Moral Kombat
Analytic Naturalism and Moral Disagreement

Edward Elliott & Jessica Isserow

Abstract

Moral naturalists are often said to have trouble making sense of inter-communal moral disagreements. The most familiar incarnation of this challenge is the Moral Twin Earth argument. The culprit is typically thought to be the naturalist's metasemantics and its implications for shared meanings and concepts across communities. We address this challenge from the perspective of analytic naturalism. We show that (contrary to a popular opinion) analytic naturalism does not trap language users within conceptual prisons that cordon off all possibility of inter-communal moral communication and disagreement. Moreover, and drawing on more general insights we develop concerning the relationship between meaning and disagreement—or better: the lack thereof—we can see that analytic naturalists are in a position to accommodate intuitions concerning the possibility of genuine inter-communal moral disagreements.

Keywords: Analytic naturalism; Moral functionalism; Moral Twin Earth; Metasemantic Challenge; Moral disagreement; Concept possession

§1. Introduction

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

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

While disputes like these can sometimes feel intractable, we tend to assume they reflect genuine disagreements. On the face of it, it's unlikely that Sonya and Johnny are merely exchanging noises with only the superficial appearance of communication and conflict. If she's right, then he's wrong.

So we take it for granted that genuine moral disagreements are possible. And if you're a moral realist, then it's easy to see how to make sense of that possibility in cases like *Sexist Johnny*. Both parties (we can safely assume) are competent speakers of the same language, they mean the same things by their terms, and whereas Sonya asserts some proposition P, Johnny flatly contradicts it by asserting $\neg P$. There's no miscommunication, the dispute isn't merely verbal, just a straightforward conflict in the beliefs expressed by their respective assertions.

But *Sexist Johnny* is most naturally read as an *intra*-communal disagreement, and the moral realist should presumably want to be able to make sense of *inter*-communal disagreements as well. Suppose now that Johnny had spent his life on Outworld, and has come to visit. People on Outworld tend to think somewhat differently about the moral acceptability of gender inequality. No doubt this raises the probability that Johnny and Sonya have simply been talking past one another, but not by much: even with these further details added in, the disagreement is still apt to strike us as *a genuine moral disagreement*. These seem possible both within and across communities—even across communities with differing views on what's morally acceptable. (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 judgements are, for the naturalist, purportedly 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 a kind of relativism of meaning to one's local (natural or linguistic) environment. For instance, in just the same sort of 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 grock onto (e.g., H_2O versus XYZ), so too one might think the meaning of 'wrong' depends upon how it's used and which natural properties happen to be around. If two communities use their moral terms differently, or if the local natural properties are somewhat different, then there's a difference of meaning.

Having established that relativism of meaning, it's then easy to imagine two otherwise similar communities that differ just with respect to whatever local metasemantic facts ground the meanings of their moral vocabularies. Since the local grounding facts differ, the meanings of their terms differ—and so, the argument goes, any apparent moral disagreement across members of those communities will be just that: *apparent*. Thus when Sonya says 'Policies promoting gender inequality are wrong', then she's asserting one thing *P*, whereas when Johnny from Outworld says 'Policies promoting gender inequality are not wrong', he's asserting $\neg Q$, and they're not really disagreeing. Cue *reductio*, reject naturalism, *QED*.

This style of argument has been popularised especially by Horgan and Timmons (2009; henceforth 'H&T'), whom we take as our primary foil. They describe a hypothetical *Moral Twin Earth* (MTE) populated by individuals with whom we purportedly cannot have any genuine moral disagreements should naturalism prove true. If the MTE argument proves successful, then there's a metasemantic fly in the naturalist's metaethical ointment. But the argument is unsuccessful. The problem, we'll argue, concerns the connection between meaning and disagreement—essentially, there isn't any. The possibility of genuine moral disagreement requires neither sameness of meaning in moral vocabulary, nor even a conflict in the propositions expressed; indeed, it doesn't directly have much anything to do with *what's said* at all, and everything to do with the *attitudes* of those doing the saying.[1]

[1] Plunkett and Sundell (2013) also deny that accounting for genuine moral disagreement requires satisfying what they call the 'Shared Meaning Task' (cf. Khoo and Knobe 2018, on the 'Exclusion Inference'). Unlike Plunkett and Sundell, we won't rely on the notion of metalinguistic negotiations to build our case—the disagreements we're interested in don't concern how a term should be used.

Most discussions of the metasemantic challenge centre upon *synthetic* moral naturalism. Ours won't. We'll approach MTE scenarios from the perspective of the analytic naturalist. It's unclear why there's been so few attempts to do so. Perhaps this can 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's not the sort of thing many will want to find skulking around their office. But every paper needs a starting point, and we won't waste space defending our own. We expect that some aspects of our response can likely be repurposed by synthetic naturalists as well, but we leave readers to draw such conclusions for themselves

We'll begin with some stage setting: §2 covers everything you need to know about analytic naturalism, while §3 goes into further detail on the MTE challenge. We then present our response in §§4–6. By the close of the paper, we hope to have shown that analytic naturalists can accommodate genuine moral disagreement across morally very different communities.

§2. Background: 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—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 found its fullest expression with Lewis (1970), we can therefore *analyse* the meaning of our term 'belief', for instance, by reference to its role within that theory. Thus, according to the analytic functionalist, it's a priori that 'belief' picks out whatever it is that plays the role that 'belief' is supposed to play within folk psychology, if anything does—or, if nothing does, then whatever comes closest to playing that role, provided it plays that role well enough.

