ARTICLE

Moral Kombat: Analytic Naturalism and Moral Disagreement

Edward Elliott 🆔 and Jessica Isserow 🆔

School of Philosophy, Religion and History of Science, University of Leeds, Leeds, UK
Corresponding author: Edward Elliott; Email: e.j.r.elliott@leeds.ac.uk

Abstract
Moral naturalists are often said to have trouble making sense of inter-communal moral disagreements. The culprit is typically thought to be the naturalist's metasemantics and its implications for the sameness of meaning across communities. The most familiar incarnation of this metasemantic challenge is the Moral Twin Earth argument. We address the challenge from the perspective of analytic naturalism and argue that making sense of inter-communal moral disagreement creates no special issues for this view.

Keywords: Moral Twin Earth; moral naturalism; analytic naturalism; moral functionalism; moral disagreement; verbal disputes

1. Title Screen

Moral disagreements are often as frustrating as they are familiar. Consider:

> **Sexist Kano.** Sonya has been known to decry the unequal opportunities extended to men and women. She asserts: 'Policies promoting gender inequality are wrong'. But Kano insists that such inequalities are acceptable, retorting, 'Policies promoting gender inequality are not wrong'. They don't seem to disagree about any relevant empirical matters.

Moral disputes like these can sometimes feel intractable. Still, we tend to assume that they reflect genuine disagreements. By the look of it, it is unlikely that Sonya and Kano are merely exchanging noises with only the superficial appearance of communication and conflict. If she is right, then he is wrong.

If you are a moral realist, then it is easy to see how you might make sense of the possibility of genuine moral disagreement in such cases. Suppose that Sonya and Kano belong to the same community and they are both competent speakers of the same language. In that case, Sonya and Kano presumably mean the same things by their terms, and whereas Sonya asserts some proposition *P*, Kano contradicts it by asserting ¬*P*. There is no miscommunication, the dispute is not merely verbal, just a straightforward conflict in the beliefs expressed by their respective assertions.

But moral realists should want to make sense of inter-communal disagreements as well as intra-communal ones. Suppose now that Kano has spent his life on Outworld, where people think somewhat differently about the moral acceptability of gender inequality, and has recently come to visit Sonya in Earthrealm. No doubt this raises the probability that Sonya and Kano have simply been talking past one another, but not overwhelmingly so: even with those additional assumptions, the disagreement is still apt to strike us as more than merely verbal. Generally speaking, real moral disagreements seem possible both within and across communities—even across

communities with different moral outlooks. (Few report imaginative resistance when watching the apparently genuine moral disputes that regularly play out across intergalactic communities on *Star Trek*.) And should some metaethical theory end up implying otherwise, then, well, so much the worse for that theory.

Naturalist moral realism (henceforth just "naturalism") is often thought to have special trouble in this regard. Very roughly, the problem is as follows. Since moral judgments are, according to the naturalist, in the business of describing mind-independent aspects of the world, we ought to be able to tell essentially the same metasemantic story for moral vocabulary as we would for any other part of our descriptive vocabulary. However, such stories will typically imply some dependence of meaning on one's local (natural or linguistic) environment. For instance, in just the same way that one might take the meaning of "water" within a community to depend upon how that term is used by its members as well as what happens to be around for that usage to latch onto (H_2O vs. XYZ), so too one might think the meaning of "wrong" depends upon how it is used and which natural properties happen to be floating about the local environment. If two communities use their moral terms differently, or if the local natural properties are somewhat different, then there is a difference in meaning. Having established that relativism of meaning, it is then possible to imagine two otherwise similar communities that differ with respect to whatever local metasemantic facts ground the meanings of their respective moral vocabularies. Since the local facts differ, the meanings of their moral terms differ—and so, the argument goes, any apparent moral disagreement across members of those communities will be just that: *merely* apparent. Thus, when Sonya says "Policies promoting gender inequality are wrong," she is asserting one thing P, whereas when Kano from Outworld says "Policies promoting gender inequality are not wrong," he is asserting $\neg Q$, and they are not really disagreeing. Cue *reductio*, reject naturalism, *QED*.

This style of argument has been popularized especially by Horgan and Timmons (henceforth "H&T"), who describe a hypothetical Moral Twin Earth ("MTE") populated by individuals with whom we purportedly cannot have any real moral disagreements should naturalism prove true (Horgan & Timmons, 1991, 1992, 2000, 2009). If this argument works, then there is a metasemantic fly in the naturalist's metaethical ointment. However, the argument *does not* work. The problem, we will argue, concerns the connection between meaning and disagreement—namely, there is none. The possibility of genuine disagreement requires neither sameness of meaning, nor even a conflict in the propositions expressed; indeed, it does not directly have anything to do with *what is said* at all, and everything to do with the attitudes of those doing the saying.

In this article, we will be responding to the metasemantic challenge from the perspective of analytic naturalism. It is unclear to us why there have been so few attempts to do so. Perhaps the lack of discussion can simply be chalked up to the fact that analytic naturalism is often dismissed in metaethical circles as yet another curious oddity of the Antipodes—much like a taipan or a funnel-web spider, it might be fascinating to look at but it is not the sort of thing many will want to find skulking around their office. To be sure, H&T's original discussion was targeted at Boyd's (1988) non-analytic version of moral realism, and most discussions since have centered upon synthetic moral naturalism. But the challenge was never intended to be so restricted in scope. In H&T's words, the "generic thought-experimental deconstructive recipe… is applicable to virtually any metaphysical naturalist version of moral realism" (Horgan & Timmons, 2009, p. 221; see also Horgan & Timmons, 2000, pp. 139–142); and in their (2009), H&T make analytic moral functionalism their primary target. In any case, every paper needs a starting point, and we will not waste space defending our own. Some parts of our response are specific to analytic naturalism, others can be repurposed by synthetic naturalists if they so choose. We leave readers to draw such connections for themselves.

We will begin with some stage setting: Section 2 covers everything you need to know about analytic naturalism, while Section 3 goes into further detail on the metasemantic challenge as it

applies to that position. We present our main response in Sections 4–6 and then deal with a related but separable issue in Section 7.[1]

2. Character Select: Analytic Naturalism

The main ideas of analytic naturalism parallel those of analytic functionalism for mental states. The concepts of BELIEF, DESIRE, PAIN, PLEASURE, and so on, are a common property—that is, they belong to our shared theory of the mind, so-called *folk psychology*—and each has a role to play within that theory. Following a familiar story that originates with Ramsey (1931) and finds its fullest expression with Lewis (1970), we can analyze the meaning of our term "belief," for instance, by reference to its role within that theory. Thus, according to the analytic functionalist, it is a priori that "belief" picks out whatever it happens to be that plays the role that "belief" is supposed to play within folk psychology, if anything does so play that role; or, if nothing does, then "belief" picks out whatever comes closest to playing that role, provided it plays the role well enough. In short: "belief" picks out the best deserver of the "belief"-role.

