by 'S ought not to do A', as on standard deontic logics.
[4] Harman is not himself strictly a content-relativist, since he advances relativism as a charitable reinterpretation of ordinary (absolutist) moral claims for the purpose of assigning truth rather than as an account of what those claims ordinarily mean (1996: 17).

with inconsistent or incompatible contents, and locating the disagreement somewhere else. (This includes my own work; see Björnsson & Finlay 2010, Finlay 2014: Ch. 8). But I think it's fair to say that philosophers not already convinced of content-relativism have generally seen this *non-content-based* strategy as an implausible and desperate move to save a theory. As some have complained (e.g. Thomson 1996: 196f, Olson 2011), the various moves made in the service of this strategy can seem gratuitously complex compared to the simpler accounts of disagreement offered by content-absolutist views.

This debate has seen two major developments over roughly the last five years. First, proponents of this non-content-based disagreement strategy have gone on the offensive, offering various kinds of evidence that as a matter of fact, much ordinary moral or normative disagreement does not involve utterances with inconsistent or incompatible contents. So it is suggested that it is actually the absolutist, not the relativist, who has some explaining to do. As Khoo and Knobe (2016) put it, "not only is it *not* problematic for a theory if it fails to predict exclusionary content in all cases of moral disagreement, but it *is* problematic for a theory if it *does* predict exclusionary content in all cases of moral disagreement."[5] I know this claim will raise some eyebrows,[6] so while it isn't my purpose in this paper to argue for the strategy, I'll quickly list some of the kinds of evidence I have in mind:

(i) Widespread intuitions that in some (especially cross-cultural) cases of apparent moral disagreement, it might be that neither party, or judgments, are "wrong" or "incorrect" (as experimentally found by Sarkissian et al. 2011), even though these were judged to be real disagreements, in which responses of "no" are appropriate (Khoo & Knobe 2016).[7]

(ii) Intuitive asymmetries in normative disagreement—i.e. A intuitively disagrees with B, but B doesn't intuitively disagree with A—whereas inconsistency is a symmetrical relation

---

[5] Also Finlay 2014: 245, Bolinger ms.
[6] An informal poll of the audience at the 2016 ChillMeta Workshop suggested that content-relativism remains extremely unpopular among contemporary metaethicists, as by my estimate less than 5% of respondents indicated sympathy for the view.
[7] To my knowledge nobody has surveyed intuitions specifically for 'that's false' in the metaethical case, although Khoo (2015) finds intuitions that it is inappropriate in the parallel case of disagreement involving epistemic modals.

(Ross & Schroeder 2013, Finlay 2014: 228, Khoo & Knobe 2016: 30; cf. Dietz 2008 on epistemic disagreement).

(iii) Experimental findings that in cases of "fundamental" normative disagreement between people who accept diverging norms, speakers often employ prosody or vocal stress in ways that are known indicators of non-content-based disagreement (Bolinger ms).

The second development is that the marketplace of ideas is newly flooded with many different relativist proposals about where the lost disagreement can be found. This includes at least Robinson 2009, Björnsson & Finlay 2010, Sundell 2011, Wong 2011, Plunkett & Sundell 2013, Khoo & Knobe 2016, Silk ms-a, ms-b, 2016, Bolinger ms.[8] Consequently, content-relativists today find ourselves confronting a new, more amiable problem about disagreement: a Problem of *Found* Disagreement.

> Found Disagreement: There are multiple competing candidates for non-content-based moral/normative disagreement in cases of intuitive disagreement where content-relativists deny there is any inconsistent content.

I'm reminded here of a sequence in Herge's graphic novel *Tintin in America*. Tasked with finding the kidnapped dog Snowy, a hotel detective turns up with an assortment of dogs: "You lost a dog? ...One single dog? Well, sir...I found you seventeen. And every one a pedigree pooch!" Which (if any) of these newly proposed kinds of disagreement is correctly identified as the *lost* disagreement—the moral or normative disagreement that is of central concern in metaethics—rather than kinds of disagreement that we weren't even looking for? This paper is an attempt at answering this question.

An initial but not necessarily simple task is to survey and classify the different species of solutions. The various proposals in the literature are formulated in different frameworks and language, and are often put forward without much acknowledgment or even awareness of each other. It isn't always easy to say how they are different or similar. What is needed is a careful

---

[8] Not to mention relevant proposals directed at parallel disagreement problems in epistemology, aesthetics, etc.

critical comparison that identifies the central distinguishing features of the different views on the playing field, and surveys their virtues and vices. In this paper I attempt to make a first step in this direction by comparing views falling on either side of a single distinction.

Extant relativist strategies for identifying non-content-based disagreement share a considerable amount in common. (1) They generally turn on the observation that utterances can communicate or *express* more than what they *say* (their content). This makes them pragmatic solutions, on a broad definition of 'pragmatics'.[9] (2) They locate the disagreement in (somehow) *conflicting attitudes*, which are pragmatically expressed or, in cases of disagreement in judgment alone, simply held. There is room for a variety of views on what kinds of attitudes those are; most accounts appeal to noncognitive or pro/con attitudes of some kind—for familiar reasons concerning the practical roles of normative judgment—but here we can stay neutral on this issue.[10] The major fault-line that particularly interests me here concerns the *object* of these conflicting attitudes.