As with the mind, so too with morality. The concepts of RIGHT, WRONG, GOOD, BAD, OBLIGATORY, 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 moral theory*. Hence the analytic naturalist proposes to analyse the terms of our moral vocabulary in terms of their respective roles within folk moral theory. Analytic naturalists will, for instance, want to say that 'right' can be analysed in terms of the best deserver of the rightness-role in our folk moral theory. Note that there are three claims being bundled together here:

1. *Metasemantic*: the meaning of 'is right' depends (in the style of Lewis 1970) on our local folk moral theory.
2. *Semantic*: 'is right' means the same thing as 'is the best deserver of the rightness-role in our folk moral theory'.
3. *Epistemic*: the foregoing semantic equivalence is knowable apriori.

If we were being more careful, we would precisify these claims in terms of *primary intensions*, as those are understood within the two-dimensionalist framework (see especially Jackson 1998 and Chalmers 2006). But such details needn't concern us, as two-dimensionalism won't play an important part in the ensuing discussion.

See Finlay (2017) for critical discussion of this 'metalinguistic strategy'. Geirsson (2005) also denies that disagreement requires interlocutors to assert incompatible propositions. However, Geirsson appeals to *speaker meanings*, and again our strategy is distinct: we don't take genuine disagreement to depend upon sameness of meaning in *any* sense.

We should also say more about our key terms, starting with 'the folk'. There's plenty of scope for variation here: perhaps 'the folk' are all human beings, provided we typically share more or less the same moral opinions, practices, and intuitions. Or perhaps there are distinct folk moral theories for different populations. There may even be one folk moral theory that covers one population, and a slightly different theory for some other overlapping population, with some indeterminacy of meaning in moral vocabulary thereby resulting for the members of those populations. But these are matters for debate *within* analytic naturalism, and shouldn't matter for our discussion. So in the interests of simplicity, we'll simply take 'the folk' to be all modern human beings.

There is likewise plenty of scope for variation in how we 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 and may in principle 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 well involve complicated ideas and principles that needn't be apparent even to those who regularly use and adhere to them. (cf. Jackson & Pettit 1995 on 'ethocentric' belief). Or we might think of folk moral theory instead as an abstraction of the principles that best explain the way the folk are disposed to act and think about morality *on average*, without at any point supposing that the complete suite of these principles constitute the contents of any individual's beliefs (tacit, ethocentric, or otherwise). Indeed, it may not even be the folk moral theory properly so-called that we're interested in, but instead a future version of the theory that's been tidied up and systematised, with the inconsistencies smoothed over and the gaps filled in—a 'mature folk morality', as Jackson (1998) puts it. Again, these are matters of debate within analytic naturalism, and shouldn't much impact our discussion. What does matter is that there's something we'll call a *folk moral theory* that's closely tied to the folk's moral opinions, practices, and intuitions, and that our moral vocabulary is analysed by reference to that theory in a manner akin to how analytic functionalists analyse folk-psychological terms by their roles within folk psychology.

Finally, we should emphasise that it's no requirement of analytic naturalism that every claim or principle within folk moral theory must end up being correct. Analytic naturalists aren't proposing to reduce moral theorising to the mere cataloguing of folk moral theories! According to the analytic naturalist, 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, *perfectly* occupy their respective roles within that theory (see Lewis 1970). Consequently, if even one of these properties *imperfectly* occupy their respective roles, even to a slight degree, then our folk moral theory must be at least somewhat mistaken.

And it very well may be mistaken. In much the same way that folk physics conflates the concepts of WEIGHT and MASS, folk morality may say that there's one moral property that plays two roles when really there are two properties that merely appear identical to the untrained eye. Or perhaps folk morality says that certain kinds of actions have the property of *being obligatory*, when in fact they end up being merely permissible or even plain wrong. One shouldn't make the mistake of supposing that, according to analytic naturalism, a term like 'obligatory' simply ends up referring to whatever disjunction of actions are deemed obligatory by our folk moral theory. There's much more to the theoretical role of 'obligatory' in that theory than just which actions it happens to get attached to, so it needn't be the case that every aspect of that role ends up being satisfiable. This will be important.

For analytic naturalists, then, the functional analyses of the terms of our moral vocabulary will secure a referent whenever our folk moral theory is at least *mostly* correct—when the roles of the terms within the theory are *mostly* satisfied. It's therefore possible to disagree with some aspect of the folk moral theory even while accepting that our moral terms are properly analysed by reference to their role within that theory. Indeed, it's not unlikely that we *should* sometimes disagree with what our folk moral theories have to say, and the moral judgements they predict. Doing so does not entail that we've somehow lost our capacity to think moral thoughts or to meaningfully converse with those who might have differing moral opinions.

§3. The Metasemantic Challenge

Now on to the metasemantic challenge. We'll begin by setting up the context. As we've noted, analytic naturalists take the meaning of a term like 'wrong' for a population to depend upon the character of that population's folk moral theory. According to H&T, this reveals that the analytic naturalist is not in fact a realist after all—she's a relativist dressed in realist garb. Specifically, H&T find analytic naturalists guilty of what they call *chauvinistic conceptual relativism*. The position is said to entail that, 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 wouldn't possess *moral* concepts at all. Analytic naturalism therefore 'chauvinistically builds the folk morality supposedly shared by all of humankind directly into moral concepts themselves', and because of this it's taken to be 'objectionably human-centred' (H&T 2009, 228).

Why is this (purported) conceptual relativism 'objectionable'? Essentially, H&T argue, because it conflicts with 'strong intuitions' (2009, 223-8) about disagreement in their MTE scenario. H&T's hypothetical world is as close to our world as possible consistent with the following stipulation: whereas we humans have our folk moral theory, T^E, the twin-humans (twumans) of Twin-Earth have some distinct theory, T^{TW}. By hypothesis, these theories are supposed to be overall similar to one another—not only with respect to which actions and people and institutions and so on that they imply will be 'right' and 'good' and so on, but also with respect to how the theories are integrated into their respective societies and broader normative theorising. Twumans' 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 Twin-Earth; [twumans] 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… (H&T 2009, 225)

The similarity between T^E and T^{TW} is crucial. Imagine, by contrast, that twumans were only ever disposed to classify an action as 'right' when it involved significant cardiovascular activity, and believed that the regular performance of 'good' actions tends to promote fat loss, muscle gain and increase life expectancy, and so on in like fashion. In this case there would be no temptation to say that T^{TW} is in any sense a *moral* theory—the twumans would simply be using words orthographically similar to those of our moral vocabulary to talk about a very different domain of phenomena.

