Moral Kombat

Analytic Naturalism and Moral Disagreement

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 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.

§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's unlikely that Sonya and Kano are merely exchanging noises with only the superficial appearance of communication and conflict. If she's right, then he's wrong.

If you're a moral realist, then it's 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're 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 $\neg P$. There's no miscommunication, the dispute isn't merely verbal, just a straightforward conflict in the beliefs expressed by their respective assertions.

But moral realists should want to make sense not only of intra-communal moral disagreements, but of inter-communal disagreements as well. Suppose now that Kano had spent his life on Outworld, where people to 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 judgements 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 versus XYZ), so too one might think the meaning of 'wrong' depends upon how it's 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's a difference of meaning. Having established that relativism of meaning, it's 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's asserting one thing *P*, whereas when Kano 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 (henceforth 'H&T'), who describe a hypothetical Moral Twin Earth ('MTE') populated by individuals with whom we purportedly

cannot have any genuine moral disagreements should naturalism prove true (H&T 1991; 1992; 2000; 2009). If this argument works, then there's a metasemantic fly in the naturalist's metaethical ointment. However, the argument *doesn't* work. The problem, we'll argue, concerns the connection between meaning and disagreement—namely, that there isn't any. The possibility of genuine disagreement requires neither sameness of meaning, 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]

Most discussions of H&T's metasemantic challenge centre upon synthetic moral naturalism. And to be sure, in H&T's original discussion the challenge was targeted at Boyd's (1988) non-analytic version of moral realism. But it was never supposed to be so restricted in scope. In H&T's words, the 'generic thought-experimental deconstructive recipe… is applicable to virtually any metaphysically naturalist version of moral realism' (2009, 221; see also H&T 2000, 139-142). And analytic moral functionalism is clearly among them; indeed, H&T (2009) explicitly subject it to their metasemantic challenge.

In this paper, we'll be responding to the metasemantic challenge from the perspective of analytic naturalism. It's unclear to us why there have been so few attempts to do so. Perhaps this can be simply 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. Parts of our response are specific to analytic naturalism. But we expect that others can be repurposed by synthetic naturalists as well. We leave readers to draw such connections for themselves.

We'll begin with some stage setting: §2 covers everything you need to know about analytic naturalism. §3 goes into further detail on the metasemantic challenge as it applies to that position. We present our

---

[1] Plunkett and Sundell (2013) also deny that accounting for substantive disagreement requires satisfying what they call the 'Shared Meaning Task' (cf. Khoo and Knobe 2018, on the 'Exclusion Inference'). Unlike Plunkett and Sundell, we don't suppose that disputees are engaged in metalinguistic negotiation—the disagreements we're interested in don't concern how a term should be used. Geirsson (2005) also denies that disagreement requires the assertion of incompatible propositions, but in doing so appeals to a distinction between public and speaker meaning whereas we don't take disagreement to depend upon sameness of meaning in *any* sense. Many others have suggested expressivist theories of disagreement, according to which disagreement need not consist in the assertion of conflicting beliefs (see Finlay 2017 for an overview). Again our strategy is distinct: we want to capture the possibility of a strong form of non-expressivist, cognitive disagreement within an analytic naturalist framework.

response in §§4–7. By the close of the paper, we hope to have shown that analytic naturalists can make sense of a strong form of genuine inter-communal moral disagreement even across the kinds of communities described in a typical MTE scenario.[2]

§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 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 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 analyse the terms of our moral vocabulary by reference of their respective roles within this folk moral theory. The analytic naturalist will, for instance, want to say that our term 'right' can be analysed in terms of the 'rightness'-role in our folk moral theory.[3]

---

[2] At the risk of spoiling a joke by explaining it, we should say something about our section headings for those who may be otherwise perplexed. *Mortal Kombat* is a game in which two characters, often from different worlds, fight. Matches have a "best two out of three" structure, and one scores a 'flawless victory' by winning the first two rounds without incurring damage. Between matches, players may engage in a 'Test your might' mini-game, and at the end of a match, the announcer may ask the winner to 'Finish him!'