The existing proposals seem largely to divide into two camps, even though many other approaches are surely possible.[11] On one hand, there are views according to which (some) utterances of 'S ought$_1$ to do A' and 'S ought$_2$ not to do A' stand in a relationship of normative disagreement by virtue of expressing conflicting attitudes towards the event-type, *S does A*. In this category we find at least Gil Harman (1996), David Wong (2011) and Gunnar Björnsson and myself (Björnsson & Finlay 2010, Finlay 2014).[12] I call this a *quasi-expressivist* account of disagreement: like expressivism it focuses on the attitude expressed by an utterance, but unlike

_____

[9] This includes but isn't limited to *conversational implicature* views, as found in Copp 1997, 2001, Finlay 2004, Strandberg 2012. Even on a broad definition, *conventional implicature* views, like those offered in Slote 1968, Barker 1999, and Copp 2001, fall into a grey area, as do *presuppositional* views, as in Perl ms. See Sundell 2011 for a catalog of kinds of pragmatic disagreement.

[10] See Perl ms. for a cognitivist account of pragmatic disagreement. Some of the views classified below as "metalinguistic" talk instead about conflicting *proposals*. Classifying these as "attitudes" may be awkward, but makes no difference for the arguments of this paper.

[11] Björnsson 2015 offers an account of non-content-based disagreement that doesn't obviously fit into either camp, for example. I do not address his account (in terms of judgments that fail to fulfill the communicative function of a claim) in this paper.

[12] Possibly also Dreier 2009, Sundell 2011: 282. Broadly parallel accounts of disagreement involving epistemic modals and other kinds of claims are offered by Dietz 2008, Montminy 2012, Huvenes 2014, Lennertz 2014.

expressivism-proper it attributes this to pragmatics rather than to semantics or meaning.[13] With tongue in cheek, I'll abbreviate this as:

> QED: Moral utterances of 'S ought/ought not to do A' with consistent contents sometimes stand in a relation of moral disagreement by virtue of pragmatically expressing conflicting attitudes towards the event-type, *S does A*.

On the other hand, there are views according to which utterances of 'S $ought_1$ to do A' and 'S $ought_2$ not to do A' sometimes stand in a relationship of moral or normative disagreement by virtue of expressing conflicting attitudes towards something like the use of the word 'ought' (or more broadly, what to do) in the conversational context. Such an approach is explicitly championed by David Plunkett and Tim Sundell (2013), Justin Khoo and Josh Knobe (2016), and Renee Bolinger (ms), and is also suggested by some claims of Denis Robinson (2009, 2010) and Alex Silk (ms-a, ms-b, 2016).[14] These views are commonly labeled *metalinguistic*, hence:

> MLD: Moral utterances of 'S ought/ought not to do A' with consistent contents sometimes stand in a relation of moral disagreement by virtue of pragmatically expressing conflicting attitudes towards the metalinguistic proposition, *if a speaker x is in context C, then x uses 'ought' with meaning M*.

This is only a rough and preliminary gloss on the metalinguistic approach, and some of the philosophers listed above may object to this characterization, if not to my classification of them itself. I work through some different formulations below in trying to do justice to their views.

It is important to acknowledge that QED and MLD are not mutually exclusive. Indeed several writers (myself included) have given voice to both views in different places. This may be partly due to a failure to distinguish between them, as their differences have so far received little if any attention, but there's also no clear reason why a proponent of one of these views should deny the existence of disagreements of the other kind. If two speakers have conflicting

---

[13] See Björnsson & Finlay 2010, Finlay 2014: 140. Harman uses the label *quasi-absolutist*, which highlights a different feature of this kind of view. Note that many metalinguistic views could be classified as broadly quasi-expressivist as well, but here I use it as a label of convenience in contradistinction to 'metalinguistic'.
[14] Pekka Väyrynen also proposed such an account to me in conversation in 2008.

attitudes towards S's doing A, then we can expect them also normally to have conflicting attitudes towards how normative words are used in relation to S's doing A, and vice versa. There is more room for reasonable debate over which is more deserving of the label of *moral*/*normative* disagreement, but even here it may be that the right verdict will ultimately turn out to be pluralistic, recognizing the existence of different species of moral/normative disagreement.[15] But as you might expect, in this paper my provisional findings are in favor of QED over MLD.

To be clear, I'm in complete agreement with the metalinguistic theorists that fundamental moral and normative discourse often *involves* metalinguistic disagreement, which contributes to the appeal of MLD accounts. What I'm less convinced of is that this is correctly identified as the lost moral/normative disagreement that motivates metaethicists' rejection of content-relativism. This lost disagreement, I believe, is more appropriately glossed as being over *whether S ought to do A*. I first consider a series of objections against MLD, some of which I argue to be decisive. I then consider some objections against QED, which I argue to be indecisive.

1. The Case Against MLD

The most general problem I see for MLD is that it intuitively mislocates the disagreement.[16] When we talk about *moral* disagreement between speakers uttering 'S ought to do A'/ 'S ought

---

[15] See also Sundell 2011, Bolinger ms. In Finlay 2014: Ch. 8, I distinguish four kinds of normative disagreement.
[16] It has been suggested that "fundamental" moral or normative disagreements fail a central linguistic test for metalinguistic negation, of resisting *neg-incorporation*. (For relevant discussion, see Plunkett & Sundell 2013: 33, Bolinger ms). Consider first:

(1) B: The miners are either in shaft A or shaft B.
     C1: No, they're not *either* in shaft A or shaft B; they're in A.
     C2: #No, they're *neither* in shaft A nor shaft B; they're in A.

