

The Metaethical Presuppositions of Conceptual Engineering¹

Preston J. Werner, Hebrew University of Jerusalem

(Pre-Print Version - Please cite final version in *Oxford Studies in Metaethics* vol. 20)

“[T]he presupposition that there are *some* moral truths cannot be avoided by those engaged in justified political resistance.”²

“[I]n virtue of refusing a problematic practice/adopting a reactionary practice, one is, de facto, involved in political resistance via linguistic means.”³

Concepts are tools. Like other tools, they can be better or worse in a variety of ways. Sometimes our tools don’t do their jobs very well. Sometimes we realize that we don’t have a tool for an important task, and we want to invent one. Sometimes we have a tool that fulfills its purpose quite well, but its purpose is unjust or oppressive. Each of these problems arise with concepts just as they do with other tools. *Conceptual engineering* is the project of adding, improving, or removing concepts.⁴

There are many projects in first-order conceptual engineering,⁵ as well as, more recently, reflection on the nature and metasemantics of conceptual change. What is more surprising, given the role of normative considerations in arguments for conceptual change, is that little if any sustained work has been done on the metanormative commitments of conceptual engineering. In this paper, I begin to fill that gap by connecting the metasemantic questions with the metanormative ones. The normative considerations of conceptual engineering are more deeply entwined with metaethics than other normative domains because of conceptual ethics’ murky relationship to metasemantic theory more generally. Ultimately, the possibility of many instances of conceptual engineering depends on normative truths that are intersubjectively shared. Or so I argue.

The plan of the paper is as follows. In section 1, I briefly discuss conceptual engineering and conceptual amelioration as a methodology. In section 2, I go over three criteria for an adequate theory of how conceptual amelioration can work (methodological, metasemantic, and semantic). This shows that such an account is far from trivial to provide. Then, in section 3, I consider the most natural approach to providing such a theory - what I call *Ameliorative Pragmatism*. However, I argue that such an approach cannot deal with some

¹ For comments and discussion on earlier versions of this paper, I’d like to thank Derek Baker, Matt Bedke, Teresa Bruno-Niño, Janice Dowell, Andy Egan, Christopher Gibilisco, Camil Golub, Ben Henke, Seungsoo Lee, Bar Luzon, Oded Na’aman, Ittay Nissan-Rosen, Caleb Perl, Joshua Schechter, Mark Schroeder, Paulina Sliwa, Christine Tiefensee, and Mark van Roojen, as well as the audience at the Madison Metaethics Conference XX. I want to especially thank Aaron Elliott for reading and commenting on several successive drafts.

² Haslanger (2017), 165.

³ Simion & Kelp (2023), 19. NB: The epigraphs were removed from the final print for copyright reasons.

⁴ The term was first used by Creath (1990).

⁵ Scharp (2007), Haslanger (2012), Cantalamessa (2021), Wakil (2021), Eder (2021).

of the semantic phenomena that surround conceptual amelioration projects. This motivates section 4, where I defend the claim that the best strategy for explaining the metasemantics of conceptual engineering is to endorse the claim that normative facts play a role in reference determination. I argue that only such a view can capture the three criteria of an adequate account of conceptual engineering. I then go on, in section 4.2, to provide two concrete and independently plausible metasemantic theories that endorse this normative-metasemantic entanglement, and then in section 5, I show how these views meet the three conditions. In section 6, I turn to the metaethical implications of normative-metasemantic entanglement, arguing that, given other constraints on semantic theory, a metaethical view must meet a condition I call *Broad Intersubjectivity*. How robust such an intersubjectivity claim is depends on other commitments in semantic theory that one might have. In section 7, after considering an objection, I briefly draw out a few more interesting implications of the view. Finally, in section 8, I sum up.

1. Conceptual Engineering: A Very Brief History

Conceptual engineering in analytic philosophy has two independent origin stories. The first goes back to Carnap's notion of "conceptual explication", found in his *Logical Foundations of Probability*. Carnap was largely concerned with conceptual precisification for the purposes of scientific theorizing.⁶ But even within the realm of 'pure' philosophy, conceptual explication will need to be done; for example, Carnap himself wants to provide explications for concepts like PROBABILITY and ENTAILMENT. Carnap, in line with his general pragmatism, proposed that we precisify our concepts such that they will best meet our theoretical goals.