So it's clearly important that T^E and T^{TW} are overall similar to one another. We might imagine, perhaps, that T^E is essentially consequentialist in character, while T^{TW} is essentially deontological but overlaps a good deal with T^E. This is how H&T themselves describe the MTE scenario, but we shouldn't get bogged down in such details—what matters most at this point is that T^E and T^{TW} are (a) similar with respect to how they're taken to relate to other things like decision-making, social policy and well-being, and (b) differ somewhat with respect to some of the actions, people and/or institutions that they end up saying are 'right' and 'good' and so on. Against this backdrop, H&T's challenge concerns whether humans and twumans are capable of having genuine, more-than-merely-verbal moral disputes. Imagine that some human (Sonya), were to encounter some twuman (Kano), perhaps at some interplanetary ethics tournament. We can suppose that 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 T^E. Kano retorts with 'Giving to charity is not obligatory', which is true according to T^{TW}.

Is this a genuine and substantive dispute, or is it merely verbal? It certainly seems like it *could* be genuine, and that it more likely than not is. At least, that's how most people's intuitions about these cases typically go (including our own). So we should very much like to be able to explain this intuition—and better, all else equal, if we can do so without having to diagnose it as some kind of widespread mistake. However, H&T seem to be of the opinion that only one style of explanation is up to the task—one which, they contend, is unavailable to analytic naturalists. In particular, H&T seem to think that in order to do justice to the intuition that this dispute is more-than-merely-verbal, we need to say that Sonya and Kano *mean the same thing* by their terms.[2] The correct diagnosis of the intuition, in their 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 twumans] would be genuine disagreements—i.e., disagreements in moral belief and in normative moral theory, rather than differences in meaning. (227)

The analytic naturalist, however, has no recourse to this explanation:

> … 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 T^E, whereas the twin-moral terms used by [twumans] designate *distinct* unique natural properties that respectively satisfy the respective conceptual analyses obtainable from T^{TW}; 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. (226-7)

T^E and T^{TW} must be similar, we said, but not *too* similar. The role of 'obligatory' in T^E must diverge enough from its role in T^{TW} that Sonya's and Kano's respective uses of the

[2] We take this interpretation to reflect how the challenge is standardly presented in the literature. Geirsson attributes to H&T the claim that 'If [moral and twin-moral terms] differ in meaning, then [humans and twumans] could not have genuine moral disagreements' (2005, 358). Rubin likewise takes H&T to assume that 'If ['right' on Twin-Earth] expresses a different meaning from ['right' on Earth] and the two predicates are not intertranslatable, then the apparent moral disagreement between [humans and twumans] is merely apparent' (2014, 290). Plunkett & Sundell read H&T similarly: '… in order to explain the possibility of genuine disagreement between [humans and twumans], one needs to understand them as meaning and referring to the same things by their moral terms' (2013, 19-20).

term have ‘incompatible’ Lewis-style analyses.[3] It’s that *incompatibility* in their analyses that’s supposed to ensure a difference in *extension*, and the difference in extension that’s supposed to explain why their assertions aren’t in conflict. Essentially: what Sonya says is true iff giving to charity has some property *p*, while what Kano says is true iff giving to charity lacks some distinct property *q*, where being *q* neither implies nor is implied by being *p*, and hence there’s no inconsistency in the two of them jointly saying that an action both has *p* and lacks *q*.

So Sonya’s term—label it ‘obligatory’^{E}—must designate a distinct property than the orthographically similar term, ‘obligatory’^{TW}, as used by Kano, and this is supposed to imply that the two are engaged in a merely verbal dispute:

> … the different parties are expressing different concepts with their moral terms, are talking past one another rather than disagreeing. (232)

But, recall, the naturalist’s problems don’t end there! If analytic naturalism is correct, H&T add, then Sonya and Kano cannot even possess one another’s concepts:

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

In summary: the core allegation is that analytic naturalism cannot accommodate the intuition that Sonya and Kano have a genuine moral disagreement, due to a difference in the meanings of their terms and the kinds of concepts each is capable of possessing. We humans are conceptually cut off from understanding the twumans’ tworal assertions, and they from our moral assertions, precisely because of the differences between our folk moral theories.

§4 Chauvinistic Conceptual Relativism?

Ultimately, we will argue that metasemantic challenge falters primarily in virtue of tying meaning and disagreement too closely together. But before that, it’ll be helpful in this section to get clear on precisely what the analytic naturalist is and (moreover) is *not* committed to saying about the MTE scenario. Analytic naturalism carries no commitment to many of the claims about *meaning*, *concepts* and *communication* that H&T burden it with—and appreciating why will be useful for understanding the line of argument that we develop below.