[3] There's 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, our preference would be to precisify these claims about 'meaning' 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 any important part in the ensuing discussion—every point we want to make can be made adequately with a one-dimensionalist theory of meaning.

It will be helpful to say a little more about our key terms, starting with 'the folk'. There's 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 shouldn't matter for our discussion. We'll 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're 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 needn't be apparent even to those who regularly use and adhere to them (cf. Jackson & Pettit 1995 on 'ethocentric' belief). 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, 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—a 'mature folk morality', as Jackson (1998) puts it. Again, though, these are matters for debate within analytic naturalism, and shouldn't much impact our discussion. What matters 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 is no requirement of analytic naturalism that *every* claim or principle within folk moral theory must end up being correct. Analytic naturalists certainly *aren't* proposing to reduce moral theorising to the mere cataloguing of folk moral theories! On 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 and so forth, all perfectly occupy their respective roles within that theory (see Lewis 1970). Consequently, if any of these properties *imperfectly* occupy their respective 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's a single moral property that plays two roles when really there's 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 plain wrong).

One therefore 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 is deemed obligatory by our folk moral theory. To suppose as much would be to misunderstand analytic naturalists. There's a lot more to the 'obligatory'-role than just which actions the term 'obligatory' does and does not get attached to, and it needn't be the case that every aspect of that role ends up being satisfied. (This will be important.)

It is, then, entirely 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 in the manner characterised above. 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. (This will also be important.)

§3. Opponent Select: Moral Twin Earth

Now onto the challenge. 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, such that if that theory had been different then the meaning of 'wrong' may well have been correspondingly different. According to H&T, this reveals that the analytic naturalist is not in fact a *realist* at all, but merely a relativist dressed in realist garb. The naturalist is, accordingly, thought to have much the same trouble accounting for genuine disagreement as relativists typically do.

H&T's MTE 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 (twumans) of Outworld have some distinct theory, M^{out}. Suppose that 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}.

By hypothesis, M^{earth} and M^{out} are similar. They are 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 theorising. 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]; [twin-humans] 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)

We might imagine, perhaps, 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 we really shouldn't get bogged down in the details—what's important is that the scenario is described such that the following are satisfied:

Divergence. M^{earth} and M^{out} differ with respect to the 'obligatory'-role, such that their respective uses of the terms have divergent analyses and extensions.

Similarity. M^{earth} and M^{out} are similar, especially with respect to how they integrate into society and general normative theorising.

In short: **Divergence** is needed to generate the required differences in meaning, whereas **Similarity** is needed to generate the intuition that the dispute isn't merely verbal. 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, 'permissible' when it involves going to the gym, and assert that the regular performance of 'obligatory' actions tends to promote fat loss, muscle gain and increase life expectancy, and so on in like fashion. Here, there would be no temptation to say that M^{out} is a moral (or tworal) theory—the Outworlders would simply be using the sounds we associate with moral properties to talk about exercise.

Against this backdrop, H&T's challenge concerns whether the people of Earthrealm and Outworld are capable of having more-than-merely-verbal disputes about morality. Are Sonya and Kano just talking past one another? It certainly seems like there *could be* something substantive underlying their dispute, and perhaps even that there *probably 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 these intuitions—and better, all else equal, if we can do so without having to diagnose them as some kind of widespread mistake.

H&T, however, seem to be of the opinion that only one style of explanation is up to the task: to do justice to the intuition that this dispute is more-than-merely-verbal, we need to be able to say that Sonya and Kano *mean the same thing* by their terms.[4] The correct diagnosis of the intuition, 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 twin-humans] would be genuine disagreements— i.e., disagreements in moral belief and in normative moral theory, rather than differences in meaning. (2009, 227)

The analytic naturalist, however, has in principle 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 M^{earth}, whereas the twin-moral terms used by [twin-humans] 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. (2009, 226-7)

And

> … The different parties are expressing different concepts with their moral terms, are talking past one another rather than disagreeing. (2009, 232)

---

[4] We take this presentation to reflect how the challenge is standardly reported 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).