The second story is much more recent, found in the feminist work on gender and race by Sally Haslanger.⁷ Haslanger considers three kinds of philosophical analysis that one may want to engage in. A *conceptual* inquiry involves what we might think of as traditional conceptual analysis — assessing our current concepts by considering a variety of actual and hypothetical cases, and whether we would place these cases within or outside of the category at an intuitive level. A *descriptive* inquiry is interested in investigating the extension of some particular concept via (largely) empirical investigation. Finally, there is an *ameliorative project*, which investigates neither our current concept nor the current extension, but rather is a normative investigation into what some particular concept *ought* to be.

To see how this distinction works in practice, consider the concept of RACE. At least common sensically, the concept of RACE picks out some kind of biological groupings of

⁶ To what extent Carnap would have been comfortable extending his notion of explication to socio-political concepts is an open question. See Pearson (2017), Dutilh Novaes (2020, Section 5) for discussion.

⁷ See the essays collected in Haslanger (2012), as well as Haslanger (2020).

individuals. On the other hand, investigation into the concept of RACE reveals that there appears to be no natural biological kind that meets this intuitive conception; on the contrary, RACE appears to track social facts. When engaged in an ameliorative project with respect to RACE, our question is what extension for our RACE concept(s) would best meet our social/political aims — in this case, the aims of anti-racist activism. It is this third kind of project, ameliorative projects, that fall under the umbrella of conceptual engineering.⁸

Since Carnap, and especially since Haslanger's early work on gender and race, work on conceptual engineering has exploded across several domains in philosophy, from philosophy of science, to political philosophy, to medical ethics.

2. Criteria for a Metasemantic Theory of Conceptual Engineering

As its name implies, conceptual engineering suggests engaging in the attempted amelioration of our concepts. An account of conceptual amelioration, especially one of *intentional* conceptual amelioration, raises metasemantic questions, as has been widely noted.⁹ The first, and perhaps most serious, problem is how conceptual amelioration is even metasemantically possible. Herman Cappelen (2018) has provided a powerful argument that it is not, and any vindication of conceptual engineering should provide a response to this argument. A second, often overlooked, desideratum on a metasemantic theory for vindicating conceptual engineering is that any such metasemantic theory must be able to make sense of the first-order methodology and practices of conceptual engineers. As we will see, this is not at all trivial. Finally, a metasemantic theory of conceptual engineering must be compatible with the felicitousness of utterances made by competent speakers in the context of what appears to be conceptual change.

I'll take each of these conditions in turn, but first, a note about my ambitions at present. As I hope will become clear, the account of conceptual engineering that I defend below is, in my view, necessary for vindicating many instances of conceptual engineering. However, I don't intend it to be applicable to any kind of project that goes under the heading of "conceptual engineering", because I don't think that there is going to be any single one-size-fits-all theory of conceptual engineering. Some cases of conceptual engineering will involve the introduction of homonyms, some will involve stipulative definition, some will be guided, and some will not. In a way this may seem a bit unfair, because it forecloses the

⁸ A final issue hovering in the background concerns the target of conceptual engineering projects. Despite its name, there has been much dispute about whether conceptual engineering is genuinely about concepts. It may, on the other hand, be about words, inferential patterns, the speaker-meaning of utterances, or even classification procedures. I am officially neutral on what to think about the difficult issue of the proper target(s) of conceptual engineering.

⁹ See, e.g. Cappelen (2018), Sawyer (2020, forthcoming), Schroeter & Schroeter (2020), Pinder (2021).

philosopher's favorite game of finding counterexamples to refute a general account. But unfortunately, I just think the relevant semantic and metasemantic phenomena in question are too complicated to expect any single universally applicable account.