As with the mind, so too with morality. The concepts of RIGHT, WRONG, GOOD, BAD, OBLIGATORY, PERMISSIBLE, and so on, are as much a common property as the concepts found within folk psychology, and each has a role to play within *folk morality.* Hence the analytic naturalist proposes to analyze the terms of our moral vocabulary by reference to their respective roles within this folk moral theory. The analytic naturalist will, for instance, want to say that our term "right" can be analyzed in terms of the "rightness"-role in our folk moral theory.[2]

It will be helpful to say a little more about our key terms, starting with "the folk." There is plenty of scope for debate here: perhaps "the folk" are all human beings, provided we humans typically share more or less the same moral opinions, practices, and intuitions. Or perhaps there are distinct folk moral theories for different populations. These are matters for debate *within* analytic naturalism and should not matter for our discussion. We will simply take "the folk" to be all human beings.

There is likewise plenty of scope for debate in how we should understand the "folk moral theory." You might think of it as a collection of *platitudes*—claims about morality with which members of the folk are disposed to acquiesce or explicitly believe. Better: you might think that folk morality is something the folk only tacitly believe, which may even conflict with what they are disposed to assert. Much like our tacit understanding of grammar, a complete expression of the folk moral theory may involve complicated ideas and principles that need not be apparent even to those who regularly use and adhere to them. Better still: you might think of folk moral theory as an abstraction of the principles that best explain the way the folk are disposed to act and think about morality *on average*, without requiring that the total suite of these principles constitute the contents of any individual's beliefs (tacit, ethnocentric, or otherwise). Indeed, it may not even be the folk moral theory properly so-called that we are interested in, but instead a future version of the theory that has been tidied up and systematized—a "mature folk morality," as Jackson (1998) puts it. Again, though, these are matters for debate *within* analytic naturalism, and should not much impact our

---

[1]At the risk of spoiling a joke by explaining it, we should say something about our section headings for those who may otherwise be perplexed. *Mortal Kombat* is a video game in which two characters (often from different "dimensions") fight one another. Matches have a best-two-out-of-three structure, and a player scores a "flawless victory" by winning the first two rounds without incurring damage. At the end of a game, the announcer may ask the winner to "Finish him!"; and following a victory, players may also sometimes play a separate "Test your might" mini-game.

[2]There are three claims bundled together here: (1) *metasemantic*: the meaning of "right" depends on our folk moral theory; (2) *semantic*: "right" designates the best deserver of the "rightness"-role in our folk moral theory; (3) *epistemic*: the foregoing equivalence in meaning is analytic, knowable a priori. If we were being more careful, we would precisify these claims about "meaning" in terms of primary intensions, as those are understood within the two-dimensionalist framework (see especially Chalmers, 2006 and Jackson, 1998). But such details need not concern us, as two-dimensionalism will not play an important part in the ensuing discussion.

discussion. What matters is that there is something we will call a *folk moral theory* that is closely tied to the folk's moral opinions, practices, and intuitions, and that our moral vocabulary is analyzed by reference to that theory in a manner akin to how analytic functionalists analyze folk-psychological terms by their roles within folk psychology.

Finally, we should emphasize that it is no requirement of analytic naturalism that *every* claim or principle within folk moral theory must end up being correct. Analytic naturalists *are not* proposing to reduce moral theorizing to the mere cataloging of folk moral theories! In their view, what the folk moral theory *says* is analytically equivalent to the claim that the moral properties of *rightness*, and *wrongness*, and *good* and *bad*, and so on, all perfectly occupy their respective roles within that theory (see Lewis, 1970). Consequently, if any of these properties *imperfectly* occupy their roles, then our folk moral theory must be at least somewhat mistaken. *And it may very well be mistaken.* In much the same way that folk physics conflates the concepts of WEIGHT and MASS, folk morality may say that there is a single moral property that plays two roles when really there are two properties that play one role each. Or perhaps folk morality says that certain kinds of actions are obligatory, when in fact they end up being merely permissible (or even just wrong).

One therefore should not make the mistake of supposing that, according to analytic naturalism, a term like "obligatory" simply ends up referring to whatever disjunction of actions should be deemed obligatory according to our folk moral theory. There is a lot more to the "obligatory"-role than just which actions the term "obligatory" does and does not get attached to, and it need not be the case that every aspect of that role ends up being satisfied. (This will be important.) It is, therefore, entirely possible to disagree with some aspect of the folk moral theory even while accepting that our moral terms are properly analyzed by reference to their role within that theory in the manner characterized above. Indeed, it is not at all likely that we *should* sometimes disagree with what our folk moral theories have to say, and the moral judgments they predict. Doing so does not entail that we have somehow lost our capacity to think moral thoughts or to meaningfully converse with those who might have differing moral opinions. (This will also be important.)

3. Opponent Select: MTE

Now the challenge. As we have noted, analytic naturalists take the meaning of a moral term like "obligation" for a population to depend upon the character of that population's folk moral theory, such that if that theory had been different then the meaning of "obligation" may have been correspondingly different. According to H&T, this reveals that analytic naturalists are not, in fact, *realists* at all, but merely relativists dressed in realist garb. The naturalist is, accordingly, thought to have much the same trouble accounting for genuine disagreement as relativists supposedly do.

H&T's scenario (adjusted to fit our theme) goes like this. Outworld is as close to Earthrealm as possible consistent with the following stipulation: whereas we humans of Earthrealm have our folk moral theory, M^{earth}, the twin-humans (Outworlders) of Outworld have some alternative theory, M^{out}. By hypothesis, M^{earth} and M^{out} must be distinct, but also similar. They will be similar not only with respect to which actions and people and institutions and so on that they imply will be "right" and "good" and "obligatory" and such, but also with respect to how the theories are integrated into their respective societies and broader normative theorizing. The Outworlders' use of terms like "good" and "right," for instance, are assumed to

> … bear all the formal marks that are usually taken to characterize moral vocabulary and moral practice. In particular, the terms are used to reason about considerations bearing on the well-being of persons on [Outworld]; [Outworlders] are normally disposed to act in certain ways corresponding to judgments about what is "good" and "right"; they normally take considerations about what is "good" and "right" to be especially important, even of overriding importance in most cases, in deciding what to do… (Horgan & Timmons, 2009, p. 225)

We might imagine that M^{earth} is essentially consequentialist in character, while M^{out} is more deontological but still overlaps a good deal with M^{earth}. This is how H&T themselves describe the scenario, but what is most important is that the following are satisfied:

Similarity. M^{earth} and M^{out} are similar, including *inter alia* with respect to how they integrate into society and general normative theorizing.
Divergence. M^{earth} and M^{out} differ with respect to the "obligatory"-role, such that Sonya and Kano's respective uses of the terms would have divergent meanings and extensions given analytic naturalism.

Given that, suppose Sonya of Earthrealm has encountered Kano of Outworld, perhaps at some interdimensional ethics tournament. We presume they each more or less accept the tenets of their respective folk moral theories. Sonya utters the words "Giving to charity is obligatory," which is true according to M^{earth}. Kano retorts with "Giving to charity is not obligatory," which is true according to M^{out}. Is there a genuine disagreement here, or is their dispute merely verbal?