This exchange features metalinguistic negation directed at the use of 'either'. Unlike ordinary content-directed negation, morphological incorporation of the negation (*not either…or ---> neither…nor*), as in C2, is impermissible when the negation is metalinguistic. However, contrast the following exchange, in which (let's assume) B and C endorse different moral standards:

(2) B: We ought to block shaft A.
     C1: No, it's not true that we ought to block A; we *ought* to block B.
     C2: No, we *oughtn't* block A; we *ought* to block B.

not to do A', we seem to be particularly interested in a conflict over what *S does* (or what *S is to do*). Identifying this with a conflict over *how to talk about what S does* seems to miss the most salient issue of disagreement.[17] By contrast, QED seems to get this right, locating the disagreement in a (broadly Stevensonian) conflict of attitudes towards S's doing A.[18]

I find this consideration prima facie compelling in favor of QED over MLD. But perhaps my articulation of the MLD claim has been uncharitable. On some statements of views that I've classified above as metalinguistic, the disagreement is characterized not as being over the use of words or sentences, but rather as being over the *context* (hence "metacontextual"). This is open to a variety of interpretations.[19] (i) Sometimes it is described as a disagreement about *what the context is*. Taken at face value this seems clearly mistaken. (It's a context where the two speakers each favor a different standard. Where's the room for disagreement about that?)[20] (ii) At other times it is described as a disagreement over what the context *ought* to be (or "is to be"): e.g. whether it ought to be a context where everybody accepts standard M_1, or one where everybody accepts standard M_2. This is more plausibly a normative disagreement, and I take it to be the more promising way to reconstruct the MLD approach. But it sharpens

---

In this case (the worry goes), incorporation of the negation (*not ought ---> oughtn't*) is fine, suggesting that the disagreement is not metalinguistic.

There are at least two reasons why MLD theorists shouldn't be troubled. First, according to many (contextualist) versions of content-relativism, the target of metalinguistic negation (i.e. the part of the first speaker's utterance that the second speaker objects to) is not the word 'ought' itself, but rather an argument or relativization of 'ought' which is left implicit. Since it is just the target of the metalinguistic negation which resists neg-incorporation, the difference between exchanges (1) and (2) is no problem. Second, 'oughtn't' is most naturally read as a contraction of 'ought not', rather than as an neg-incorporated version of 'not ought', and so C2 plausibly isn't a case of neg-incorporation at all. Additionally, Plunkett & Sundell (2013: 33) argue that metalinguistic disagreement need not involve metalinguistic negation. So we can dismiss this worry.

[17] It's tempting to also complain that MLD fails to give an account of disagreement over *whether S ought to do A*, but this might overreach, since metalinguistic disagreements are familiarly described in object-level terms; e.g. it seems natural enough to describe a metalinguistic disagreement over the use of 'bald' as a disagreement over *whether S is bald*. However, notice that it isn't similarly natural to describe this as a disagreement over *how much hair S has*, which corresponds, in the metaethical case, to describing a metalinguistic disagreement as being over *what S is to do*.

[18] I omit the motive to change the other person's attitudes that Charles Stevenson includes in his account of disagreement in attitude. Thanks to Giulia Pravato here.

[19] Cf. Sundell 2011: 279.

[20] 'Context' is likely being used with a different sense here, but other interpretations don't seem to make this claim any more plausible. E.g. if it is interpreted as meaning a set of parameters accepted as salient by all parties to the conversation, or as the "common ground", then there is (relevantly) no context in these scenarios, and both parties to such a disagreement would be mistaken.

the mislocation worry. The question of what context the speakers ought to be in seems *prima facie* too obliquely related to S's doing A to be the question at issue in moral disagreement over whether S ought to do A.[21]

Concommitantly, MLD faces difficulties in accounting for the intuitive *extension* of fundamental normative disagreement.[22] First, there is a problem of *interconversational disagreement*: two speakers can intuitively be in fundamental normative disagreement while participating in two separate conversations at different places and/or times, in which case (we might expect) they could not be disagreeing over what context to be in. (The context of what conversation?) Suppose Jeremy says 'One ought not to tell the truth to axe-wielding maniacs' while talking to John, and Immanuel says 'One ought to tell the truth even to axe-wielding maniacs' while talking to Gottfried. We can reasonably think that they have thereby disagreed over what to do when confronted by axe-wielding maniacs, without assuming that either has any view at all about how to talk when in the conversational circumstances of the other. (We might even suppose that Jeremy agrees with Immanuel about how to talk with Gottfried, because he believes that Gottfried perversely always tries to do what he thinks he ought not do.) By contrast, QED seems safe from this objection: two speakers can express conflicting attitudes towards S's doing A without participating in the same conversation.

Second, people can apparently be in fundamental normative or moral disagreement *in judgment*, without speaking at all. One could, perhaps, try to accommodate this on MLD as involving possession of conflicting attitudes towards what *kinds* of linguistic contexts to enter, or how to talk about S's doing A in general.[23] But this creates further problems, because people

---

[21] There is also a regress problem here: how is the content-relativist to analyze these second-order 'ought's? It seems likely that any reason to be skeptical that there is always a common contested content in first-order cases would also be a reason for skepticism about there always being a common contested content in second-order cases. Providing a metalinguistic diagnosis of this second-order disagreement (e.g. disagreement over what context to be in when talking about what context to be in when talking about whether S "ought" to do A) looks extremely unattractive, but if the MLD theorist offers a different kind of non-content-based strategy at this level, she must now explain why we shouldn't simply apply this different model of non-content-based disagreement at the first-order level instead.