In other words, the role of 'obligatory' in M^{earth} must *diverge enough* from its role in M^{out} that Sonya's and Kano's uses of the term have *incompatible* analyses, in which case Sonya's term 'obligatory'^{earth} will designate a distinct property than 'obligatory'^{out} does in Kano's mouth. Note that it's not enough that the terms simply have *different* analyses, since terms with different analyses may still fix on the same extension. It's the *incompatibility* that matters: incompatibility guarantees *a difference in extension*, and it's the difference in extension that explains why their assertions aren't really in conflict.[5] Essentially: what Sonya says is true iff giving to charity has some property *f*, and what Kano says is true iff 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 merely verbal dispute. (Spoiler alert: it doesn't.)

But the problems don't end there! If analytic naturalism is correct, H&T add, then Sonya and Kano *cannot even possess one another's concepts*, and hence cannot even understand what one another are saying:

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

Specifically, H&T argue that analytic naturalists are guilty of what they call *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 wouldn't 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-centred' (2009, 228).

In summary: H&T's 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 that each is capable of possessing. We humans are conceptually cut off from understanding the twumans' tworal assertions, and they from understanding our moral assertions, precisely because of the differences between our folk moral theories.

---

[5] We assume that two predicates have *incompatible* analyses just in case those analyses are such that, at every possible world, there are things in the extension of each 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 actually explain what they mean by 'incompatible', but this is the definition that seems to make the most sense of their argument. The details don't matter so much, though, so long as the analyses are such as to guarantee an appropriate difference of extension.

Our aim in what follows will be to argue that the challenge fails, primarily in virtue of tying meaning and disagreement too closely together. But first we want to firmly reject the charge of chauvinistic conceptual relativism. Analytic naturalism carries no commitment to many of the claims about meaning and concepts that it's so often burdened with; pace H&T, it is indeed possible for Sonya and Kano to share one another's concepts and to fully understand one another (§4). After dispatching the relativism charge, we will argue that it's *more likely than not* that Sonya and Kano's dispute is substantive. In doing so, we do not assume that Sonya and Kano fully understand one another—they *might*, given what we say in §4, but *whether* or not they do isn't essential to making sense of their dispute being genuinely meaningful. After supplying the relevant background on disagreement and disputes (§5), we present our main argument (§6), followed by some final mopping up (§7).

§4. Test Your Might: Chauvinistic Conceptual Relativism

Let's accept that Sonya and Kano associate different meanings—different concepts—with the terms of their respective moral and tworal vocabularies. It's much less clear whether we should also accept that their terms have 'incompatible' analyses and thereby refer to distinct properties. As others have observed, there is a delicate balance to be struck between **Divergence** and **Similarity** (cf. Merli 2002; Levi 2011; 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 plausible it is 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 disagreement here that needs accommodating. It's not entirely obvious that this tightrope can be walked.

But let us grant for the sake of argument that the scenario is coherent, and that Sonya and Kano's respective uses of 'obligatory' do indeed come with incompatible analyses. Does it follow that, according to analytic naturalism, the twumans 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 aren't locked up in there with them?

Not even a little. There's nothing about analytic naturalism that makes it impossible for humans and twumans to understand one another, nor to express and critically evaluate one another's folk moral theories

with full comprehension thereof. The truth is the opposite: there's 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.

Before getting to the meat of our argument, some ground-clearing will be helpful. First, we understand *concepts* to be 'parts of propositions', loosely construed.[6] 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 recognise 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 isn't.

We take it that this is a fairly ordinary conception of concept possession in philosophy. Note that on this conception, one needn't 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 flavours 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 (and frankly baffling) assumptions regarding the relationship between what an individual's words mean in her own community and which concepts she *lacks*.

---

[6] 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 representational structures that are sometimes hypothesised to explain our recognitional and categorisational capacities. Analytic naturalism includes no commitment regarding the existence of concepts in the latter sense, and has nothing much to say about what it takes to possess a certain kind of representational vehicle or share such vehicles between peoples.

Now to be sure, there are plentiful 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 Connection. 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 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'$^{\text{out}}$.