It certainly seems like there *could be* something substantive underlying their dispute, and perhaps even that there *probably is.* This is how many people's intuitions go, including our own, and such intuitions are at least in part a result of **Similarity**. Compare a case where **Similarity** is *not* satisfied—for instance, one where Outworlders are only disposed to classify an action as "right" if it involves significant cardiovascular activity, and "permissible" when it involves going to the gym, and they assert that the regular performance of "obligatory" actions promotes fat loss, muscle gain and increased life expectancy, and so on. In this case, it would be much more tempting to say that the Outworlders are just using the words that we associate with moral properties to instead talk about exercise, and we would be much more inclined to classify the dispute as merely verbal. So **Similarity** is there to ensure the firm intuition that Sonya and Kano could, and indeed probably do, have a substantive disagreement.

However, from **Divergence** it follows that if analytic naturalism is correct, then Sonya and Kano mean different things by their respective uses of "obligatory." This is the basis for the challenge. The correct explanation for the intuition of disagreement, in H&T's view, is that:

… moral and twin-moral terms do *not* differ in meaning or reference, and hence… any apparent moral disagreements that might arise between [humans and Outworlders] would be genuine disagreements—i.e., disagreements in moral belief and in normative moral theory, rather than differences in meaning. (Horgan & Timmons, 2009, p. 227)

But the analytic naturalist *cannot* say this:

The moral terms used by [humans] designate the unique natural properties that respectively satisfy the Lewis-style conceptual analyses of those terms obtainable from theory [M^{earth}], whereas the twin-moral terms used by [Outworlders] designate *distinct* unique natural properties that respectively satisfy the respective conceptual analyses obtainable from [M^{out}]; hence, because corresponding moral and twin-moral terms have different, incompatible, conceptual analyses, moral and twin-moral terms *differ in meaning*, and are not intertranslatable. (Horgan & Timmons, 2009, pp. 226–227)

And so:

The different parties are expressing different concepts with their moral terms, are talking past one another rather than disagreeing. (Horgan & Timmons, 2009, p. 232)

In more detail: the role played by "obligatory" in M^{earth} must *diverge enough* from its role in M^{out} that Sonya's and Kano's respective uses of the term have *incompatible* analyses, in which case

Sonya's "obligatory"^earth will designate a distinct property than "obligatory"^out does in Kano's mouth. We assume that two predicates have *incompatible* analyses just in case those analyses are such that, at least at the worlds in question, there are things in the extension of each that are not in the extension of the other.[3] Thus incompatibility guarantees a difference in extension, and the difference in extension explains why Sonya and Kano's assertions are not in conflict. Essentially: what Sonya says is true if and only if giving to charity has some property *f*, and what Kano says is true if and only if giving to charity lacks some property *g*, where presumably being *g* neither implies nor is implied by being *f*—and hence they can both be *correct* when between them they assert that an action both has *f* and lacks *g*. This is supposed to entail that the two are engaged in a verbal dispute.

The upshot is that analytic naturalists are unable to vindicate the intuition that Sonya and Kano likely have a genuine disagreement, precisely because analytic naturalism predicts a difference of meaning and therefore that the dispute is merely verbal. Or to put the same point in the other direction: the intuition of genuine disagreement is supposed to provide good evidence that Sonya and Kano *mean the same thing*, and thus that intuition counts against those views which (like analytic naturalism) predict a difference in meaning.

But the problems do not end there! If analytic naturalism is correct, H&T add, then Sonya and Kano cannot even possess one another's concepts, and hence in principle cannot understand one another:

> … agents who have a [folk moral theory] different from that of humans would not possess the concepts of GOODNESS, RIGHTNESS, etc., at all. (Horgan & Timmons, 2009, p. 228)

Specifically, H&T argue that analytic naturalists are guilty of *chauvinistic conceptual relativism*: were we to encounter some non-human population with a folk moral theory distinct from our own, then members of this population would possess different moral concepts than we do—or perhaps would not even possess *moral* concepts at all, properly so-called, but *tworal* concepts instead. Analytic naturalism therefore "chauvinistically builds the folk morality supposedly shared by all of humankind directly into moral concepts themselves" and is as such "objectionably human-centered" (Horgan & Timmons, 2009, p. 228). We humans are conceptually cut off from understanding the Outworlders' tworal assertions, and they from understanding our moral assertions, precisely because of the differences between our folk moral theories.

In summary, the core allegation is that (1) analytic naturalism cannot accommodate the intuition that Sonya and Kano have a genuine disagreement, due to a difference in the meanings of their terms; and a secondary allegation is that (2) Sonya and Kano are not even *capable* of understanding one another, because they must lack the requisite concepts. The latter allegation is not essential to the former, but it is an independently important challenge that we will address separately in Section 7. (Spoiler alert: we reject it.) Our main response—that is, to the core allegation—will be in Sections 4 and 5, with a bit of mopping up in Section 6. We will argue that, given analytic naturalism, it is not just *possible* but indeed *more likely than not* that Sonya and Kano's dispute is substantive. In doing so, we will *not* assume that Sonya and Kano fully understand one another—they may or may not share moral concepts, but *whether* they do is not crucial to making sense of their disagreement as substantive.

4. Round One: Disagreement and Disputes

The key step of the MTE argument is to treat the intuition of genuine disagreement as compelling evidence that the disputees mean the same thing by their terms. As Finlay (2017, pp. 188–90) notes,

---

[3]H&T never explain what they mean by "incompatible," but this is the definition that makes the most sense of their argument. The specifics do not matter so much, though, as long as the analyses determine divergent extensions such that Sonya and Kano can both be speaking truths.

the typical naturalist response has therefore been to question whether the explanation of that intuition really does require us to posit utterances with incompatible or inconsistent contents, and more generally whether it requires positing sameness of meaning for the terms involved (in this or that sense of "meaning"). Indeed, as Khoo and Knobe recently 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. (2018, p. 110)

This is usually coupled with a proposal to locate the sense of disagreement in a conflict of *attitudes* rather than a direct conflict in *what is said.* (For some examples, see Björnsson & Finlay, 2010; Bolinger, 2022; Finlay, 2014; Geirsson, 2005; Khoo & Knobe, 2018; Plunkett & Sundell, 2013; Silk, 2017; see also the further references in Finlay, 2017.)

There are two main versions of this response, what Finlay (2017) calls the *quasi-expressivist* and *metalinguistic* responses. According to the former, Sonya and Kano's utterances can pragmatically express conflicting attitudes (usually non-cognitive) regarding the act of giving to charity, even if they differ in meaning and even if they do not have conflicting contents. According to the latter, their utterances might instead indirectly express conflicting attitudes (usually cognitive) toward some metalinguistic proposition concerning how the terms ought to be used. We are going to take a third route. We will argue that, from the analytic naturalist perspective, it is possible to make sense of Sonya and Kano's dispute as substantive—that is, as likely indicative of a conflict in their beliefs, and *not* just their beliefs regarding metalinguistic propositions—even if they mean different things by their terms and even if they are not expressing incompatible contents. In saying this, we do not mean to suggest that the existing quasi-expressivist and/or metalinguistic strategies are mistaken. We see our response as complementary, not necessarily competing; our point is that analytic naturalists do not *need* to appeal to either metalinguistic negotiation or quasi-expressivist disagreement to explain the intuition of substantive disagreement in MTE scenarios.