Let’s start by noting that we are happy to grant that Sonya and Kano will associate different (primary) intensions with the terms of their respective moral and tworal vocabularies. This is indeed an implication of analytic naturalism, and in this sense Sonya and Kano *mean* different things by ‘obligatory’. But it’s less clear whether we should also accept that they must be *referring* to different properties. Recall that it’s crucial for T^{E} and T^{TW} to be similar to one another. (If they’re too different, then intuitions of genuine disagreement are less likely to arise.) But to say that T^{E} and T^{TW} are similar *just is* to say

[3] We assume that two predicates have *incompatible* analyses just in case those analyses are such that, at every world, there’s things in the extension of each one that aren’t in the extension of the other. So neither predicate’s extension subsumes the other’s, but they can overlap. H&T never explain what they mean by ‘incompatible’, but this is the definition that makes the best sense. ‘Incompatible’ in this context cannot mean that the extensions never overlap at any world. By hypothesis, humans and twumans call *some* of the same things ‘good’, ‘right’, ‘obligatory’, and so on, so presumably there must be some overlap. Nor can ‘incompatible’ merely mean that the predicates *might* have different extensions at *some* worlds—for then how could we justify the confidence that Sonya and Kano’s respective uses of ‘obligatory’ *will* as a matter of fact have divergent extensions?

that the terms within those theories play similar theoretical roles. And if ‘obligatory’^{E} plays a similar role in T^{E} as ‘obligatory’^{TW} does in T^{TW}, then we might expect that the Lewis-style conceptual analyses of ‘obligatory’^{E} and ‘obligatory’^{TW} could share a *best deserver*—and hence wouldn’t be incompatible after all! So it remains open for us to legitimately question whether the MTE set-up really makes sense, or at least whether intuitions about the case are being muddied by the precariously thin line that H&T are trying to walk here. (For similar points, see Merli 2002, Levi 2011, and Väyrynen 2018.)

But we won’t dwell on this issue further; we have bigger fish to fry. In particular, we want to thoroughly deny the charge of chauvinistic conceptual relativism. It’s simply false that twumans must lack our moral concepts, and we theirs. More generally, there’s nothing about analytic naturalism that makes it *impossible* for humans and twumans to fully understand one another, or to express and critically evaluate one another’s folk moral theories.

First some ground-clearing. We understand *concepts* to be ‘parts of propositions’, loosely construed.[4] 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 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 recognise the difference between those actual and hypothetical economic systems that use a conventional medium of exchange versus those that merely involve bartering. To *lack a concept*, then, is 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 isn’t.

Note that on this conception of concept possession, one needn’t explicitly associate a particular word or specific phrase in one’s spoken language with a concept in order to possess that concept. Someone might, for example, have the capacity to reliably differentiate between two subtly distinct flavours of wine, even if they struggle to express that difference in words. And we presume that many English speakers would have had the concept of SCHADENFREUDE long before it came to be associated with the German loanword. This 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, and which concepts she *lacks*.

Now to be sure, there are some close links between the language one speaks and the concepts one possesses. For example, the following seems eminently plausible:

> A speaker fully understands a sentence which expresses the proposition *P* only if she’s capable of entertaining *P*. Likewise, a speaker fully understands a word which expresses the concept C only if she possesses C.

We can presume that Sonya and Kano are competent speakers of their respective languages. So Sonya possesses the concept OBLIGATORY expressed by ‘obligatory’^{E} while Kano possesses the concept TWOBLIGATORY expressed by ‘obligatory’^{TW}. So far so good—but this only tells us about the concepts they *have*, and nothing of what concepts

[4] 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 emphatically do *not* take ‘concepts’ in this context to designate representational *vehicles*, such as words in a language of thought or the mental representations that are sometimes hypothesised to explain our recognitional and categorisational capacities.

they *lack*. A further premise is needed if we're going to derive the conclusion that Sonya and Kano *do not* and *cannot* possess one another's concepts.

It's clear what the missing premise is supposed to be: if the *only* way to possess the OBLIGATORY concept is to belong to a linguistic community whose folk moral theory is *our* theory, then the fact that Kano's theory isn't ours would imply that he must lack that concept. This would also imply, by *tollens* on the plausible link above, that Kano is incapable of really understanding Sonya's 'obligatory'-sentences. So something like this premise is evidently at the foundation of H&T's objection—that one can possess the concepts characterised by their role in a folk moral theory T *only if* T is *their* folk moral theory. Individuals are, in effect, trapped inside the conceptual prison of their community's folk theory, forever cut off from understanding the concepts of those who aren't locked up in there with them.

But analytic naturalism is committed to no such premise. Quite the opposite, in fact: analytic naturalism implies that anyone who fully understands our moral concepts must *ipso facto* possesses the resources needed to understand innumerably many other folk moral theories and the concepts defined therefrom. What H&T get right is that analytic naturalism renders our moral terms thoroughly theory-laden; their mistake is to think that this implies a lack of understanding between those who hold different theories. Analytic naturalists know better, for they've internalised the lessons of the Lewisean method of defining theoretical terms—the point of which was to provide theory-*neutral* analyses of our theory-*laden* terminology.

According to the analytic naturalist, for Sonya to fully understand what her own theory T^E says *just is* for her to know that T^E posits some suite of properties *x, y, z*, …, 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. She fully possesses the concept OBLIGATORY, then, inasmuch as she understands how 'obligatory'^{E} relates to the other moral terms in T^E (the theory's 'T-terms'), how they all relate to one another, and how they relate to the terms used to describe the non-moral world (the 'O-terms'). But notice what follows: if Sonya has the conceptual resources to understand T^E, she *also* has the resources needed to understand many other theories too—including at least any theory gotten 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 folk moral theory, if she cannot even imagine other ways a moral theory might go?

More importantly, if it turns out that T^E and T^{TW} can be characterised in terms of how the properties they posit are related to one another and to the non-moral and non-tworal world, then assuming that Sonya and Kano can at least share their non-moral and non-tworal concepts, there's no reason to think they shouldn't also be capable of fully understanding one another's moral theories. In this case, Sonya can fully understand T^{TW}, and possess any of the concepts characterizable therefrom, simply by recognising that T^{TW} posits its own distinct suite of properties *x*′, *y*′, *z*′, …, that are related to one another and to the non-moral world in somewhat different ways than her own theory T^E is. Of course, Sonya will associate different meanings with her terms than Kano does with his, but she is nevertheless capable of understanding Kano's alternative theory and of possessing Kano's TWOBLIGATORY concept. Indeed, she's even capable of fully and accurately expressing the content of T^{TW} in her own language.