2.1 The Metasemantic Control Challenge

A helpful way of modeling metasemantics (for our purposes) is as an interpretation function, taking "as input all of the potentially relevant empirical facts about the use of a term and [delivering] as output the reference of that term (its extension and intension)".¹⁰ Given this, successful conceptual engineering requires, in some way, changing the relevant empirical facts that feed into the use of a term to change the relevant output reference. This isn't to say that the conceptual engineer needs to have a fully worked out metasemantic theory, or even has to know what metasemantics *is*, although perhaps that would help. Rather, it is just to say that whatever facts feed into the interpretation function must be in some sense in the control of agents or groups engaging in an engineering project.¹¹ However, given independently plausible metasemantic theory, it is unclear how this could be possible. This is the Metasemantic Content Challenge (MCC).

The most detailed defense of the challenge comes from work by Herman Cappelen (2018, 2020). As Cappelen presents the challenge, it combines an epistemic point and a metaphysical point; I will focus on the metaphysical one. Here is Cappelen's brief statement of the metaphysical point:

The process of conceptual engineering is governed by factors that are not within our control: no individual or group has a significant degree of control over how meaning change happens. Even if we could overcome our epistemic limitations—and know all about the relevant factors for a particular case—what we would have knowledge of would be something we had little control over.¹²

In order to assess the strength of this claim, we need to consider what kinds of facts feed into the interpretation function. Which facts these are will depend on our metasemantic theory, but consider a few plausible contenders:

- The relation some term has to the demonstrated object or property, or reference-fixing description, of some "initial baptism event".

¹⁰ Schroeter & Schroeter (2020), p.567-568.

¹¹ Sometimes this challenge gets called the *Implementation Challenge*. Unfortunately, in the literature on conceptual engineering, there is a common conflation between two kinds of challenges which both get called the "Implementation Challenge". On the one hand there is the challenge I am calling the Metasemantic Control Challenge. On the other hand, there is a more practical and political challenge of how to get widespread political uptake for one's ameliorated concept. I am not concerned with this latter challenge here, which presupposes an answer to the former one. See Nimtz (2024) for discussion of the latter objection.

¹² Cappelen (2018), 72-73.

- The causal chain of tokenings of some term or concept both within and between speakers throughout historical use of the term up until the present.
- Which object or property is the dominant source of information¹³ for beliefs associated with the term/concept.
- Which object or property is the extension of the term/concept under conditions of omniscience in a canonical language (which doesn't itself contain the term/concept, on pain of circularity).
- The object or property which, if interpreted to be the referent of the term/concept, would make speakers as rational as possible (and in accordance with the interpretation of surrounding terms/concepts).
- An ordering of properties from most natural/fundamental to least natural/fundamental (and most gerrymandered).
- The object or property that a term/concept was "set up to be set off by", either in a creature's evolutionary history or their particular developmental history.

Different metasemantic theories will adopt some subset of these conditions, or conditions resembling them, as the proper inputs to the interpretation function. Now take some term/concept *C*. Suppose we want to ameliorate *C* - the current referent of *C* is *F*, but it would be better if it had the referent *F**. Getting *C* to refer to *F** instead of *F* requires us to change the facts which constitute inputs to the interpretation function. But looking at the list above, this looks like a nearly hopeless enterprise: We are not in a position to change the past. We are not in a position to change facts about which properties are more natural or fundamental. Nor are we in a position to change the teleology of the term or concept. Most of the facts that figure into the interpretation function for *C* are not in our control, even in principle. So ameliorating our concept is metasemantically impossible. (Note that this is about more than just getting broad social uptake.)

We can summarize this argument as follows:

The Metasemantic Control Challenge (MCC)

1. Conceptual amelioration involves altering either the intension of a semantic unit, the extension of a semantic unit, or both.
2. The intension and extension of semantic units are fixed by the interpretation function plus the inputs to that function.
3. It is not in our power to change the interpretation function (viz. metasemantic facts are not in our control).
4. It is not in our power to change enough of the inputs into the interpretation function to significantly and purposely change the outputs of the interpretation function (viz. The intension or the extension of the semantic unit).

So,

¹³ Evans (1973).