So far so good. But this only tells us about the concepts they *have*, and nothing at all about the concepts they *lack*. A further premise is clearly needed if we're going to derive the conclusion that Sonya and Kano *do not*—and indeed *cannot*—possess one another's concepts. And it's 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. So we need something like:

Possession-to-Theory Connection. A person possesses the concept c expressed by a term whose analysis can be given in terms of the role it plays in folk moral theory M only if that person belongs to a community whose folk moral theory is M

This would also imply, by *tollens* on the **Understanding-to-Possession Connection**, 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 charge of chauvinistic conceptual relativism is premised on something like the **Possession-to-Theory Connection**.

But analytic naturalists should emphatically reject the **Possession-to-Theory Connection**; for 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. What H&T get right is that analytic naturalism renders our moral terms thoroughly theory-laden. Their mistake is to think this entails any lack of comprehension between those with different theories. Analytic naturalists know better, for they've internalised 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!

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, …,* 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 OBLIGA-TORY 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 follows: if Sonya has the resources to understand the concepts of M^{earth}, then she *also* has 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?

If it turns out, then, that M^{earth} and M^{out} 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 there's no reason to think that Sonya and Kano shouldn't also be capable 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 recognising that M^{out} posits its own distinct suite of prop-erties *x′, y′, z′, …,* 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. It's simply not true that, according to analytic naturalism, a person can possess the concepts analysed by reference their role in a folk moral theory M only if M is *their* folk moral theory. If analytic naturalism were committed to anything as implausible as the **Possession-to-Theory Connection**, then MTE would be the least of its worries.

(Some readers may worry here about *conceptual holism*: perhaps the differences between M^{earth} and M^{out} will 'infect' Sonya's and Kano's other concepts, such that they cannot share a common base of *non*-moral and *non*-tworal O-terms. We'll set that matter aside here, and return to it in §7. Suffice it to say for now that if the charge of chauvinistic conceptual relativism rests on a presupposition of conceptual holism, then our response is simply to reject holism.)

§5. Round One: Disagreement and Disputes

We've argued that it's consistent with analytic naturalism that Sonya and Kano are *in principle* capable of fully understanding one another's assertions. But our 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 substantive moral dispute. That will be our goal for the remainder of the paper. The point of this section is to get clear on what makes a dispute feel intuitively *real*, or *genuine*, or *substantive*, or *more-than-merely-verbal*. Before we get to that, though, it'll be helpful to begin with some key distinctions and a bit of stipulated terminology.[7]

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 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 doesn't quite capture his thoughts. Following discussion, it comes to light that what he said isn't 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's what Kano *doesn't* 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 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 answers, and we say that agents *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 interlocutors—the *disputees*—dispute one another. To *dispute* (verb) is to attempt to express what one

---

[7] Parts of the following discussion overlaps in various ways with points made by Geirsson (2005), Chalmers (2011), Plunkett & Sundell (2013), and Jenkins (2014). We note again that our interest here is in cognitive disagreement—essentially: conflict in belief. We do not assume that this is the only legitimate way to understand disagreement; we do, however, argue that analytic naturalists don't *need* appeal to any expressivist notions of disagreement to account for MTE-style intuitions.

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 one any question of interest will typically affect how one answers many other questions, and so disputes will often relate to clusters of questions. These we say are the questions 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 turn out to be something rather different than what the disputees take to be the questions under dispute. This does not, however, automatically render the dispute non-substantive. Consider:

Pro-Life Kano. *Kano 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 later discover they agree entirely about those norms, but disagree about a related question—viz., when a foetus develops consciousness.*

In this case the dispute is grounded in a mistake about the nature of the disagreement; they are aligned in their conception of which questions are under dispute, but do not in fact disagree about those questions. In this case, though, Sonya and Kano still disagree about some substantive matters of nearby relevance to the questions that each 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 in total misalignment, and, moreover, there's 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.

(In saying this, we are supposing that disagreement about the meanings of words usually won't 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 won't be especially relevant to most of your beliefs about *prime ministers*. Obviously, this doesn't apply in cases where the disputees take themselves to be disagreeing about the meanings of words, or how their words ought to be used.)

That, at least, is how we're conceiving of merely verbal disputes. Perhaps you disagree with us. But no worry. All that's required for our argument in §6 is the following two theses:

Miscommunication is Insufficient. A dispute can be non-verbal even if differences in meaning lead to misunderstandings and miscommunication.