(Some readers may worry here about *conceptual holism*: perhaps the differences between T^E and T^{TW} will 'infect' Sonya's and Kano's other concepts, such that they cannot share a common base of non-moral/non-tworal O-terms. But we mention this possibility only to set it aside; we've neither the space nor the inclination to go chasing down the holist's rabbit hole. Let us briefly note two things instead. First: analytic naturalists are by no means committed to conceptual holism. Furthermore: the metasemantic challenge isn't usually taken to depend on the premise that humans and twumans cannot even share *non*-moral/*non*-tworal concepts. If it really does rest on that premise, then we take that to be a significant limitation for the challenge, and our response is straightforward: we reject that premise.)

§5. Disagreement and Disputes

We've shown that it's consistent with analytic naturalism that Sonya and Kano are capable of fully understanding one another's assertions; neither need be trapped in the conceptual prison of their local linguistic community. But our dialectical task is not yet complete: we still need to explain the *specific* intuition that Sonya and Kano are (more likely than not) engaged in a dispute that's more-than-merely-verbal. That will be our goal for the remainder of the paper. The point of this section in particular is to get clear on what makes a dispute feel intuitively *real* or *genuine*, or at least *more-than-merely-verbal*.[5] The hypothesis is that the presence or lack of specific kinds of *defectiveness* in a dispute is what explains these intuitions of genuineness. But before we get to that, it'll be helpful to begin with some key distinctions and a bit of stipulated terminology.

First, we want to distinguish between the mental phenomena that we'll refer to as *disagreements*, and the linguistic phenomena that we'll call *disputes*. To get a feel for the distinction, consider:

> *Academic Johnny*. While Johnny presents his latest research, Sonya has a complicated thought: she gets the sense that she disagrees with his thesis, but struggles to verbalise exactly why. Sonya attempts to explain her disagreement—and fails. She means every word she says, but what she says doesn't quite capture her thoughts. Following some discussion, what she said is shown to be consistent with Johnny's thesis. Still, Sonya has the feeling that had she expressed her ideas better, an inconsistency would have been apparent.

Here's what Sonya *doesn't* say in this case: "Well what I *said* isn't inconsistent with Johnny's thesis, so I guess we don't disagree after all." After all, the *disagreement* and Sonya'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 or more subjects regarding some question. A *question* we take to be just a way of partitioning logical space into a set of possible answers, and we say that subjects *disagree on a question* whenever they have incompatible beliefs regarding which of those answers is correct. They needn't 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 two or more 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, perhaps more often, on some group of related questions. Disagreements usually come in clusters, since how one answers one question will typically affect how one answers many other related questions, and so disputes will likewise often relate to clusters of questions. These are the questions the disputees take to be *under dispute*.

[5] Parts of the following discussion overlap with points made by Geirsson (2005), Chalmers (2011), Plunkett & Sundell (2013), and Jenkins (2014).

The *Sexist Johnny* case (from §1) has all the hallmarks of what might be thought of as an *ideal* dispute—it's most naturally read such that the following are satisfied:

Incompatible Assertions. The propositions expressed by the disputees are incompatible with one another; at most one can be true.

Perfect Understanding. The disputees fully understand one another's expressions; there is no miscommunication between them.

Perfect Alignment. The disputees have the very same conception of what questions are under dispute.

Identified Disagreement. The disputees do disagree on all, or at least many of, the questions they jointly take to be under dispute.

However, we should resist the temptation to think that every instance of an intuitively real or meaningful dispute must have this structure. Indeed, *none* of the above properties are necessary for a dispute to be more-than-merely-verbal.

Start with **Incompatible Assertions**, and consider:

> *Meteorological Johnny.* Johnny looks out the window and sees some dark clouds off to the East. He says: 'It will rain soon'. Sonya shakes her head and responds: 'No, the wind isn't blowing from the East'. They both believe that it will rain just in case the wind is blowing from the East.

The propositions expressed aren't incompatible: rain often follows a Westerly wind. Indeed, Sonya and Johnny might each said something *true*—they may just be wrong to believe that it'll rain only if wind is blowing from the East. But the dispute feels perfectly genuine, and there's no sense at all that there's something defective about how the dispute has been conducted. Sonya and Johnny disagree about whether it will rain, Sonya recognises this and communicates said disagreement via implication relative to Johnny's background beliefs (which need not be accurate). So the **Incompatible Assertions** condition needn't be satisfied for a dispute to seem real.

We can kill both **Perfect Understanding** and **Perfect Alignment** with one stone:

> *Evil Johnny.* Johnny 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 shouldn't 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 where Johnny is from, 'conscious' just means *is able to speak*.

Sonya and Johnny mean very different things by their respective uses of 'conscious'. And while Sonya recognises that not everything that speaks is conscious, and not every conscious thing can speak, still the dispute strikes us as very much meaningful. Why? Because while there's some miscommunication, and some misalignment in what each disputee takes to be the question under dispute, and therefore something slightly defective about how the dispute was conducted, they nevertheless *do* disagree on those questions. Johnny takes the issue to relate to the permissibility of torturing things that speak for fun (P or $\neg P$), and expresses his belief that P: it's permissible to torture anything that speaks for fun. P implies R: it's permissible to torture any conscious being that speaks for fun. Sonya takes the question to be about the permissibility of torturing conscious beings for fun (Q or $\neg Q$), and expresses her belief that $\neg Q$: it's not permissible to torture any conscious beings for fun. $\neg Q$ implies $\neg R$, and therefore implies $\neg P$. Neither party is fundamentally mistaken about the nature of their disagreement. That seems to suffice to make the dispute feel *real*. So **Perfect Understanding** and **Perfect Alignment** aren't necessary either.