[22] On this issue for content-relativism more generally, see also Lasersohn 2005, MacFarlane 2007: 20, Stephenson 2007, and discussion in Sundell 2011: 284f.

[23] For example, Khoo & Knobe suggest analyzing such disagreements "in terms of something about A's disposition to reject certain claims B has made or is disposed to make." (2016: 30n21)

can favor different contexts or speech behaviors for different reasons, and not merely because of which moral standards they accept. For example, I might prefer that Donald Trump continues to speak in a racist context or way (e.g. when he says, 'Undocumented immigrants ought to be rounded up and deported') for the reason that I want him to continue alienating reasonable Americans, while you might prefer that he switches to a different context for the reason that you fear the ratcheting up of racial tension. While there is certainly some disagreement between us in this scenario, it is surely wrong to say that we thereby disagree over *whether undocumented immigrants ought to be rounded up and deported*. So, it seems, we should not attempt to analyze the latter kind of moral disagreement as a disagreement over what context to be in.[24] This is a predictable kind of problem if we analyze disagreement over whether it ought to be that *p* in terms of attitudes toward something other than *p*. By contrast, QED is easily extended to accommodate these cases. Although no attitudes towards S doing A might be *expressed* in the absence of speech, we can still plausibly identify conflicting attitudes *held* towards that event-type, and identify fundamental normative disagreement in judgment about whether it ought to be that *p* with a conflict in attitudes held towards *p*.

At this stage, I anticipate one of two responses from MLD theorists. The first response concedes the points made above, but clarifies that MLD is proposed as an account only of one particular sense of disagreement, "activity" disagreement—which occurs when people are actively engaged in dispute with one another, contrasting with mere "state" disagreement.[25] This focus is explicit in the treatments of Silk, who labels it "discourse" disagreement (ms-a, ms-b), and Khoo and Knobe, who label it "conversational" disagreement (2016: 3). These philosophers may therefore protest that the above objections mistake the target of their analyses. In this case, my reply is simply that their MLD theories then do not really address the traditional metaethical Problem of Lost Disagreement, which extends to interconversational

---

[24] It might be objected that the Trump scenario doesn't involve distinctly *moral* attitudes on both sides, and for this reason doesn't involve a *moral* disagreement. But we can construct structurally identical scenarios involving moral attitudes, such as a dispute between a "Government House" Utilitarian who opposes invocation of genuine principles of morality in ordinary discourse, and a regular (transparency-favoring) Utilitarian.

[25] A distinction due to Cappelen & Hawthorne 2009. This focus is rejected in favor of state disagreements by (ironically) the account that most explicitly embraces the metalinguistic label, that of Plunkett and Sundell (2013: 10-11), precisely for the kinds of extensional issues raised above.

and judgment disagreements, and so is more plausibly interpreted as a problem about state disagreements.

The second response rejects my interpretation of MLD strategies as overly narrow. A champion of MLD might respond to these objections in the following way:

> "By *context*, I just mean a set of parameters or indices, of the kind that modal claims take as values. One such parameter is a *standard or end* parameter. So to have or express an attitude towards "the context", in the sense I intend, is simply to have or express an attitude (relevantly) towards a standard or end. Two people can express conflicting attitudes of this kind without being in the same conversation, and can possess such attitudes without being in a conversation at all."

This response would indeed blunt the objections raised above. But the cost of this route is that it threatens to collapse the distinction between MLD and QED, in a way that leaves it simply misleading to characterize fundamental normative disagreement as "metalinguistic", or concerned with "linguistic context". In no way are these attitudes helpfully understood as being directed towards anything "linguistic". (Some philosophers I classified above as MLD theorists might not see this as a cost, in which case they will turn out to be on the QED side of this dispute after all.)[26]

There is, however, a difference between having or expressing an attitude toward a standard/end, and having or expressing an attitude toward an event-type like S's doing A. Since I've here defined QED narrowly in terms of the latter, there remains a gap between QED and

---

[26] This paper began as a response to Alex Silk (particularly ms-a, but see also ms-b, 2016), trying to identify a difference between our otherwise very similar views. Silk now explains that his account of normative disagreement is neither "metalinguistic" nor "metacontextual", since "more fundamentally, [the] disagreement concerns…what moral norms to accept and why." However, this concession to QED at a more fundamental level does not, so far as I can see, license his denial that the primary account of ("discourse") disagreement in Silk ms-a, ms-b, and 2016 is metalinguistic/metacontextual. For example, the "precise sense" he identifies in which such speakers disagree, is in their conflict "over the grammatically backgrounded content of what value for the contextual deontic premise set variable is determined by the concrete conversational situation" (ms-b). Also: "[This] disagreement is given a precise representation: their utterances carry incompatible assumptions about what body of moral norms is operative in their context. The locus of their disagreement concerns the very contextual features which determine the contents of their deontic modal utterances" (2016: 128-9). One passage even seems to imply that all normative disagreement is metacontextual: "Their disagreement, rather, is fundamentally normative. It concerns what body of moral norms is operative in their conversational situation."