Compatibility is Insufficient. A dispute can be non-verbal even if the propositions asserted by the disputees are compatible, and indeed even if both propositions happen to be true.

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

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 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 on Outworld, "conscious" means the same as Earthrealm's "is able to speak".*

The dispute strikes us as very much substantive. While there's *some* miscommunication, and *some* misalignment in what each disputee takes to be the questions under dispute, they nevertheless *do* disagree regarding those questions. Kano 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. 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. P implies R: it's permissible to torture any conscious being that speaks for fun. $\neg Q$ implies $\neg R$, and therefore implies $\neg P$. There's something defective about how the dispute has been conducted, but neither party is grossly mistaken about the nature of their disagreement. That seems to suffice to make the dispute feel more than merely verbal.

Next **Compatibility is Insufficient**:

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

Kano asserts *P*, and Sonya replies with ¬*Q*. But Sonya disagrees with Kano about when the match will begin, among other things, and though compatible with *P* her assertion of ¬*Q* expresses that disagreement via implication *relative to* what she takes him to believe. That is, Sonya believes that Kano has certain background beliefs *B*—that the match will begin when it rains, that there's a nearby storm to the East, and so on—and she believes (and believes that Kano will recognise) that ¬*Q*∧*B* implies ¬*P*.

Note in this case that *P* and ¬*Q* might not just be compatible but also both *true*. Kano may be mistaken about the match starting when it rains, or it may be there's rainclouds off to the West that neither of them saw. Even so, the dispute would still be substantive. The point of the example is that Sonya and Kano's disagreement need not be apparent in the truth conditions of *what's said*; one also needs pay attention to the context—and in particular the beliefs—of those doing the saying.[8]

With that in hand, it's time we return to MTE. Essentially what we'll argue is that, given the way the scenario is constructed—in particular, given **Similarity** and **Divergence**—the analytic naturalist has good reason to suppose that Sonya and Kano's dispute over the obligation to give to charity is more likely than not more-than-merely-verbal. There is likely to be some miscommunication (similar to the **Evil Kano** case). It may even be that they express compatible propositions. Nevertheless, given very reasonable presumptions about their respective background beliefs, the likely substantive nature of their moral disagreement will be apparent (similar to the **Weatherman Kano** case).

§6. Round Two: Interdimensional Disagreement

Let's start with some assumptions. We assume that Sonya and Kano are competent speakers of their respective languages, and aren't debating in bad faith. We assume that Sonya more or less accepts the tenets of her folk moral theory M$^{\text{earth}}$, and is generally similar to other humans with respect to how she thinks about and behaves in response to moral judgements. Likewise for Kano, *mutatis mutandis*. We assume—

---

[8] One might want to say that Sonya and Kano's assertions are incompatible *given that context*, and that *this* kind of contextual incompatibility is still necessary for a dispute to be substantive. That might be so. If it is, then read our argument in §6 as saying that Sonya and Kano's assertions are incompatible *in context*.

for the sake of argument—that the analysis of 'obligatory'$^{\text{earth}}$ is *incompatible* with the analysis of 'obligatory'$^{\text{out}}$, 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. Sonya, for instance, doesn't much care about which meanings get attached to which sounds as uttered by Outworlders' mouths—intuitively, what she really cares about is that Kano doesn't fully recognise 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 all of these assumptions are an intended part of the scenario; some more explicitly than others. From these assumptions we're licensed to infer that when Sonya says 'Giving to charity is obligatory', she's expressing some content P she happens to believe, and that's true according to M$^{\text{earth}}$; and when Kano says 'Giving to charity is not obligatory', he's expressing some content $\neg Q$ he believes, and that's true according to M$^{\text{out}}$. Furthermore, P and $\neg Q$ are compatible; indeed they're both *true*. But we are *not* thereby licensed to infer that the dispute is merely verbal. Quite the opposite: given these assumptions, it's more likely than not that their dispute is more than verbal; indeed, the dispute is probably *about* more or less what the disputees take it to be about, or something in the very nearby vicinity.[9]

Here's a quick overview of the reasoning. Though P and $\neg Q$ are both true, it's 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. In any case, Sonya and Kano disagree on this matter of R versus $\neg R$. Worry not what R is just yet; we'll get to that very soon. What matters is that from this disagreement over R we can go on to infer a likely *further* disagreement regarding some matter directly relating to the moral—yes, *moral*—import of giving to charity. Or that's what we'll argue.