Nor, finally, do we need **Identified Disagreement**. Consider:

> *Pro-Life Johnny*. Johnny is in favour of anti-abortion laws, while Sonya isn't. They initially take themselves to be disagreeing primarily about ethical norms relating to the suffering of conscious beings. They eventually discover they agree entirely about those norms but disagree about a related empirical question—*viz.*, when a foetus develops consciousness.

This dispute is grounded in a more severe mistake about the true source of the dispute than the *Evil Johnny* case. But it's still apt to strike us as at least *more-than-merely-verbal*, given that Sonya and Johnny did at least disagree about some substantive matters of direct relevance to the questions they initially took to be under dispute. By way of contrast, consider a paradigmatically verbal dispute:

> *Playful Johnny*. Johnny seems to be saying 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 where Johnny is from, 'torture' just means *play with*.

There's something very seriously defective about this dispute. For one, Sonya's and Johnny's conceptions of the questions under dispute are totally misaligned. Moreover, they (presumably) don't disagree about any of the questions either of them take to be under dispute, nor even on any particularly substantive questions in the vicinity. They do disagree about *something*—of course they do. Any ongoing dispute must have its causal origins in, and be maintained by, some conflict in disputees beliefs *somewhere*. But the true source of their dispute turns out to be something quite distant from what either Sonya or Johnny took it to be, and of comparatively little import.[6]

To summarise: a dispute can be wholly non-defective, and feel perfectly genuine, even when the disputees express compatible contents; indeed, even when they each say something *true*. Disputes that involve mere miscommunication and/or misalignment have something defective about them, but they can still feel very much real. Where the disputees are mistaken about what questions they truly disagree upon are apt to strike us as defective in a deeper way. But such mistakes come in degrees, and a dispute will seem at least more-than-merely-verbal whenever there's some (non-metalinguistic) disagreement in the nearby vicinity. The seriously defective cases arise when neither disputee takes themselves to be having a disagreement about the meanings of words, but it is precisely such a disagreement that is the ultimate source of their dispute.

With all that in hand, in the next section we'll argue that (given the way the MTE scenario is described), it's reasonable to think that Sonya and Kano's dispute is *at least* more-than-merely-verbal. In particular, they (probably) disagree *at least* on the matter of which if either of their moral theories is correct. They may or may not be miscommunicating slightly, and they may or may not have slightly misaligned conceptions of exactly which questions are under dispute; however, they each *rightly* take themselves to disagree about the truth of their respective moral theories, or at least the disagree on

[6] In saying this, we're supposing that where the apparent topic of the dispute isn't itself about the meanings of words, then beliefs about the meanings of words won't usually count as 'directly relevant' 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 won't be especially relevant to most of your beliefs about *prime ministers*.

some substantive issue in the nearby vicinity. Either way, their disagreement is not merely about the meanings of words.

§6. Conflicts across worlds

Let's start with some assumptions. First: we can reasonably suppose that Sonya and Kano are fully competent speakers of their respective languages, and they're not arguing in bad faith. Second: we can also suppose—though only for the sake of argument—that the Lewis-style analysis of 'obligatory'^{E} is *incompatible* with the analysis of 'obligatory'^{TW}, and specifically such that the former but not the latter applies to the act of giving to charity. Finally, we assume that neither Sonya nor Kano take themselves to be having a dispute about the meanings of words. This includes 'metalinguistic disputes' about the *best* way to use the word 'obligatory' (*pace* Plunkett & Sundell 2013). Sonya, for her part, doesn't especially care about which meanings the twumans attach to which sounds—she cares that they've got the wrong moral theory, as evidenced by their failure to treat the act of giving to charity with the appropriate gravitas!

We've loaded the dice very much in our opponent's favour with these assumptions. Along with the description of the MTE scenario, together they imply:

1. When Sonya says 'Giving to charity is obligatory', she's expressing some content *P* that she believes, and is true according to T^{E}.
2. When Kano says 'Giving to charity is not obligatory', he's expressing some content $\neg Q$ that he believes, and is true according to T^{TW}.
3. *P* and $\neg Q$ are both true, and therefore jointly consistent.

But Sonya and Kano's dispute can still be more-than-merely-verbal. Indeed, they can be more or less aligned in their conceptions of which questions are under dispute, and they can both be *correct* in believing that they disagree about those very questions. For Sonya and Kano might each *rightly* believe (and believe that the other believes) that what Sonya said implies some further and closely related content *R relative to* Sonya's background beliefs, while what Kano said implies $\neg R$ *relative to* Kano's background beliefs, and thus they disagree at least on the question of *R* versus $\neg R$. The conflict, in other words, can be implied by what they each said *along with* other things they respectively believe by virtue of their differing moral theories.

The hard part is for us to say just what those further implications *R* and $\neg R$ could be. The problem is that the most obvious implications of what Sonya said (relative to the background theory she believes) are *moral* implications, whereas the most obvious implications of what Kano said (relative to the background theory he believes) are *tworal* implications. For example, what Sonya said might be naturally thought to imply:

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

However, what Kano said implies:

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

You see the problem: *S* and $\neg T$ are jointly inconsistent if and only if *S* implies some proposition *U* while $\neg T$ implies $\neg U$; but if *S* only has moral implications, and $\neg T$ only has tworal implications, then we run into the same problem again when we go searching for U and $\neg U$. So what we need are some relevant *non*-moral and *non*-tworal implications, about which Sonya and Kano can plausibly be said to disagree.

It's here that the description of the MTE scenario works in our favour. Recall that it's crucial for T^E and T^{TW} to be similar to one another *not only* with respect to the relations they hypothesise to hold between the moral (or tworal) properties they each posit, *but also* with respect to how those theories are integrated into their respective societies and broader normative theorising—how, in other words, their use of moral (or tworal) concepts and language hooks up to the non-moral (and non-tworal) world. And (here's the key move of the argument) if T^E and T^{TW} are similar in these respects, then it's likely there will be *some* relevant non-moral (and non-tworal) questions regarding which Sonya and Kano disagree.