This section is mostly setup; we want to establish some terminology and some initial key points concerning substantive versus merely verbal disputes.[4] First, it will be helpful to separate the mental phenomena that we will refer to as *disagreements* from the linguistic phenomena that we will call *disputes.* To get a feel for the distinction, consider:

> **Academic Kano.** While Sonya presents her latest research on interdimensional travel, Kano has a complicated thought: he gets the sense that he disagrees with Sonya's thesis, but struggles to verbalise why. During the Q&A, Kano attempts to explain his disagreement— and fails. He understands and means every word he says, but what he says does not quite capture his thoughts. Following discussion, it comes to light that what he said is not inconsistent with Sonya's thesis. Nevertheless, Kano has the distinct feeling that had he expressed his thoughts better, then an inconsistency would have been apparent.

Here is what Kano *does not* say in this case: "Well, what I *said* wasn't inconsistent with what Sonya said, so I guess we don't disagree after all." The disagreement itself, and Kano's attempts at *expressing* that disagreement, are two very different things.

So let us henceforth say that a *disagreement* is a relation that holds between two agents regarding some question. A *question* we take to be just a way of partitioning logical space into a set of possible

---

[4]Parts of the discussion in this section overlaps in various ways with points made by, *inter alia*, Chalmers (2011), Geirsson (2005), Jenkins (2014), Khoo and Knobe (2018), and Plunkett and Sundell (2013). We have not spent a great deal of space trying to establish that substantive disagreement is possible even where the contents expressed are not inconsistent, since we take that point to be thoroughly established already in the literature—see the citations above.

answers, and we say that agents *disagree on a question* whenever they have incompatible beliefs regarding which of those answers is correct. (They need not be aware of this incompatibility; disagreement does not imply awareness of disagreement.) On the other hand, a *dispute* (noun) is a kind of linguistic interaction in which interlocutors—the *disputees*—dispute one another. To *dispute* (verb) is to attempt to express what one believes to be a disagreement with another regarding some question. Or, better, a number of related questions. Disagreements usually come in clusters, since how one answers any one question of interest will normally affect how one answers many other questions, and so disputes will often revolve around a cluster of questions. These we say are the questions that the disputees take to be *under dispute.*

Any ongoing dispute will presumably have its causal origins in, and be maintained by, some disagreement *somewhere.* But the true source of a dispute may be something rather different than what the disputees take to be the questions under dispute. This does not automatically render the dispute non-substantive, or merely verbal. Consider:

> **Pro-Life Kano.** Kano is in favor of anti-abortion laws, while Sonya is not. They initially take themselves to be disagreeing primarily about ethical norms relating to the suffering of conscious beings. They later discover they agree entirely about those norms, but disagree about a related question—viz., when a fetus develops consciousness.

The dispute is grounded in a mistake about the nature of the disagreement: Sonya and Kano start with the same conception of which questions are under dispute, but do not, in fact, disagree about any of those questions. So there is something clearly defective in how the dispute is being carried out. However, Sonya and Kano still disagree regarding some substantive matters of close relevance to the questions that they initially took to be under dispute, and that seems to suffice for intuiting that the dispute is more-than-merely verbal. By contrast,

> **Playful Kano.** Kano seems to be asserting that it's ok to torture puppies for fun. Horrified, Sonya tries to convince him that puppies should never be tortured, and certainly not for fun. Eventually, she discovers that on Outworld, "torture" means the same as Earthrealm's "play with."

We take it that this is a paradigmatic example of a verbal dispute. Sonya and Kano's conceptions of the questions under dispute are misaligned, and (moreover) there is no apparent disagreement regarding any of the questions that either of them take to be under dispute nor any particularly substantive questions in the vicinity. The basic source of the dispute is no more than a disagreement about the meanings of some words.[5]

That, at least, is how we are conceiving of merely verbal disputes. Perhaps you disagree with us. But no matter—what is ultimately required for our argument in the next section will be the following two theses:

> **Miscommunication**. A dispute can be substantive even if differences in meaning lead to misunderstandings and miscommunication.
> **Compatibility**. A dispute can be substantive even if the propositions asserted by the disputees are compatible, and indeed even if both propositions happen to be true.

---

[5]We presume that disagreements about the meanings of words usually will not count as "directly relevant" or "in the vicinity" in the appropriate sense. For instance, while your beliefs about the meaning of the phrase "prime minister" will be relevant to how you *express* your beliefs about prime ministers, they usually will not be especially relevant to most of your beliefs about *prime ministers.* Obviously, this does not apply in cases where the disputees themselves take the disagreement to be about the meanings of words (e.g., conscious metalinguistic disagreement about how the words ought to be used).

We support each by means of an example, starting with **Miscommunication**:

> **Evil Kano.** Kano seems to be saying that it's ok to torture conscious beings for fun. Horrified, Sonya discusses the matter further, trying to convince him that conscious beings should not be tortured at all, except perhaps under extreme circumstances—e.g., where the fate of Earthrealm depends on it—and certainly never for fun. Eventually, she discovers that on Outworld, "conscious" means the same as Earthrealm's "is able to speak."

While there is *some* miscommunication due to the differences in meaning, and *some* misalignment in what each disputee takes to be the questions under dispute, they nevertheless *do* disagree regarding the very questions that each takes to be at issue. Kano takes the issue to relate to the permissibility of torturing things that speak for fun (*P* or ¬*P*), and expresses his belief that *P*: *it is permissible to torture anything that speaks for fun.* Sonya takes the question to be about the permissibility of torturing conscious beings for fun (*Q* or ¬*Q*), and expresses her belief that ¬*Q*: *it is not permissible to torture any conscious beings for fun.* *P* entails *R*: *it is permissible to torture any conscious being that speaks for fun.* ¬*Q* entails ¬*R*, and therefore entails ¬*P*. So while there is something defective about how the dispute has been conducted, neither party is grossly mistaken about the nature of their disagreement. As such, the dispute strikes us as very much substantive.[6]

Next up is **Compatibility**:

> **Weatherman Kano.** Sonya and Kano are discussing their upcoming tournament match, and Kano informs her that it'll start the next time it rains. A short time later, Kano sees storm clouds off to the East, and says: "The match will begin soon." Following his gaze, Sonya shakes her head and replies: "The wind is blowing from the West."

Kano believes and thereby asserts *P*, and Sonya replies with ¬*Q*, which is consistent with *P*. Sonya disagrees with Kano about *P* (among other things), and though compatible with what Kano said her assertion expresses that disagreement via implication *relative to* what she reasonably takes him to believe. That is, Sonya believes that Kano will have certain background beliefs *B*—that *the match will begin when it rains*, that *there is a nearby storm to the East*, that *the wind is blowing from the East*, and so on—and she believes (and believes that Kano will believe) that ¬*Q*∧*B* entails ¬*P*. With that as the context, Sonya believes that her assertion will be taken, as intended, as expressing disagreement with the belief that Kano expressed.