5. Conceptual amelioration is impossible.

Premise 1 is simply a restatement of what conceptual engineers themselves seem to be up to.¹⁴ Premise 2 is, I take it, true by definition given how “interpretation function” is being used here. Premise 3 is a bit trickier, since the metaphysical status of metasemantic facts has, to my knowledge, almost never been discussed. I won’t defend this in detail here for reasons of space, but I do think something like premise 3 is presupposed in the (non-conceptual engineering) metasemantic literature. Riggs (2019) and Thomasson (2021), both working within the conceptual engineering literature, reject 3; however, because their rejection of 3 only succeeds by way of rejecting premise 4, I set it aside.¹⁵ Premise 4 has already been defended above.

2.2 Making Sense of the Method

We are engaged in the project of providing a vindictory metasemantics for conceptual engineering. An often neglected aspect of doing this is making sense of exactly *how* conceptual engineers engage in their projects. If our response to the MCC suggests a methodology for conceptual engineering which is wildly divergent from the methodology used in the actual literature, this would at best vindicate some other kind of conceptual ameliorative methodology, not conceptual engineering as practiced by Carnap, Haslanger, or the many who have followed in their footsteps. As far as I know, the only person to make this point explicitly in the literature is Derek Ball:

[I]n general, when a debate seems unconfused, we should try to understand what the participants in the debate are doing in a way that makes them unconfused. In this kind of case, we want to be able to say that the parties to a debate are rational, that their contributions to the debate make sense and contribute in a recognizable way to their aims.¹⁶

In a way, this desideratum on a theory of conceptual engineering is the most important. Only by looking at the first-order methodology of conceptual engineers can we get a sense of the philosophical presuppositions their practice is making. Once we have those in hand, we can ask what is the most plausible (in our case, metasemantic) theory that can vindicate those presuppositions, thus (conditionally) vindicating the practice as a whole.

The practice of conceptual engineers shows us, I think, two important facts that a theory of conceptual engineering must capture. First, conceptual engineers, virtually without exception, defend their amelioration proposals by way of *normative* premises: Epistemic, but also moral and socio-political premises as well. A theory of conceptual

¹⁴ See, e.g. Isaac et al. (2022) and the references therein.

¹⁵ Both Riggs (2019) and Thomasson (2021) aim to reject premise 3 by conceptual engineering on semantic units that refer to metasemantic features (“concept”, “metasemantics”, “reference”), which requires independent reason to think that premise 4 is false. See section 3.1 below for discussion.

¹⁶ Ball (2020b), p.40.

engineering needs to be responsive to this widespread fact: How can these normative premises be understood as related to the metasemantics of amelioration? A corollary of this is that an adequate account of conceptual engineering must capture *the way* that those normative premises operate in conceptual engineering discourse. I'll return to this latter point in the next section.

Second, conceptual engineers aim at a range of ameliorative targets. First, there is the distinction between normative improvements in concepts as opposed to normative improvements in our conceptions. The distinction between concepts and conceptions is clearly articulated by Sarah Sawyer:

Concepts are mental representations that are constituents of thoughts. Conceptions, in contrast, are sets of beliefs. Specifically, the set of beliefs a subject associates with a concept is her conception of the subject matter that her concept represents.¹⁷

Even beyond these two, conceptual engineers also engage in ameliorative proposals regarding the *function* of a concept. Consider the historical function of MARRIAGE. It is not at all implausible to think that the historical institution had an inherently oppressive function. But one may still want to preserve the concept of MARRIAGE, albeit with a different and normatively permissible function. (We could read Elizabeth Brake's notion of "minimal marriage" in this light, for example.¹⁸) A fully vindicatory theory of conceptual engineering should be able to accommodate this flexibility in ameliorative targets.

2.3 Ameliorative Reports and The Structure of Ameliorative Arguments

It is a general condition on a semantic theory that it can capture the ways that competent speakers converse using terms to express the concepts that they have a grasp of. This is of course not to say that speakers are infallible, because they can be ignorant or misinformed. And of course speakers *can* utter sentences that are ungrammatical or internally incoherent. But it is a mark against a semantic theory if it entails that speakers do this systematically and pervasively.

This condition relates to our current question because of a particular kind of utterance that ordinary speakers do tend to make. Here are a few examples:

"I once believed that same-sex marriage was impossible (and said things like, "A man can only marry a woman and vice-versa"), but I was convinced...that same-sex marriage is possible after all."¹⁹

We used to think of race as a biological concept, but now we understand that it is actually a social one.²⁰

¹⁷ Sawyer (2020), p.6.