the metalinguistic thesis so interpreted. But if MLD leaves things there, it still hasn't solved its mislocation problem. We cannot plausibly identify moral disagreement over whether S ought to do A with disagreement over whether to accept standard M_1.[27] First, an *individuation problem*: standards/ends stand to event-types in a one-to-many relationship. This identification would therefore imply that claims about whether S_1 ought to do A_1 in C_1 can be in direct disagreement with claims about whether S_2 ought to do A_2 in C_2, where S_2, A_2, and C_2 are completely different agents, actions, and circumstances, respectively. To illustrate with my toy example from above, Jeremy would be disagreeing with Immanuel about whether one ought to tell the truth to axe-wielding maniacs no less if, instead of saying 'One ought not tell the truth to axe-wielding maniacs,' he were to say, 'Magistrates ought sometimes to convict those they know to be innocent, for the greater societal good'. He would still be disagreeing with Immanuel, in the same respect and to the same degree, even if he mistakenly believed that his favored moral standard did call for telling the truth even to axe-wielding maniacs, so that he would assent to Immanuel's sentence about axe-wielding maniacs if prompted. This seems quite implausible.

Second, the above response fails to escape another, more general problem for MLD, concerning cases of moral/normative *agreement* with diverging contents. Suppose speakers B and C accept competing moral standards, which as they recognize, happen to converge in licensing utterance of 'S ought to do A'. It seems natural to say that B and C *agree* that S morally ought to do A, whatever else they might disagree about. This can't be content-based agreement, since by hypothesis what each says is consistent with the negation of what the other says. This Problem of *Lost Agreement* is a widely overlooked corollary of the Problem of Lost Disagreement, and one which MLD theorists have generally failed to address at all. The difficulty for MLD is that B and C *dis*agree about what standard or end to accept, so it seems we can't locate their moral agreement in a consensus about the context. By contrast, the QED treatment of fundamental disagreement extends without any difficulty to these cases of agreement: B and C express or hold converging attitudes towards the event-type of S doing A, putting them in (roughly) a Stevensonian agreement in attitude toward it.

---

[27] This proposal can now be found in Silk ms-b and Perl ms.

There is a possible solution to both the *individuation* and *lost agreement* problems, which is to identify the disagreement/agreement with B and C having conflicting/identical attitudes toward the action-type, *accepting a standard that calls for S's doing A*. This solves the individuation problem, because such an attitude has a different object from an attitude toward accepting a standard that calls for S_2's doing A_2. And it solves the lost agreement problem, because B and C can agree on this much even while differing on which particular standard to prefer.[28] However, this solution introduces its own problems.[29] One is that it threatens to render normative claims trivially tautologous: '[In view of a standard that requires one to tell the truth even to axe-wielding maniacs,] one ought to tell the truth even to axe-wielding maniacs.'[30] At this point it becomes unclear why one would opt for content-relativism rather than simply an expressivist form of content-absolutism. (One way to resist this result would be to give diverging treatments of how the relevant standard is picked out for the semantic content and how it is picked out for the attitudinal content. But this looks seriously ad hoc.)

But if the MLD theorist has come this far, then I see no reason for not making one further step, which resolves both the individuation and the lost agreement problems without this further unpalatable consequence. If someone has/expresses an attitude favoring standard M, and believes/asserts that S's doing A is called for by M, then they can also be expected to have/express an attitude favoring S's doing A (as I've argued in Finlay 2004, 2014, and elsewhere). A conflict/consilience of such attitudes can then be identified with a moral disagreement/agreement over whether S ought to do A. But this just is QED. So if this is the right way to understand the MLD proposal, then there isn't really any difference between MLD and QED at all, except that MLD is misleadingly presented.

---

[28] Solutions of roughly this kind have now been embraced in both Silk ms-b and Perl ms.

[29] A reviewer points out that this solution may itself be vulnerable to a further "Government House" objection: conceivably someone might accept a standard M, and on this basis judge that S ought to do A, while at the same time being opposed to *acceptance* of standards which, like M, call for S to do A. If so, we shouldn't analyze agreement/disagreement over whether S ought to do A in terms of consensus/conflict over whether to accept such a standard. This problem is avoided by QED.

[30] Thanks to Zoë Johnson-King here. While this result would strike many as intolerable, it might seem hypocritical for me to press this objection, as I have myself suggested that some fundamental normative claims are, in respect of their semantic content, tantamount to useful tautologies (Finlay 2009, 2014: Ch. 7). However, the present suggestion on behalf of MLD is much more extreme, as it seems committed to analyzing all moral and normative claims as tautologous—since apparently every moral/normative claim is a possible target for fundamental or non-content-based disagreement and agreement.

Before I turn to examine the other side of the coin—potential advantages of MLD over QED—I'll briefly sketch one other, unconnected worry I have about MLD, about *transparency*. The concern is that metalinguistic analyses are poorly positioned to accommodate the fact that it is often—perhaps even usually—very difficult to know whether a moral or normative dispute is ultimately grounded in a difference in the basic standards or ends the disagreeing parties accept, or rather is grounded in a difference in their beliefs about what their standards or ends call for. (Ought-claims don't wear subscripts in public). It follows that it will be even more rare that it is mutually recognized by the parties to a moral or normative disagreement that their disagreement is of this kind. On MLD, this implies that it's very rare that we ever know whether we're arguing about what S is to do, or about what context we are to be in. This seems a difficult implication to swallow.[31] By contrast, the nontransparency of fundamental disagreement doesn't pose a serious problem for QED, because regardless of whether the *source* of the disagreement is a difference in standards/ends or a difference in beliefs about what conforms with those standards/ends (which may well not be transparent), there will be the same kind of disagreement in attitude toward S's doing A.