Here's one possibility—not the only one, but the one we'll be zooming in on. Given some natural background beliefs of the sort we could reasonably expect most humans to share—e.g., that the moral and the pragmatic reflect distinct normative domains, each 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—we can reasonably expect Sonya to believe there are compelling *non-prudential* reasons to give to charity.[7] Relative to that background, what Sonya said implies:

> R = *Typically, one ought all-things-considered to give to charity unless one has compelling pragmatic reasons to do otherwise.*

Kano has similar background beliefs, *mutatis mutandis*—e.g., that the tworal and the pragmatic reflect distinct normative domains, each of which yields distinct sorts of normative reasons for action, and that reasons stemming from tworal obligations are typically weightier. But unlike Sonya, he finds no compelling non-prudential reasons to give to charity. At best, doing so is supertwerogatory (twupererogatory?), and so pragmatic reasons can readily outweigh whatever tworal reasons we have to give to charity. Relative to that background, then, what Kano said implies:

> $\neg R$ = *It's not the case that, typically, one ought all-things-considered to give to charity unless one has compelling pragmatic reasons to do otherwise.*

Perhaps the question of R versus $\neg R$ is one among the cluster of questions that Sonya and/or Kano explicitly take to be under dispute. But we needn't commit to saying that; the $R/\neg R$ question is likely to be in the nearby vicinity of whatever questions they take themselves to be disagreeing about, and that will suffice. We can reasonably expect that Sonya will (at least tacitly) believe R, or something to the same effect, while Kano will (at least tacitly) believe $\neg R$, or something to the same effect. We can also reasonably expect that Sonya will recognise, on the basis of what Kano said, that Kano (at least tacitly) believes $\neg R$ or something like it, while Kano will recognise, on the basis of what Sonya said, that Sonya (at least tacitly) believes R or something like it. Consequently, we can expect that they will each *rightly* recognise that there's probably some conflict in their beliefs closely related to the moral/tworal status of charity-giving, and so they engage in what they each *rightly* take to be a more-than-merely-verbal dispute.

[7] The present example is designed to make sense on the hypothesis that the moral and the pragmatic *exhaust* the domain of normative reasons for action, or on the weaker hypothesis that these tend to be *the weightiest* reasons at play by a considerable margin. If the reader thinks there may be other kinds of weighty reasons—aesthetic reasons, say—then we can just replace 'strong pragmatic reasons' with 'strong pragmatic *or* aesthetic *or* … reasons' and the upshot will be unchanged.

If the example works, then it works because *R* is a *non*-moral and *non*-tworal proposition—i.e., neither specifically *moral* nor *tworal* concepts are required for one to have thoughts with the contents *R* or ¬*R*. If so, then the concepts needed to think such thoughts can be analysed independently of variations between Sonya's and Kano's folk moral theories, and we (*qua* analytic naturalists) can happily say that Sonya and Kano can have *R* and ¬*R* thoughts. That's how we can show that Sonya and Kano can have conflicting beliefs about the *R*/¬*R* question without getting tied up in squabbles over whether Sonya's *moral* beliefs (like *P* and *S*) are or are not inconsistent with Kano's *tworal* beliefs (like ¬*Q* and ¬*T*).

But *does* the content *R* count as non-moral and non-tworal in this sense? We argue that it does: aside from a bit of basic logic, the key concepts required to think a thought with the content *R* seems to be just the concepts of a PRAGMATIC REASON and of the ALL-THINGS-CONSIDERED OUGHT, and possessing *these* concepts doesn't seem to require the possession of any specifically *moral* or *tworal* concepts.

Consider first the concept of a PRAGMATIC REASON. This concept isn't naturally analysed 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. (We note that the same is true under a hedonistic analysis of PRAGMATIC REASON.) More generally, it's not at all difficult to imagine a wholly amoral (and atworal) society of *homines economici* whose members make frequent use of the concept of a PRAGMATIC REASON. So *even if* Sonya and Kano do indeed lack one another's moral and tworal concepts—and remember, *they need not*—they can still both have thoughts about pragmatic reasons.

Much the same seems to be true for ALL-THINGS-CONSIDERED OUGHT. This can (and should) be analysed without reference to anything specifically moral. 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; but in doing so, she doesn't say anything about any *specific* types of reasons. To have the concept of the ALL-THINGS-CONSIDERED OUGHT, Sonya need possess no more than the generic concept of a REASON and some idea of how different reasons can carry different weights that balance off against one another. None of this presupposes the possession of any specifically moral or tworal concepts. Again, the members of a wholly amoral and atworal society could still have the concept of the ALL-THINGS-CONSIDERED OUGHT, which they employ when considering trade-offs between (say) pragmatic, epistemic, and aesthetic reasons. Of course, those with different theories about which *kinds* of reasons there are will have different *conceptions* of the all-things-considered ought. But a difference in conception (i.e., what one happens to think about how a concept is instantiated as a matter of fact) is not a difference in concept. Two agents can have the same concept of CAR, for example, in the sense that they agree exactly about what that concept picks out at each and every possible world, yet have different *conceptions* by virtue of believing themselves to inhabit different worlds—e.g., one where all cars are red and manufactured by Ford, versus one where all cars are blue and manufactured by Toyota.