¹⁸ Brake (2012).

¹⁹ Ball (2020), 52.

²⁰ See, e.g. Benoist (2021).

Prior to the work of Carolus Linnaeus, it was known that whales had live birth, breathed air, and had other characteristics similar to land mammals. But nonetheless, the scientists thought they were fish; Linnaeus provided the classification system that showed that this was a mistake.²¹

Call these *Ameliorative Reports*. If such reports are felicitous (as their prevalence suggests that they are) and taken at face value, this constrains adequate theories of conceptual engineering. If, as in Ball's case, people take themselves to have changed their mind *about marriage*, and not simply changed which word they use when they say "marriage", we need an account of amelioration which captures this.

By themselves, the existence of Ameliorative Reports does not require sameness of meaning over time.²² We can imagine speakers asserting such sentences as elliptical for some more complicated metalinguistic claim. For example, imagine someone saying "I used to think that Moroccan food was spicy, but after traveling in Mexico, I realize that Moroccan food isn't spicy at all." They haven't really changed their descriptive beliefs about the absolute spiciness of Moroccan food, but have readjusted their threshold for what counts as spicy.

However, and returning to the methodological point made above, Ameliorative Reports are even more probative when combined with another feature of conceptual engineering discourse. Compare someone being convinced by Haslanger's arguments in favor of her ameliorative definition of WOMAN with someone adopting some other ameliorative definition of WOMAN because, say, Elon Musk paid them \$50 to do so. Both of these motivations involve some sense of responding to a reason that one has for adopting a new definition. However, unlike the Haslanger case, the Musk case appears to be the *wrong kind of reason* to adopt an ameliorative concept of WOMAN. So what does it take for an ameliorative concept to be adopted for the 'right' kind of reason? It looks as though, as in the Haslanger case, the agents and institutions see their changes as being *reason-responsive*: The amelioration of the concept is adopted on the basis of concept-centered normative reasoning, as opposed to boredom, bribery, or some other 'wrong kind of reason'.²³ As Viktoria Knoll says, "[w]hen involved in ameliorative disputes, we see ourselves responding to reasons that are good and convincing independently of whether somebody (ourselves included) will eventually acknowledge them as such."²⁴ We can see this in the structure of many

²¹ Often this claim is framed much more simply: "We used to think whales were fish, but now we know that they are not." - This could also be an ameliorative report. However, framed in this way, the claim looks more like the fact that whales are not fish was a straightforward empirical discovery. Things were much more complicated, in reality. For a detailed history, see Romero (2012).

²² Thanks to Caleb Perl, Joshua Schechter, and Mark Schroeder for pressing me on this point.

²³ Eklund (forthcoming, Section 5), briefly discusses the way in which engineering projects can run into "wrong kind of reasons" problems.

²⁴ Knoll (2020), p.25.

ameliorative reports (“I was convinced”; “now we understand”, etc.). This distinction between reasons for accepting an amelioration constrains an adequate theory of conceptual engineering: Such an account must make sense of the ways in which conceptual engineering aims to be *reason-responsive*.

2.4 Summing up the Criteria

Let’s sum up the three conditions on an adequate metasemantic theory of conceptual engineering:

Metasemantic. Conceptual engineering is only a viable method if we can explain what has gone wrong with the MCC; this requires a metasemantics for conceptual engineering that motivates rejection of one of the premises of the argument.

Methodological. The methodology of conceptual engineers and ameliorators involves rational argumentation which appeals to normative (epistemic, moral, and socio-political) premises, can target concepts, conceptions, and the function of concepts.

Semantic. A viable theory of conceptual engineering must allow for the felicitousness of Ameliorative Reports, as well as the right kinds of reasons which lead people to endorse an ameliorative proposal.