2. The Case Against QED

I am only aware of one advantage that MLD is alleged to have over QED as an account of normative disagreement, which concerns their treatments of "expressions of linguistic denial" (Plunkett & Sundell 2013, Silk ms-a, ms-b, 2016, Khoo & Knobe 2016, Bolinger ms). Contrast the following, adapted from Silk (ms-a):

(1) B: Elmer ought to receive the inheritance.

　　C1: No, Elmer ought not to receive the inheritance.

　　C2: That's false, Elmer ought not to receive the inheritance.

---

[31] But contrast Sundell 2011: 277, Plunkett & Sundell 2013: 20.

(2) B: I like Mexican.

     C1: #? No, I like Thai.

     C2: # That's false, I like Thai.

Exchange (1) is a paradigmatic case of normative disagreement, and a variety of kinds of negation markers, including 'No' and 'That's incorrect/wrong/false' are (it's claimed) perfectly fine. Exchange (2)—occurring in a context of conversation about where B and C are going to have dinner together—is a paradigmatic case of an expressed conflict in attitudes about an event-type. (Note that the relevant conflict here consists not in B liking Mexican food and C liking Thai food, but rather in B's preferring that B and C together have Mexican for dinner and C's preferring, incompatibly, that B and C together have Thai for dinner.) In this case, negation markers that seem fine in (1) are clearly out of place.

It's important to note that this data is ambiguous and contested. First, although some philosophers (e.g. Silk ms-a) claim that 'no' is infelicitous in an exchange like (2), I think it can be perfectly fine once an appropriate conversational context has been made clear. For example, B: "Where shall we go for dinner? I like Mexican." C: "No, I like Thai." On the other hand, responses like 'You're wrong/mistaken' and 'That's false/incorrect' are much more clearly infelicitous in (1) than in (2) (McKenna 2014, Dowell ms). We've already observed, however, that empirical studies have found that many people still find them infelicitous in many normative exchanges (Sarkissian et al 2011, Khoo & Knobe 2016); I'll return to this complication below.

In any case, the difference in felicity of 'that's incorrect/wrong/false' between exchanges like (1) and those like (2) poses a challenge for QED. As (1) shows, mere conflicts in (noncognitive) attitudes do not in general license use of these negation markers. But according to QED, fundamental normative disagreements just are such conflicts of attitudes. So it seems potentially embarrassing for QED that use of these markers is (more) felicitous for expressing fundamental normative disagreement. (In Björnsson & Finlay 2010 we label this the "semantic assessment problem"). By contrast, it is well-established that metalinguistic disagreement licenses a full range of negation markers. So this looks like a point in favor of MLD. Can QED

provide a satisfactory explanation why normative disagreements differ from expressed conflicts of attitude generally, in licensing these negation markers? Here I'll explore three different responses to the challenge.

A first, simple response is for QED theorists simply to help themselves to the metalinguistic explanation of these markers. As I've observed, there is no incompatibility between quasi-expressivist and metalinguistic forms of disagreement, and QED needn't deny that there is sometimes, often, or even ubiquitously metalinguistic disagreement going on when speakers engage in non-content-based moral disagreement. This response does not collapse QED into MLD, because here metalinguistic disagreement is invoked merely as an explanation of the use of negation markers like 'incorrect', 'wrong', and 'false', and not as an proposal to locate the lost disagreement over whether S ought to do A. To the extent that this semantic behavior is all that a particular MLD theorist is trying to explain, there therefore needn't be any dispute between us. However, an appeal to metalinguistic disagreement might arguably be insufficiently general. It seems that speakers can also use 'incorrect', 'wrong' or 'false' in evaluating people's moral *judgments* or opinions, and their moral claims in other conversations, and so the extensional problems for MLD raised above might create difficulties here too. In this respect, a QED-based solution might even be superior.[32]

A second idea is to identify the target of 'incorrect', 'that's false', etc. as something else that can be associated with the other's utterance, besides the asserted content. Pointing to the *attitude* expressed is a nonstarter here, given that the challenge is to explain the divergence between exchanges like (1) and (2). But there are other possibilities, such as (i) the sentence, (ii) something subsentential or subpropositional, like the predicate (e.g. "that's not true *of x*"), and (iii) the speech act. In earlier work (2010), to try to locate a propositional target for 'that's false' Gunnar Björnsson and I proposed that it engages with the sentence used, by targeting the

---

[32] One might also wonder whether MLD faces an equal but reverse problem of explaining the *in*felicity of C2 in (2). In principle, the metalinguistic story looks like it could extend to such conversations. By saying 'That's false; *I* like *Thai*', C would thereby reject B's assumption that their context is (or should be) one where *B* is the individual whose preferences determine where they go, instead proposing that the determinative preferences be those of C. One might argue that some element(s) of my QED-based solution below are necessary conditions for felicitous MLD disagreement using semantic negation markers.

proposition that the sentence would semantically express if used in the responder's preferred context (standards/ends) rather than the speaker's. In the case of exchange (1), this is the proposition that Elmer ought-relative-to-M_C (C's standard) to receive the inheritance. Since this is the proposition that C most relevantly (from C's point of view) rejects, we argued that C can felicitously respond to B by targeting this proposition with 'that's false'—thereby also pragmatically expressing a negative attitude toward Elmer receiving the inheritance.