So to express her disagreement with *P*, Sonya need not assert ¬*P* nor something that directly entails ¬*P*. In fact, *P* and ¬*Q* might not just be *compatible* but also both *true.* Kano may be mistaken about when the match will begin, or it may be there are also some rainclouds off to the West that neither are aware of. The dispute would still be substantive despite this, because whether it is substantive does not turn on the truth conditions of *what is said*—on whether their sentences happen to express conflicting propositions—but instead on the attitudes of those doing the saying.[7]

5. Round Two: Interdimensional Disagreement

It is time we return to MTE. We are not going to question whether Sonya and Kano associate different meanings—different concepts—with their terms. That they do is a natural implication of

---

analytic naturalism. What is much less clear is whether we should also accept that their terms have incompatible analyses and thereby pick out distinct properties. As others have observed, there is a delicate balance to be struck between **Divergence** and **Similarity** (cf. Levy, 2011; Merli, 2002; Väyrynen, 2018). To the extent that the two folk theories are similar, "obligatory" will play a similar role in M^{earth} as "obligatory" plays in M^{out}—and the more similar those roles are, the more likely it will be that their Lewis-style conceptual analyses will end up picking out a *common best deserver*. So M^{earth} and M^{out} cannot be *too similar* if Sonya and Kano's terms are going to have incompatible analyses. But M^{earth} and M^{out} cannot be *too different* either; if they are, then intuitions may no longer counsel that there truly *is* any real disagreement here that needs accommodating.

It is not *obvious* that this tightrope can be walked. Maybe it cannot be. But these are points that have already been made, so we want to tackle the MTE from a different angle. As such, we are going to assume for the sake of argument that there are some possible cases where **Similarity** and **Divergence** both hold, such that the Lewis-style analysis of "obligatory"^{earth} is incompatible with the Lewis-style analysis of "obligatory"^{out}, and more specifically such that the former but not the latter applies to the act of giving to charity. We want the deck stacked against us here—the idea is to establish that we do not *need* to posit a common best deserver, since we can still explain the intuition of disagreement *even if* there is no common best deserver.

Along the same lines, we also assume that Sonya and Kano are competent speakers of their respective languages, and are not arguing in bad faith. We assume they more or less accept the tenets of their respective folk moral theories, and are otherwise generally similar to others of their kind with respect to how they think about and behave in response to "moral" judgments and assertions. And finally, we assume that neither disputee takes themselves to be engaged in a dispute about the meanings of words or some other metalinguistic matter. Sonya, for instance, does not much care about which meanings get attached to which sounds as uttered by Outworlders' mouths—intuitively, what she *really* cares about is that Kano does not fully recognize the important moral reasons there are for giving to charity, as evidenced by his failure to treat such actions with the appropriate gravitas!

We take it that these assumptions are an intended part of the scenario; we are just making them more explicit. Together, they imply that when Sonya says "Giving to charity is obligatory," she is expressing some content P that she happens to believe, and that is true according to M^{earth}. Likewise, when Kano says "Giving to charity is not obligatory," he is expressing some content $\neg Q$ that he believes, and that is true according to M^{out}. Furthermore, P and $\neg Q$ are not only compatible but both *true*. However, in light of **Miscommunication** and **Compatibility**, we are *not* thereby licensed to infer that the dispute is merely verbal. Quite the opposite: at a scenario where **Similarity** and **Divergence** hold, the analytic naturalist has good reason to suppose that Sonya and Kano's dispute is more likely than not indicative of a substantive moral disagreement. Indeed, the dispute is probably *about* what the disputees take it to be about, more or less, or something in the nearby vicinity. There is likely to be *some* degree of miscommunication and misalignment in what they take to be at issue as a result of the differences in meaning (similar to the **Evil Kano** case); nevertheless, given very reasonable presumptions about their background beliefs, the substantive nature of their moral disagreement will be apparent (similar to the **Weatherman Kano** case).

Here is a quick overview of the reasoning. Though P and $\neg Q$ are both true, it is reasonable to expect that both Sonya and Kano will believe, and believe that the other believes, that what Sonya said (P) implies some content R in the context of Sonya's relevant background beliefs, whereas what Kano said ($\neg Q$) implies some content $\neg R$ in the context of Kano's relevant background beliefs. (Those background beliefs will include beliefs arising directly from their respective folk moral theories, as well as other beliefs that, while not a part of those theories proper, are required to draw particular inferences therefrom—for instance, beliefs about what charities are, about the costs and consequences of giving to them, and so on.) So Sonya and Kano disagree at least on this question of R versus $\neg R$. Worry not what this content R is just yet; we will get to that soon. What matters is that from this disagreement regarding R, we can go on to infer a likely *further* disagreement regarding

some matter more directly relating to the moral import of giving to charity. Or that is what we will argue.

The first step is to say what those implications R and $\neg R$ could be. Given **Divergence**, this is a little tricky. The difficulty arises from the fact that the most obvious implications of what Sonya said (relative to her background beliefs) are *moral* implications, whereas the most obvious implications of what Kano said (relative to his background beliefs) are *tworal* implications. For example, what Sonya said implies:

S = If a person can give to charity and does not, and satisfies the requirements for moral responsibility, that person has done something morally blameworthy.

However, what Kano said implies:

$\neg$T = It's not the case that, if a person can give to charity and does not, and satisfies the requirements for tworal responsibility, that person has done something tworally blameworthy.

You see the problem: we still need to show that S and $\neg T$ are inconsistent. To establish *that*, we would need to find some R such that S implies R and $\neg T$ implies $\neg R$; but of course that is just what we were trying to do in the first place. So, clearly, what we need to find is some *non-moral* (and *non-tworal*) question about which Sonya and Kano can plausibly be said to disagree.

It is here that **Similarity** works in our favor. M^{earth} and M^{out} are similar *not only* with respect to the theory-internal relations they hypothesize to hold between the moral (or tworal) properties they posit, *but also* with respect to how those theories are integrated into their respective societies and broader normative theorizing—how, in other words, their use of moral (or tworal) concepts hook up to the non-moral (and non-tworal) world. And if M^{earth} and M^{out} are similar in those respects, but diverge in how "obligatory" ought to be applied in particular cases, then it is reasonable to expect that there will be *some* relevant non-moral (and non-tworal) questions regarding which Sonya and Kano disagree.

For example, given some very natural beliefs of the sort we could expect most humans to share— for example, that the moral and the pragmatic reflect distinct normative domains yielding distinct sorts of normative reasons for action, and that reasons stemming from moral obligations typically carry presumptively heavier weight than other kinds of reasons, and so on—we can reasonably expect Sonya to believe, for example,

R_1 = There are usually especially weighty non-prudential reasons to give to charity.

R_2 = If one does not give to charity, then one has probably not acted in accordance with the weightiest of their reasons.

R_3 = One typically ought all-things-considered give to charity unless one has strong prudential reasons to do otherwise.

Kano, for his part, will have similar background beliefs, *mutatis mutandis*, inasmuch as is consistent with his acceptance of a similar-but-distinct folk moral theory.[8] So he will believe that the tworal and the pragmatic reflect distinct normative domains yielding distinct sorts of normative reasons for action, and that reasons stemming from tworal obligations typically carry presumptively heavier weight than other kinds of reasons, and so on. But unlike Sonya, Kano is not disposed to give to

---

[8]The example here is designed to make sense under the simplifying assumption that the moral and the pragmatic *exhaust* the domain of normative reasons for action. If the reader thinks there may be other kinds of reasons—esthetic reasons, say—then just replace our "strong prudential reasons" with "strong prudential, or esthetic reasons, or …" and the upshot will be unchanged.

charity, nor to advocate that others must do so. At best, he thinks, giving to charity is twuperer-ogatory. (Supertwerogatory?) So Kano presumably believes $\neg R_1$, and $\neg R_2$, and $\neg R_3$, and so on. Since their background beliefs are similar in the relevant respects, we can reasonably expect that Sonya and Kano will be able to draw these implications out.