---

[9] Why 'more likely than not' and 'probably', rather than more forceful assertions? Because there are indefinitely many ways we could spell out the details of the folk moral theories and the possibly relevant background beliefs Sonya and Kano might have consistent with the extremely sparse description of the scenario, all of which might make a difference to our intuitions about the case. In the absence of more details, the best we can reasonably offer are generic probabilistic reasons.

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 doesn't, 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 doesn't, and satisfies the requirements for tworal responsibility, that person has done something tworally blameworthy*

You see the problem: it's not obvious that S and $\neg T$ are inconsistent. To establish *that*, we'd need to find some R such that S implies R and $\neg T$ implies $\neg R$; but of course that's 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) issue about which Sonya and Kano can plausibly be said to disagree.

It's here that **Similarity** works in our favour. M^{earth} and M^{out} are similar *not only* with respect to the theory-internal relations they hypothesise 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 theorising—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's 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—e.g., 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 *there are especially weighty non-prudential reasons to give to charity*, and as such *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 beliefs, *mutatis mutandis*, inasmuch as is consistent with his acceptance of a similar-but-distinct folk moral theory. 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 isn't disposed to give to charity, nor to advocate that others must do so. At best, he thinks, giving to charity is twupererogatory. (Supertwerogatory?) So Kano presumably believes *there are no especially weighty non-prudential reasons to give to charity*, and *it's not the case that one typically ought all-things-considered give to charity unless one has strong prudential reasons to do otherwise*. Since their background beliefs are similar in the relevant respects, we can reasonably expect that Sonya and Kano will draw these implications out of what each party asserted.[10]

That is a non-moral (and non-tworal) 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 doesn't *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 in relation 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 inconsistent with M^{earth}. 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 theory is making false claims about how 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 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 weight that Sonya thinks it does.[11]

Now, if our illustrative example works, it works because *there are weighty non-prudential reasons to give to charity*, and *one typically ought all-things-considered give to charity unless one has strong prudential reasons to do otherwise*, are non-moral and non-tworal propositions—neither specifically moral nor tworal concepts are required for one to have thoughts with these contents. Are they non-moral and non-tworal in this sense? We argue that they are. The key concepts required to think the relevant thoughts seem

---

[10] The example is designed to make sense under the hypothesis 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—aesthetic reasons, say—then just replace our 'strong prudential reasons' with 'strong prudential *or* aesthetic reasons' and the upshot is unchanged.

[11] We do not assume Kano possesses any moral concepts, and hence don't assume that Kano will be able to express his disagreement with Sonya using concepts they both possess. He may or may not be able to do so; what's required to establish a conflict in belief is that Kano believes something that entails the negation of what Sonya believes. Disagreement implies neither awareness of that disagreement nor the capacity to precisely verbalise the disagreement in some common language.

to be just the concepts of PRAGMATIC REASON and 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—it's 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. (We note that the same is true under a hedonistic analysis of PRAGMATIC REASON.) So *even if* Sonya and Kano do indeed lack one another's moral and tworal concepts— and remember, *they need not lack one another's concepts*—still they might both have thoughts about pragmatic reasons.