Whatever you might think of this proposal, it successfully identifies a difference between exchanges (1) and (2). For in the case of (2), B's sentence ('I like Mexican') cannot be used, in anybody's mouth, to *assert* the proposition that C is most relevantly interested in rejecting, which concerns rather B and C's together going to dinner at a Mexican restaurant. This is a meaningless victory, however, since we can easily come up with a modified version of (2) where this difference evaporates:

(3)   B: I prefer that we have Mexican for dinner.
        C: #That's false.

Clearly, there isn't an available reading of C's response as denying the proposition that *C* prefers that they have Mexican for dinner, even though that is what B's sentence would semantically express in C's mouth, given the indexicality of 'I'. But without some further story, that's what our proposal would seem to predict (Lennertz 2014).

For this and other reasons (including the scruple that '*that's* incorrect/wrong/false' should refer to something that the responder believes the speaker actually put forward by her utterance), I no longer favor that solution—at least by itself. A further (complementary) solution is to appeal to *semantic opacity*.[33] Even if a content-relativist (e.g. contextualist) theory of the semantics of normative language is correct, it is evident that ordinary speakers do not in general have a reflective or theoretical awareness of this fact. Observe how controversial content-relativism is in metaethics, even among the supposed experts. Given (i) this semantic

---

[33] Often discussed under the label of 'semantic blindness'. For a fuller discussion see Finlay 2014: 236-45, and for parallel appeals to address problems of lost disagreement involving 'knows' and 'might', see Montminy 2009 and Bach 2009, respectively.

opacity, (ii) that according to QED there really is a kind of normative disagreement being expressed whenever one speaker makes the moral claim 'S ought$_1$ to do A' and another makes the moral claim 'S ought$_2$ not to do A', *and* (iii) that these utterances do have ordinary propositional contents, we can reasonably expect that in ordinary moral discourse speakers would assume that they are disagreeing over a single proposition. If this is right, it suffices to explain the use of negation markers like 'incorrect', 'wrong', and 'false' in exchanges like (1)— an explanation that clearly doesn't overgeneralize to conversations like (3) or (2), since no sensible person fails to recognize that content-relativism is true for the pronoun 'I'. This solution is also easily extended to the assessment of private moral judgments and moral claims made in other conversations.

To further support this explanation, recall the final objection I raised against MLD theories, concerning the nontransparency of whether or not a normative disagreement is fundamental (involves a difference in standards/ends). This nontransparency works in favor of QED here, as it provides a further reason why speakers engaged in a fundamental disagreement might fail to recognize that there is no content-based disagreement between them. Often, and perhaps even usually, moral or normative disagreement is (at least in part) disagreement over a common proposition, and it is often hard to distinguish when this is and when it isn't the case. This lends further credibility to the hypothesis that speakers assume their moral disagreement is content-based even when it isn't.

Finally, to keep this semantic assessment problem in perspective it is important not to forget that the evidence for the felicity of semantic negation markers is far from univocal. Recall the experimental findings of a statistically significant resistance in ordinary subjects toward the use of markers like 'incorrect', 'wrong', and 'false' in the context of (clearly and explicitly) fundamental disagreements (Sarkissian et al. 2011, Khoo & Knobe 2016). Such a reticence fits well with the QED account of these disagreement, given the observed datum that mere conflicts in attitudes (as in exchange (2)) do not in general license such use.

Of course, throwing the kitchen sink at the negation marker problem as I am doing here raises further issues, since some of the responses I've just offered seem to cut in opposing directions:

18

my appeals to metalinguistic disagreement and the observed resistance to negation markers suggest a certain level of semantic *transparency*, in prima facie tension with my appeal to semantic opacity. However, it seems fair to characterize the overall evidence as ambiguous, in a way that QED is well-placed to explain but that is potentially awkward for rival treatments of moral and normative disagreement including MLD. Metaethical controversy over content-relativism, together with the fact that the experimental findings are of a merely *statistically significant*—but far from universal—reticence with negation markers, point toward an intermediate *degree* of semantic opacity, such that awareness of the truth of content-relativism apparently varies from person to person, and on the spectrum between tacit and explicit. It seems a plausible hypothesis that the extent to which individual speakers are resistant to markers like 'incorrect', 'wrong', and 'false' varies corresponding to the degree they are aware of the content-relativity of normative utterances.

The conclusion of this paper is therefore that whereas metalinguistic accounts of moral or normative disagreement face serious problems—problems that force them in the direction of quasi-expressivism—the main challenge raised against quasi-expressivist accounts is amenable to a well-motivated solution. With respect to the choice between these two strategies, content-relativists should identify fundamental moral and normative disagreement over whether S ought to do A with a conflict of attitudes towards S doing A: QED.[34]

References

Bach, K. (2009) "Perspectives on Possibilities: Contextualism, Relativism, or What?" In A. Egan & B. Weatherson (eds.) *Epistemic Modality*. New York: Oxford University Press, 19-59.

Barker, S. (1999) "Is Value Content a Component of Conventional Implicature?" *Analysis* 60: 268-79.