That is arguably a non-moral disagreement, but it is not the *extent* of their disagreement—better to think of it as an inroad by which to reach the real source of their dispute. For Sonya does not *just* believe that *giving to charity is morally obligatory*; she believes more generally that the property of *being morally obligatory* plays a certain kind of theoretical role that ties it *inter alia* to the non-moral world. Kano's moral theory is different, and posits roles that, despite defining distinct concepts, nevertheless tie those concepts to the non-moral world in ways that are inconsistent with M^{earth}. After all, if moral reasons are *not* tworal reasons, then they cannot both serve as the weighty non-prudential reasons; and if moral reasons *are* tworal reasons, then at least one of their theories is making false claims about how and when those reasons apply. In the present case, since Kano believes that *tworal* reasons are the especially weighty non-prudential reasons, and that such reasons do *not* attach to the act of charity-giving, he must therefore believe—or believe something that entails—that either giving to charity is not *morally* obligatory, or *if it is*, then that fact does not carry the import that Sonya thinks it does.[9]

Now, if our illustrative example works, it works because those R_1, R_2, R_3, and so on, are *non-moral* and *non-tworal* propositions—neither specifically moral nor tworal concepts are required for one to have thoughts with these contents. So we should say more about this. Are they non-moral and non-tworal propositions in this sense? We argue that they are.

The key concepts required to think the relevant thoughts seem to be just the concepts of PRAGMATIC REASON and ALL-THINGS-CONSIDERED OUGHT, and possessing *these* concepts does not seem to require the possession of any specifically moral or tworal concepts. Consider first the concept of a PRAGMATIC REASON. This concept is not naturally analyzed in terms of moral concepts at all; better, we think, to understand pragmatic reasons instead as being determined by an agent's preferences, intrinsic desires, or personal goals. No reason to tie the analysis of a PRAGMATIC REASON to any specifically moral concepts—it is not difficult to imagine a wholly amoral society of *homines economici*, for example, whose members make frequent use of the concept of a PRAGMATIC REASON despite having no thought for morality or tworality. (The same would be true under a hedonistic analysis of PRAGMATIC REASON.) So *even if* Sonya and Kano do indeed lack one another's moral and tworal concepts, still they might both have thoughts about pragmatic reasons.

And much the same seems to be true for ALL-THINGS-CONSIDERED OUGHT. This can (and should, we think) be analyzed without reference to any specifically moral concepts. For Sonya to say "I all-things-considered ought to give to charity" is for her to say something about the balance of *all* her reasons for and against giving to charity, whatever they may be; but in doing so, she does not say anything about any *specific* types of reasons. To have the concept of the ALL-THINGS-CONSIDERED OUGHT, Sonya needs a concept of a REASON and some idea of how different reasons of different kinds might carry different weights that can be balanced off against one another. None of this seems to presuppose the possession of any specifically moral or tworal concepts. Again, one can imagine that the *homines economici* could have the concept of the ALL-THINGS-CONSIDERED OUGHT, which

---

[9]We have not assumed that Kano possesses any of Sonya's moral concepts, and hence we have not assumed that Kano will be able to express his disagreement with Sonya using concepts they both possess. Kano may or may not be in a position to *fully* understand Sonya's moral assertions—we will talk about that later, where we will argue that they *can* possess one another's concepts, but for the present argument it does not seem to matter either way. Disagreement *qua* conflict in belief implies neither awareness of that disagreement nor the capacity to verbalize the disagreement in some common language. What is required is that Kano believes something that *is* or otherwise *implies* the negation of something Sonya believes, and one proposition can entail the negation of another even if they are expressed using distinct concepts (as with the **Evil Kano** case).

they employ when considering trade-offs between (say) pragmatic and epistemic reasons without ever having moral or tworal reasons get into the mix.[10]

In any case, the point is *not* to rest our argument on this single illustrative example. Connections between moral (and tworal) hypotheses and theses about pragmatic reasons and the all-things-considered ought are just one of *a myriad of* connections between the moral and the non-moral world posited by our folk moral theories. And these connections need not be confined to the prudential either. (That is simply a feature of our example.) Folk moral theories also posit connections between the moral realm and the realms of epistemology, belief, affect, motivation, and behavior. The ascription of certain moral virtues to people, for instance, might yield predictions about how they are likely to behave, or be offered as explanations of why they behaved as they did. *The reader should not mistake us as claiming that the particular illustrative example above is what explains our intuitions in the MTE scenario.*

The explanation is provided instead by the much more general rationale for supposing that *some* relevant disagreement will likely be present given the way the scenario has been characterized. That general rationale goes like this: since M^{earth} and M^{out} must be *similar* with respect to how they link moral (and tworal) properties to the non-moral and non-tworal world, the fact that they also *diverge* with respect to what they deem "obligatory" is a reason enough to think that those theories are likely to have numerous inconsistent implications. OBLIGATORY and TWOBLIGATORY are distinct concepts, to be sure, but **Similarity** plus **Divergence** makes it likely those concepts will be tied to the non-moral (and non-tworal) world in inconsistent ways *given* M^{earth} and M^{out} respectively. Consequently, if *anything* perfectly satisfies the role of "obligatory"^{earth} in M^{earth}, then *nothing* perfectly satisfies the role of "obligatory"^{out} in M^{out}, and vice versa, because those roles place *inconsistent* demands on the *non*-moral and the *non*-tworal world—demands that cannot both be perfectly satisfied. Perhaps nothing perfectly satisfies either role. In any case, at most one of M^{earth} or M^{out} can be true—and, presumably, where they run into conflict will have something to do with what they say about the import of the reasons in favor of giving to charity. Sonya and Kano recognize this, and so they argue.

6. Finish Him!

We should address a natural follow-up concern. Rather than simply analyzing *moral* terms by reference to their roles within a folk *moral* theory, some analytic naturalists may wish to extend the scope of their account to provide an analysis for all kinds of *normative* terms by reference to their roles within a more general folk *normative* theory N^{earth} that includes M^{earth} as a proper part, but encompasses other normative domains as well. The concern, then, is that if M^{earth} is embedded in this broader normative theory N^{earth}, then this will affect the analysis of Sonya's non-moral-but-still-normative concepts as defined by their role therein (such as PRAGMATIC REASON and ALL-THINGS-CONSIDERED OUGHT). The same will apply to M^{out}, embedded now in some broader twormative theory N^{out}. If so, then we can no longer presume that Sonya and Kano will share a base of non-moral-but-still-normative concepts, as we did above.

---

[10]One may be concerned about conceptual holism here, according to which the differences between M^{earth} and M^{out} will "infect" Sonya's and Kano's "total theory" such that they cannot even share a common base of *non*-moral and *non*-tworal concepts from which we might draw some inconsistent implications of the form *R* versus ¬*R*. We have put this concern in a footnote because we have neither the space nor the inclination to go chasing down the holist rabbit hole. So let us quickly note two things instead. First: analytic naturalists are not committed to conceptual holism, and moreover it is not a particularly natural position for them to adopt. The analytic naturalist's strategy crucially requires a distinction between theoretical T-terms and independently understood O-terms—a distinction that does not make a good deal of sense for the holist. Second: the metasemantic challenge is not usually taken to depend on the premise that humans and twin-humans cannot even share their *non*-moral (or *non*-tworal) concepts. As such, if the metasemantic challenge for analytic naturalism really does end up resting on some version of conceptual holism, then our response is straightforward: we reject holism.