Much the same seems to be true for ALL-THINGS-CONSIDERED OUGHT. This can (and should, we think) be analysed 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; 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 of different kinds might carry different weights that can be balanced off against one another. None of this presupposes 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 they employ when considering trade-offs between (say) pragmatic and epistemic reasons without ever having moral or tworal reasons get into the mix.[12]

In any case, the point is not to rest our argument on this one 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

---

[12] Those with different theories about reasons will have different *conceptions* of the all-things-considered ought. But a difference in conception—i.e., what one believes about how the concept is instantiated—is not a difference in concept. Two agents can have the same concept of CAR, in the sense that they agree exactly about what that concept picks out at each possible world, yet have different *conceptions* by virtue of believing themselves to inhabit different worlds—e.g., one where all cars are manufactured by Ford, versus one where all are manufactured by Toyota. Again, some readers may worry here about conceptual holism, and again we'll defer until §7.

our example.) Folk moral theories also posit connections between the moral realm and the realms of epistemology, belief, affect, motivation, and behaviour. 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 general rationale for thinking that *some* relevant disagreement will likely be present given the way the scenario is characterised. It 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 *diverge* with respect to what they deem 'obligatory' is a reason to think that those theories are likely to have 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 *something* 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 *non*-tworal world—demands that cannot both be perfectly satisfied. Perhaps nothing perfectly satisfies either. In any case, at most one of M^{earth} or M^{out} is true—and, presumably, where they come apart will have something to do with what they say about the import of the reasons in favour of giving to charity. Sonya and Kano recognise this, by virtue of recognising the inconsistent implications of their theories, and so they argue.

§7. Finish Him!

There is still the issue of conceptual holism, 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 $\neg R$. We've been putting off discussing holism because we've neither the space nor the inclination to go chasing down this rabbit hole. So let us quickly note two things instead. First: analytic naturalists are not committed to conceptual holism, and it's 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 doesn't 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 twumans cannot

even share their *non*-moral (or *non*-tworal) concepts. If the metasemantic challenge really does end up resting on some presumption of conceptual holism, then our response is straightforward: we reject the holism. And good riddance.

But perhaps there's a middle-ground. Rather than simply analysing *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}. The upshot is that we can no longer presume a shared base of non-moral-but-still-normative concepts.

It's 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. As far as the *original* MTE scenario is concerned, it doesn't much change our response. We don't have to establish that Sonya's non-moral-but-still-normative terms will be *analysed* the same way as Kano's non-tworal-but-still-twormative terms. The *analyses* are irrelevant. What matters, recall, are the *extensions*; the similarity of their respective broader normative and twormative theories as required in the original scenario presents plenty of reason to expect that those analyses will latch on to the very same extensions.

So we must assume instead that we're considering a revision of the original scenario: not just the *Moral* Twin Earth, but some Normative Twin Earth (cf. Eklund 2017). Of course, for the scenario to work it will need to satisfy the same general properties, *mutatis mutandis*, that were required for the moral case—namely, **Divergence** and **Similarity**. So, for instance, we'll need to presume that N^{earth} and N^{out} diverge with respect to the 'pragmatic reason'-role in such a way as to generate incompatible analyses and thus divergent extensions. And we will need to presume that N^{earth} and N^{out} are similar *qua* normative theories in the required manner. So, 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's* the case, then the same general rationale as above applies. 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 the sorts of things 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, there will more likely than not be genuine disagreements between those who hold to N^{earth} and those who instead go along with N^{out}.

§8. Flawless Victory

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 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 MTE 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 over 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 intercommunal moral disputes can reflect genuine, more-than-merely-verbal disagreement.

Bibliography

Boyd, R. 1988. How to be a Moral Realist. In Sayre-McCord (ed.), *Essays on Moral Realism*. Cornell University Press.

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.

Eklund, M. 2017. *Choosing Normative Concepts*. Oxford University Press.

Finlay, S. 2017. Disagreement Lost and Found. In Shafer-Landau (ed.), *Oxford studies in Metaethics* 12. Oxford University Press.

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

Horgan, T. and Timmons, M. 1991. New Wave Moral Realism Meets Moral Twin Earth. *Journal of Philosophical Research* 16: 447-465.

————. 1992. Troubles on Moral Twin Earth. *Synthese* 92: 221-260.

————. 2000. Copping Out on Moral Twin Earth. *Synthese* 124: 139-152.

————. 2009. Analytical Moral Functionalism Meets Moral Twin Earth. In Ravenscroft (ed.), *Minds, Ethics, and Conditionals* (221-236). Oxford University Press.

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

Jackson, F. & Pettit, P. 1995. Moral Functionalism and Moral Motivation. *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. *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.