Björnsson, G., & Finlay, S. (2010). "Metaethical Contextualism Defended." *Ethics* 121(1), 7-36.

Björnsson, G. (2015). "Disagreement, Correctness, and the Evidence for Metaethical Absolutism." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 10. Oxford: Oxford University Press.

Bolinger, R. (ms). "Metalinguistic Negotiations in Moral Disagreement."

Cappelen, H. & Hawthorne, J. (2009). *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Copp, D. (1997). *Morality, Normativity, and Society*. Oxford: Oxford University Press.

Copp, D. (2001). "Realist-Expressivism: a Neglected Option for Moral Realism." *Social Philosophy & Policy* 18(2): 1-43.

Dietz, R. (2008). "Epistemic Modals and Correct Disagreement." In. M. García-Carpintero & M. Kölbel (eds.), *Relative Truth*. Oxford: Oxford University Press.

Dowell, J. (ms). "Flexible Contextualism about Deontic Modals: Puzzles about Moral Disagreement."

Dreier, J. (2009). "Relativism (and Expressivism) and the Problem of Disagreement." *Philosophical Perspectives* 23: 79-110.

Finlay, S. (2004). "The Conversational Practicality of Value Judgement." *Journal of Ethics* 8(3): 205-23.

Finlay, S. (2009). "Oughts and Ends." *Philosophical Studies* 143(3): 315-40.

Finlay, S. (2014). *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press.

Harman, G. (1996). "Moral Relativism." In G. Harman and J.J. Thompson (eds.) *Moral Relativism and Moral Objectivity*. Blackwell: 3–64.

Huvenes, T. (2014). "Disagreement Without Error." *Erkenntnis* 79(1): 143-54.

Khoo, J. (2015). "Modal Disagreements." *Inquiry* 5: 1-24.

Khoo, J., & Knobe, J. (2016). "Moral Disagreement and Moral Semantics." *Noûs*.

Lasersohn, P. (2005). "Context Dependence, Disagreement, and Predicates of Personal Taste." *Linguistics and Philosophy* 28: 643-86.

Lennertz, B. (2014). *Reasoning with Uncertainty and Epistemic Modals*. PhD dissertation, University of Southern California.

Lyons, D. (1976). "Ethical Relativism and the Problem of Incoherence." *Ethics* 86: 107-21.

MacFarlane, J. (2007). "Relativism and Disagreement." *Philosophical Studies* 132(1): 17-31.

MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.

McKenna, R. (2014). "Disagreeing about 'Ought'." *Ethics* 124 (3): 589-597.

Montminy, M. (2009). "Contextualism, Invariantism, and Semantic Blindness." *Australasian Journal of Philosophy* 87 (4): 639-657.

Montminy, M. (2012). "Epistemic Modals and Indirect Weak Suggestives." *Dialectica* 66 (4): 583-606.

Moore, G. E. (1922). *Philosophical Studies*. London: Kegan Paul.

Olson, J. (2011). "In Defense of Moral Error Theory." In M. Brady (ed.), *New Waves in Metaethics*. New York: Palgrave MacMillan, 62-84.

Perl, C. (ms). *Deontic Positivism*. PhD dissertation, University of Southern California.

Plunkett, D., & Sundell, T. (2013). "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13 (23).

Robinson, D. (2009). "Moral Functionalism, Ethical Quasi-Relativism, and the Canberra Plan." In D. Braddon-Mitchell & R. Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*. MIT Press.

Robinson, D. (2010). "Reflections on Moral Disagreement, Relativism, and Skepticism about Rules." *Philosophical Topics* 38 (2): 131-56.

Ross, J., & Schroeder, M. (2013). "Reversibility or Disagreement." *Mind* 122(485), 43-84.

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). "Folk Moral Relativism." *Mind and Language* 26(4): 482-505.

Silk, A. (ms-a). "Normativity in Language and Law." Forthcoming in D. Plunkett, S. Shapiro & K. Toh (eds.), *Legal Norms, Ethical Norms: New Essays on Metaethics and Jurisprudence*. Oxford: Oxford University Press.

Silk, A. (ms-b). "Normative Language in Context." Forthcoming in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 12. Oxford: Oxford University Press.

Silk, A. (2016). *Discourse Contextualism: A Framework for Contextualist Semantics and Pragmatics*. Oxford: Oxford University Press.

Slote, M. (1968). "Value Judgments and the Theory of Important Criteria." *Journal of Philosophy* 65(4): 94-112.

Stephenson, T. (2007). "Judge Dependence, Epistemic modals, and Predicates of Personal Taste." *Linguistics and Philosophy* 30: 487-525.

Strandberg, C. (2012). "A Dual Aspect Account of Moral Language." *Philosophy and Phenomenological Research* 84(1): 87-122.

Sundell, T. (2011). "Disagreements about Taste." *Philosophical Studies* 155: 267-88.

Thomson, J. J. (1996). "Moral Objectivity." In G. Harman and J.J. Thompson (eds.) *Moral Relativism and Moral Objectivity*. Blackwell.

Westermarck, E. (1906-08). *The Origin and Development of the Moral Ideas*. New York: MacMillan.

Westermarck, E. (1932). *Ethical Relativity*. New York: Kegan Paul.

Wong, D. (2011). "Relativist Explanation of Interpersonal and Group Disagreement." In S. Hales (ed.), *A Companion to Relativism*. Malden, MA: Wiley-Blackwell, pp. 411-29.