It is not obvious to us that the analytic naturalist *should* extend their account in this way but it *is* a very natural move to consider and worth discussing on that basis alone.[11] As far as the *original* MTE scenario is concerned, we note simply that we do not have to establish that Sonya's non-moral-but-still-normative terms will be analyzed the same way as Kano's non-tworal-but-still-twormative terms, since they may nevertheless share a common best deserver. After all, the stipulated similarity of their respective broader normative and twormative theories as required in the *original* MTE scenario presents plenty of reason to expect that those analyses will latch on to the very same extensions.

So for the present challenge we must imagine instead a *revision* of the original scenario—not just the MTE, but instead some new Normative Twin Earth (cf. Eklund, 2017). That is, for the scenario to work we will need the same general properties that were required for the moral case—namely, **Divergence** and **Similarity**—but this time holding with respect to Sonya and Kano's broader normative theories rather than their moral theories. So, for instance, we will need to presume that N^{earth} and N^{out} diverge with respect to the "pragmatic reason"-roles and "all-things-considered ought"-roles in such a way as to generate incompatible analyses and thus divergent extensions for those terms. And we will need to presume that N^{earth} and N^{out} are similar *qua* normative theories. Consequently, in just the same way that M^{earth} and M^{out} needed to be similar with respect to how they hook up to the non-*moral* world, this time N^{earth} and N^{out} will need to be similar with respect to how they hook up to the non-*normative* world.

But if *that is* the case, then the same general rationale as above applies, *mutatis mutandis.* One cannot say that N^{earth} and N^{out} are similar with respect to how they connect their theoretical terms up to the non-normative world, and yet diverge with respect to what they take those terms to apply to, without *ipso facto* making it a priori likely that those two theories will place inconsistent demands on the non-normative world. Given that, we have the same resources for explaining genuine disagreements between those who hold to N^{earth} and those who go along with N^{out}.

7. Test Your Might: Chauvinistic Conceptual Relativism

Let us move on to the second of the two allegations raised—the problem of *chauvinistic conceptual relativism.* For many readers, this will have been lurking in the background of the discussion so far. ("How can Sonya and Kano be genuinely disagreeing with one another, if they are not even capable *in principle* of possessing one another's concepts?") We want to address this allegation directly, since it is an important challenge. If analytic naturalism really were committed to a chauvinistic conceptual relativism, then that would be reason enough to be concerned regardless of what we have to say about the MTE.

So, again, let us again assume for the sake of argument that Sonya and Kano's folk moral theories generate inconsistent Lewis-style analyses for their respective uses of "obligatory." Does it follow that, according to analytic naturalism, the Outworlders of Outworld must lack our moral concepts, and we theirs? Are individuals really trapped inside the conceptual prison of their community's folk morality, forever cut off from understanding those who are not locked up in there with them? Not even a little. There is nothing about analytic naturalism that makes it impossible for Sonya and Kano to understand one another, nor to express and critically evaluate one another's folk moral theories with full comprehension thereof. In fact, there is good reason to expect that Sonya and Kano *could* fully understand one another, *could* possess each other's concepts, and *could* knowingly communicate their differences of opinion on moral (and tworal) matters. Analytic naturalists may be many things, but chauvinistic conceptual relativists they are not.

---

[11]One of the authors would prefer to resist this extension from moral functionalism to a broader normative functionalism, and the other is uncertain. But that is a topic for another paper.

Some brief ground-clearing will be helpful to begin with. First, we understand *concepts* to be "parts of propositions," loosely construed.[12] More importantly, we take it that to *possess a concept* is to have the capacity to entertain (non-trivial) propositions which have that concept as a part. To put the same idea in another way, to possess a concept is to have the capacity to appreciate certain kinds of divisions in logical space. So one (fully) possesses the concept of MONEY, for example, when one understands and can recognize the difference between those actual and hypothetical economic systems that use a conventional medium of exchange versus those that merely involve bartering. It follows that to *lack a concept* is thus to lack the corresponding recognitional and categorisational capacities—to not be in a position to appreciate the difference between those scenarios where the concept in question is instantiated versus those where it is not.

We take it that this is a fairly ordinary conception of concept possession in philosophy. Note that on this conception, one need not explicitly associate a particular word or phrase in one's spoken language with a concept in order to count as possessing that concept. Someone might, for example, have the capacity to reliably differentiate between two subtly distinct flavors of wine, even if they struggle to express that difference in words. And presumably, many monolingual English speakers would have had the concept of SCHADENFREUDE long before it came to be commonly associated with the German loanword. This clarification is important, because the charge of chauvinistic conceptual relativism makes sense only against a background of some strong assumptions regarding the relationship between what an individual's words mean in her own community and which concepts she *lacks*.

Now to be sure, there are some interesting relations between the languages one speaks and the concepts one possesses. For example, the following seems like an eminently plausible connection for any theorist to draw:

Understanding-to-Possession. A person fully understands a word which expresses the concept c only if she possesses c.

We can safely presume that Sonya and Kano are supposed to be competent speakers of their own languages. Sonya therefore possesses the concept OBLIGATORY expressed by "obligatory"$^{\text{earth}}$, while Kano possesses the concept TWOBLIGATORY expressed by "obligatory"$^{\text{out}}$. But this only tells us about the concepts they *have*, and nothing at all about the concepts they *lack*. A further premise is needed if we are going to derive the conclusion that Sonya and Kano *do not*—and indeed *cannot*—possess one another's concepts.

It is clear what the missing premise is supposed to be: if the *only* way to possess the concept expressed by "obligatory"$^{\text{earth}}$ is to belong to a community whose folk moral theory is M$^{\text{earth}}$, then the fact that Kano's theory is not M$^{\text{earth}}$ would imply that he must lack that concept; hence, Kano cannot think OBLIGATORY-thoughts (only TWOBLIGATORY-thoughts). So we need something like:

Possession-to-Theory. Let c be a concept expressed by a term whose analysis can be given in terms of the role it plays in folk moral theory M; then a person possesses c only if they belong to a community whose folk moral theory is M.

This would also imply, by *tollens* on **Understanding-to-Possession**, that Kano is incapable of understanding Sonya's moral assertions; and likewise, by the same connections, that Sonya is incapable of understanding Kano's tworal assertions. So we take it that the conceptual relativism allegation is premised on **Possession-to-Theory**, or something to essentially the same effect.

---

[12]Less loosely, we take concepts to be the kinds of entities (e.g., functions from worlds to extensions) that might serve as the meanings of certain subsentential expressions (e.g., names and predicates), and which compose to determine truth-conditions. We do *not* take "concepts" in this context to designate representational *vehicles*. Analytic naturalism includes no particular commitments regarding the existence, structure, and possession conditions of concepts in this latter vehicular sense.