3. Interest-Relative Metasemantics: Ameliorative Pragmatism

My aim in this paper is to develop and defend a metasemantic theory that requires normative truths - construed in a minimally realistic sense - as playing a role in the metasemantic interpretation function. But first, it's worth considering a natural and less committal alternative, which I'll call *Ameliorative Pragmatism*. Ameliorative pragmatists claim that there are no objective constraints on conceptual amelioration. Pragmatists will admit, of course, that conceptual engineering presupposes that “there are at least some standards for evaluating concepts”.²⁵ However, the pragmatist allows such standards to be parochial. As long as there is significant overlap in shared goals, the direction of our ameliorative projects can be *interest-relative* without being arbitrary.

The pragmatist view is attractive because it is minimally committal but also prima facie plausible. It also aligns with some of the ways that conceptual engineers frame their projects.²⁶ However, the view faces serious problems. To see this, first return to the MCC. The most natural strategy for the ameliorative pragmatist to take in responding to this argument is to deny premise 4:

4. It is not in our power to change enough of the inputs into the interpretation function to significantly and purposely change the outputs of the interpretation function (viz. The intension or the extension of the semantic unit).

²⁵ Thomasson (2020) 437.

²⁶ Haslanger (2006) - but see also Haslanger (2005), where she appears to be more open to a realist interpretation harnessing semantic externalism and semantic ignorance.

As a first pass, the ameliorative pragmatist must reject 4 since she explicitly allows “our” political and theoretical goals to shape the ameliorated meaning of our concepts. For the ameliorative pragmatist, sometimes meaning change will explicitly require going against the historical use and function of a term since that historical use and function is morally, politically, or theoretically problematic. This appears to require a very unorthodox metasemantic theory, and it is initially hard to see why it would not run up against the kinds of serious problems that caused people to move away from simple internalist metasemantic theories in the time of the externalist turn in the 1960s and 1970s.²⁷

Perhaps the most attractive approach for the Ameliorative Pragmatist is what I’ll call the *Homonym Interpretation*. The Homonym Interpretation rejects premise 1, because it denies that conceptual amelioration involves any conceptual change. Instead, “conceptual amelioration” is a misnomer: What really occurs in an ameliorative project is the coining and baptism of a new concept which uses a lexically identical symbol as some previously existing term. Since the kind of metasemantic control required for baptizing a new term is, even on externalist metasemantics, quite easy, the MCC can be met. Furthermore, since there are few constraints on the content of newly baptized terms (one can in principle introduce a name for any reference-fixing description whatsoever), we retain the freedom of the pragmatist theory as well.

The homonym interpretation is plausible for some cases of conceptual engineering. For example, in naive set theory, sets were defined to obey the principle of unrestricted composition: For any well-defined property, there will be a set of all and only the objects that have that property. Russell famously showed that using this principle to define sets leads to paradox. This led to multiple reworkings of how to define sets to restrict how they can be composed. The homonym interpretation is plausible here because, as with many concepts in mathematics, we are dealing with a stipulative definition rather than fixing a referent in a more standard way (such as by ostension). When one stipulative definition is seen to lead to paradox or other problems, mathematicians can just generate another stipulative definition, and continue using the same term in terms of its role in the theory and family resemblance.²⁸

Stepping outside of math, consider a trickier case: the Carnapian explication of HARDNESS. HARDNESS has a folk conception with which we are all familiar. However, materials science has long had an operationalized and formal definition of HARDNESS in

²⁷ Barcan Marcus (1961), Putnam (1975), Kripke (1980).

²⁸ This is one important theme of Lakatos (1973). I thank Joshua Schechter and Christopher Gibilisco for their discussion on this point.

terms of the Mohs scale.²⁹ On the one hand, it is very plausible to interpret the folk use of *HARDNESS* as univocal with the experts, but lacking information, much as in Burge's famous *ARTHRITIS* case. The homonym interpretation, on the other hand, would say that there are actually distinct *HARDNESS* concepts, one for the folk conception and another for the materials science conception. Although I'm inclined to hear this as more naturally a version of Burge's *ARTHRITIS* case, and thus incompatible with the Homonym Interpretation, I don't think semantic intuitions are strong enough in either direction for this kind of case to be conclusive.

On the other hand, the homonym interpretation has trouble making sense of the felicitousness and structure of Ameliorative Reports. Consider again Ball's example:

"I once believed that same-sex marriage was impossible (and said things like, "A man can only marry a woman and vice-versa"), but I was convinced...that same-sex marriage is possible after all."