(Those sympathetic to conceptual holism will probably here want to disagree with us again, since holists will want to say that differences in conception will—typically, if not always—imply a difference in concept. If that's you, then we refer you back to what we said in §4. To that we can here also add: our present task is to show that Sonya and Kano have a more-than-merely-verbal disagreement. For this we need to find some implications *R* and ¬*R* of the appropriate kind—i.e., expressible in a shared *non*-moral and *non*-tworal language. But should this prove impossible, it does *not* follow that Sonya and Kano are *not* having a more-than-merely-verbal disagreement. For all we've said, S and $\neg T$ above might still be inconsistent even if we cannot prove it. More generally, the default presupposition cannot be that all of Sonya and Kano's relevant beliefs are consistent unless we prove otherwise—that's something our opponent needs to show if the MTE argument is to be compelling. So if it turns out that P and $\neg Q$ have zero implications that can be expressed in a shared non-moral/non-tworal language, then we appear to arrive at a stalemate: we cannot prove that Sonya and Kano *are* having a more-than-merely-verbal disagreement, but neither can our opponent prove they *are not.*)

In any case, we don't want to rest too heavily on this one example, and we certainly don't want to give the impression that our whole argument rests on whether Sonya and Kano can mean the very same thing by their respective uses of 'pragmatic reason' and 'all-things-considered ought'. It doesn't—even if they *do* mean slightly different things, *still* they might be expressing conflicting beliefs. Consider again the *Evil Johnny* case: sameness of meaning in any sense just isn't a prerequisite for genuine disagreement! More importantly, connections between moral claims and claims 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. There are also connections to epistemology, affect, motivation and behaviour—would that we could discuss them all without making this paper inordinately long!

What *really* matters here is the general reason for thinking that such examples are likely to arise. It goes like this: since T^{E} and T^{TW} must be *similar* with respect to how they link the moral (and tworal) properties they posit to the non-moral and non-tworal world, the fact that they *diverge* with respect to what they deem 'obligatory' (and 'good' and 'right' and so on) is then a reason to think that *those theories* are likely to have *some* inconsistent implications. *That's* what our example is supposed to be an example of: OBLIGATORY and TWOBLIGATORY are distinct concepts, to be sure, but they're tied to non-moral (and non-tworal) normative theorising in inconsistent ways *given* T^{E} and T^{TW} respectively. Consequently, if something *perfectly* satisfies the role of 'obligatory'$^{\mathrm{E}}$ in T^{E}, then nothing *perfectly* satisfies the role of 'obligatory'$^{\mathrm{TW}}$ in T^{TW}, and vice versa, because those roles place at least some inconsistent demands on the *non*-moral and *non*-tworal world. Perhaps nothing perfectly satisfies either role. Either way, at most one of T^{E} or T^{TW} is true.[8] Sonya and Kano recognise this, and so they argue.

§7. Conclusion

There's 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

[8] In more detail: we've assumed P and $\neg Q$ are both true; hence, there's no R such that P implies R and $\neg Q$ implies $\neg R$. However, where B_{S} captures Sonya's relevant background beliefs, and B_{K} captures Kano's, it's entirely possible that P & B_{S} implies R while $\neg Q$ & B_{K} implies $\neg R$. If so, then at most one of B_{S} or B_{K} is true. Since the key difference between them is that B_{S} includes T^{E} while B_{K} includes T^{TW}, we can reason that at most one of T^{E} or T^{TW} is true; in this case, presumably because there's an incompatibility in the divergent roles for 'obligatory' (e.g., in relation to the calculation of all-things-considered oughts). Those roles may still secure *imperfect deservers* for those roles (i.e., such that P and $\neg Q$ are both *true*, per our stipulation), but *any* such imperfection implies the falsity of the relevant theory (as noted in §2).

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’d be worried about the Moral Twin Earth scenario. On this understanding, Sonya and Kano are simply applying orthographically similar labels to overlapping but distinct disjunctions of actions. There’s no incompatibility in their beliefs about those actions *per se*, nor even in the vicinity. Their dispute boils down to nothing more than a disagreement about labels. However—and thankfully—that’s *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, behaviour, and non-moral normative theorising. 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 Sonya and Kano’s dispute is, *at least*, more-than-merely-verbal.

Bibliography

Chalmers, D.J. 2006. Two-dimensional semantics. In Lepore & Smith (Eds.), *Oxford Handbook to the Philosophy of Language*. Oxford University Press.

————. 2011. Verbal Disputes. *Philosophical Review*, 120(4): 515-566.

Finlay, S. 2017. Disagreement Lost and Found. *Oxford Studies in Metaethics* 12: 187-205.

Geirsson, H. 2005. Moral Twin-Earth and Semantic Moral Realism. *Erkenntnis* 62: 353-378.

Horgan, T. and Timmons, M. 2009. Analytical Moral Functionalism Meets Moral Twin Earth. In Ravenscroft (ed.), *Minds, Ethics, and Conditionals* (221-236). OUP.

Jackson, F. 1998. *From Metaphysics to Ethics*. Oxford University Press.

Jackson, F. & Pettit, P. 1995. Moral Functionalism and Moral Motivation. *The Philosophical Quarterly* 45: 20-40.

Jenkins, C.S.I. 2014. Merely verbal disputes. *Erkenntnis* 79: 11-30.

Khoo, J. & Knobe, J. 2018. Moral disagreement and moral semantics. *Noûs* 52: 109-143.

Levy, N. 2011. Moore on Twin Earth. *Erkenntnis* 75: 137-146.

Lewis, D. 1970. How to Define Theoretical Terms. *The Journal of Philosophy* 67: 427-446.

Merli, D. 2002. Return to Moral Twin Earth. *Canadian Journal of Philosophy* 32: 207-240.

Plunkett, D. & Sundell, T. 2013. Disagreement and the Semantics of Normative and Evaluative Terms. *Philosopher’s Imprint* 13: 1-37.

Rubin, M. (2014) Biting the Bullet on Moral Twin Earth, *Philosophical Papers* 43: 285-309,

Väyrynen, P. 2018. A Simple Escape from Moral Twin Earth. *Thought* 7, 109-118.