But analytic naturalists should reject **Possession-to-Theory**, and firmly so! According to analytic naturalism, anyone who understands any folk moral theory and the concepts therein must *ipso facto* possess the resources needed to understand indefinitely many other moral theories as well. According to the analytic naturalist, for Sonya to *fully* understand what her own theory M^{earth} says *just is* for her to know that M^{earth} posits some suite of properties $x, y, z, \ldots$, that are related to one another and to the *non*-moral world in such-and-such a way, and to know which of those properties are supposed to be designated by which terms in her moral vocabulary. Sonya therefore fully possesses the concept OBLIGATORY inasmuch as she understands how "obligatory"^{earth} relates to the other moral terms in M^{earth} (the theory's "T-terms"), how those other terms all relate to one another, and how they all relate to the terms used to describe the non-moral world (the "O-terms"). Notice what also follows: if Sonya has the resources to understand the concepts of M^{earth}, then she must also have the resources needed to understand many other theories too—including at least (but not restricted to) any theory derived by rearranging how the T-terms and/or O-terms are related to one another. As well it should be—for how else could Sonya ever *disagree* with *her own* folk moral theory, if she cannot even entertain other ways that theory might go?

What H&T get right is that analytic naturalism renders our moral terms theory-laden; their mistake is to think this entails any lack of comprehension between those with different theories. Analytic naturalists know better, for they have internalized the lessons of the Ramsey-Carnap-Lewis method of defining theoretical terms—a central point of which was to provide theory-*neutral* analyses of our theory-*laden* terminology!

Indeed, if it turns out, then, that M^{earth} and M^{out} can be characterized in terms of how the properties they posit are related to one another and to the *non-moral* (and *non-tworal*) world, then there is no reason to think that Sonya and Kano *should not* also be capable in principle or in practice of fully understanding one another's theories and possessing one another's concepts. Under that assumption, Sonya will fully understand M^{out}, and possess any of the concepts characterizable therefrom, simply by recognizing that M^{out} posits its own distinct suite of properties $x', y', z', \ldots$, that are related to one another and to the non-moral world in somewhat different ways than her own theory M^{earth} posits. And if she wanted to, Sonya could even fully and accurately express the content of M^{out} in her own language.[13]

So it is simply not true that, according to analytic naturalism, a person can possess the concepts analyzed by reference their role in a folk moral theory M only if M is *their* folk moral theory. If analytic naturalism implied otherwise, then we would have more than enough reason to reject it already even without the metasemantic challenge.

8. Flawless Victory

There is a conception of analytic naturalism according to which, if it were true, then assigning meanings to our moral terms so as to render our folk moral theory true is a more or less trivial exercise. What does "obligatory" mean? Why, just whatever disjunction of actions are deemed obligatory by the theory! The total role of "obligatory" within our moral theory amounts to nothing more than a label we arbitrarily attach to certain actions but not others. Likewise for "good," "right," and so on. So *of course* the theory will turn out true—how could it *not*?

If that were indeed how analytic naturalism works, *then* we would be worried about MTE. On this very simple understanding of the theory, Sonya and Kano are simply applying orthographically similar labels to overlapping but distinct disjunctions of actions. There is no incompatibility in their beliefs about those actions *per se*, nor even in the vicinity. Their dispute boils down to nothing more

---

[13]Again, readers may worry here about conceptual holism. And again, we will point out that holism is neither an implication of analytic naturalism nor a very natural position for analytic naturalists to adopt. If there is a concern arising from conceptual holism, then we reject the holism.

than a disagreement over labels. However—and thankfully—that is *not* how analytic naturalism works. "Obligatory" is more than just a label for an arbitrary disjunction of actions. The theoretical roles of our moral terms stretch out into the non-moral world, with connections to psychology, behavior, and non-moral normative theorizing. And where two moral theories posit *similar* such connections but *diverge* with respect to which actions should be considered "obligatory," we should expect to find conflict. Analytic naturalism gives us every reason to think that intercommunal moral disputes can reflect genuine, more-than-merely-verbal disagreement.

Edward Elliott is an associate professor at the University of Leeds and will be associate professor at the University of Notre Dame as of Fall 2024. Elliott specialises in theoretical philosophy but occasionally joins forces to defend the philosophical truth.

Jessica Isserow is an associate professor at the University of Leeds and will be associate professor at the University of Notre Dame as of Fall 2024. Isserow specialises in moral philosophy but occasionally joins forces to defend the philosophical truth.

References

Björnsson, Gunnar, & Finlay, Stephen (2010). Metaethical contextualism defended. *Ethics*, 121(1), 7–36.

Bolinger, Renee (2022). Metalinguistic negotiations in moral disagreement. *Inquiry*, 65, 352–380.

Boyd, Richard (1988). How to be a moral realist. In Geoff Sayre-McCord (Ed.), *Essays on moral realism*. Ithaca: Cornell University Press.

Chalmers, David (2006). Two-dimensional semantics. In Ernest Lepore & Barry C. Smith (Eds.), *Oxford handbook to the philosophy of language*. Oxford: Oxford University Press.

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

Eklund, Matti (2017). *Choosing normative concepts*. Oxford: Oxford University Press.

Finlay, Stephen (2014). *Confusion of tongues: A theory of normative language.* Oxford University Press.

Finlay, Stephen (2017). Disagreement lost and found. In Russ Shafer-Landau (Ed.), *Oxford studies in metaethics* 12. Oxford: Oxford University Press.

Geirsson, Heimir (2005). Moral twin-earth and semantic moral realism. *Erkenntnis*, 62, 353–378.

Horgan, Terence, & Timmons, Mark (1991). New wave moral realism meets moral twin earth. *Journal of Philosophical Research*, 16, 447–465.

Horgan, Terence, & Timmons, Mark (1992). Troubles on moral twin earth. *Synthese*, 92, 221–260.

Horgan, Terence, & Timmons, Mark (2000). Copping out on moral twin earth. *Synthese*, 124, 139–152.

Horgan, Terence, & Timmons, Mark (2009). Moral functionalism meets moral twin earth. In Ian Ravenscroft (Ed.), *Minds, ethics, and Conditionals* (pp. 221–236). Oxford: Oxford University Press.

Jackson, Frank (1998). *From metaphysics to ethics*. Oxford University Press.

Jenkins, Carrie (2014). Merely verbal disputes. *Erkenntnis*, 79, 11–30.

Khoo, Justin, & Knobe, Joshua (2018). Moral disagreement and moral semantics. *Noûs*, 52, 109–143.

Levy, Neil (2011). Moore on twin earth. *Erkenntnis*, 75, 137–146.

Lewis, David (1970). How to define theoretical terms. *Journal of Philosophy*, 67, 427–446.

Merli, David (2002). Return to moral twin earth. *Canadian Journal of Philosophy*, 32, 207–240.

Plunkett, David, & Sundell, Timothy (2013). Disagreement and the semantics of normative and evaluative terms. *Philosopher's Imprint*, 13, 1–37.

Ramsey, Frank P. (1931). Theories. In R. B. Braithwaite (Ed.), *The foundations of mathematics*. Routledge.

Silk, Alex (2017). Normative language in context. In Russ Shafer-Landau (Ed.), *Oxford studies in metaethics 12*. Oxford: Oxford University Press.

Väyrynen, Pekka (2018). A simple escape from moral twin earth. *Thought*, 7, 109–118.