On the Homonym Interpretation, such a speaker is deeply conceptually confused, as though they did not realize that there are two *MARRIAGE* concepts at work. More generally, two people arguing over how to think about marriage are, unbeknownst to them, having a metalinguistic dispute about how to use the word "marriage". The fact that people can have a metalinguistic dispute without knowing it is a cost for the Homonym Interpretation. However, there is a growing literature on metalinguistic negotiation which aims to explain how this can occur even when speakers would reject that their dispute is metalinguistic.³⁰ Plunkett & Sundell have pointed out that speakers, for example, often commit errors distinguishing between the semantics and pragmatics of their utterances, between what is expressed and what is implicated by a given assertion. On the other hand, as Joanna Odrowąż-Sypniewska has pointed out, this error appears to occur at the margins of the semantic/pragmatic distinction, and there is little precedent for the expansive amount of error that Plunkett & Sundell's account implies.³¹ I'm inclined to side with Odrowąż-Sypniewska (and others³²) that the metalinguistic negotiation interpretation entails an implausible amount of error to reflective and competent speakers, but I won't attempt to settle that dispute here. A further problem here is that it is not clear how the Homonym Interpretation, even supplemented by the notion of metalinguistic negotiation,

²⁹ I am simplifying here; there are actually multiple precise ways of measuring hardness of a material, with conversion scales between them.

³⁰ Plunkett & Sundell (2013, 2021), Plunkett (2015).

³¹ Odrowąż-Sypniewska (2023), pp.1020-1021.

³² Marques (2017), Abreu (2023).

can make the distinction between what ameliorators take as the right and wrong kinds of reasons for a particular interpretation.

I want to be clear that I don't think that these problems for the Homonym Interpretation make it a non-starter. However, all of these considerations, taken together, at least puts pressure on the conceptual engineer to assess whether there are other theoretical possibilities for the metaseantics of conceptual engineering, ideally with fewer problems and complications. This, I think, is enough to motivate looking to a metaseantic view which incorporates substantively normative facts.

4. Substantively Normative Metaseantics

4.1. Initial Motivations

Because it will help point us in the right direction, let's pause to take stock. We are looking for a vindication of conceptual engineering that is on the one hand metaseantically plausible, and on the other fits with how conceptual engineers engage in the practice. The former condition is extremely difficult to meet because it appears as though neither the interpretation function, nor the inputs into it, are sufficiently under our control for engineering projects to be possible. The key to the second condition is to notice that conceptual amelioration projects, as varied as they otherwise are, invariably appeal to normative premises, either epistemic, moral, or socio-political. Conceptual engineering is an inherently normative project. Taken together, we saw how these desiderata make ameliorative pragmatism, and in particular, the Homonym Interpretation, an initially attractive option. However, I've argued that the Homonym Interpretation fails, at least for many cases of conceptual engineering.

As we saw, the semantic data underlies two lessons about the kind of account we are looking for. First, we need to make sense of agents who change their own conception of the concept they are using: In ameliorative reports, agents, or institutions, begin to use concepts in different ways than they had before, but they nonetheless see themselves as using the same concept they were before. Second, the agents and institutions, at least at the center of such reports, see their changes as being *reason-responsive*. And notice that ameliorative reports can occur because of two kinds of normative changes in belief - They may arise as a result of a change in belief about how best to advance one's current normative aims, but they may also arise as a result of a change in belief about *which aims to have*. This is important since it shows that an interest-relative approach to explaining ameliorative reports will not generalize.

The answer that suggests itself at this point is clear: We need a metaseantic theory that incorporates substantive normative facts. Only such a theory, I argue, can explain the

data at each level of theory that we've discussed up until this point. Of course, there is also the most basic consideration, of showing that the theory is *prima facie* plausible. I suggest the best way to demonstrate that is to develop the theory, so I turn to that first.

4.2 Two Approaches to Normative-Metasemantic Entanglement

We can state the general thesis that I claim is necessary for an adequate account of conceptual engineering as follows:

Normative-Metasemantic Entanglement (NME): The outputs of the interpretation function are in part determined by substantive normative facts (including epistemic, moral, and socio-political normative facts).

The challenge is to provide an independently plausible metasemantic theory that vindicates NME. I will discuss two possible approaches: J.R.G. Williams' defense of normative reference magnets, and Schroeter & Schroeter's *Connectedness* model, supplemented with normative idealization.

4.2.1 Normative Reference Magnets

J.R.G. Williams' Radical Interpretivism, with its incorporation of *Normative Reference Magnets*, is ready-made to undergird a theory of conceptual engineering that captures the criteria mentioned above. As with others in the interpretivist tradition,³³ Williams argues that representational content is fixed by rational norms of interpretation. That is, the content of an agent's beliefs and desires are grounded in the most reason-responsive possible interpretation of her actual and counterfactual behavior. But there are two novel aspects to Williams' account which are connected to the questions raised in this paper.

First, Williams argues that radical interpretation explains reference magnetism. This is novel because previous interpretivists took reference magnets to be an added ornament to the theory that helped it to avoid objections; Williams instead argues that it follows straightforwardly from a proper understanding of the radical interpretivist story.³⁴

A second novel aspect of Williams' theory involves the question of what feeds into the norms of interpretation. According to traditional interpretivism, assigning content to an agent's mental states is a matter of what best rationalizes her (actual and counterfactual) behavior, where the rationalizing process here is cashed out in terms of epistemic norms plus Humean, means/ends reasoning to action.³⁵ Williams argues that moral norms, as action-guiding reasons, must be incorporated into the interpretation of mental states.³⁶ So Williams' view is that the norms of interpretation should be given in terms of *substantive* rationality, including substantive epistemic norms as well as substantive moral norms.

³³ Davidson (1973, 1994), Lewis (1974), Mölder (2010).

³⁴ Williams (2020, Ch.3).

³⁵ Davidson's (1974) view is more complex than this, but it does not incorporate moral norms.

³⁶ Williams (2020, Ch.4)

From these two novel features of Williams' view, we get an independently plausible argument for *Normative Reference Magnets*: Concepts' semantic content are fixed in part in terms of the eligibility of referents (certain properties *magnetize* reference), and part of what determines a property is more eligible is that it is *normatively*, including *morally*, *better* for that concept to pick out one property rather than another.

To see how this works, let's walk through how Williams' theory would make sense of Haslanger's amelioration of race concepts. Here is the schema for Haslanger's ameliorative analysis of BEING-RACIALIZED:

"A group is racialized if and only if its members are socially positioned as subordinate or privileged along some dimension (economic, political, legal, social, etc.), and the group is "marked" as a target for this treatment by observed or imagined bodily features presumed to be evidence of ancestral links to a certain geographical region."³⁷

Assuming this analysis picks out an eligible referent for the concept of *being racialized*. There will be other eligible referents - perhaps biological, but also alternative social analyses of BEING-RACIALIZED. There will also be ineligible referents (race concepts could not pick out features related to *being-an-electron*, for example). Among the remaining eligible referents, use will underdetermine referent. It is here that normative reference magnets - epistemic, political, and moral - kick in. Our interpretation of people's race concepts, then, will map on to the substantively normatively best interpretation of the function that those concepts play in our cognitive lives. This makes it the case that Haslanger's analysis is correct iff it is the normatively best of the eligible referents of race concepts. And whether that is true or not depends *in part* on normative argumentation.

Williams' view is an attractive place to point when looking for a metasemantics of conceptual engineering, since it is motivated by methodological and semantic principles far removed from solving the problems we are concerned with here. However, interpretivism and radical interpretation always have been, and continue to be, contentious (albeit plausible) theories of reference-fixing. So it's worth mentioning that the theory of conceptual engineering I am defending here does not require a commitment to Williams' interpretivism. While it is true that Williams provides us with one compelling route to the conclusion that such *Normative Reference Magnets* exist, there may be others.

4.2.2 Connectedness + Normative Idealization

In a series of articles, Laura Schroeter and François Schroeter have developed an account of co-reference that allows for changes in concept's conception and meaning over time. Schroeter & Schroeter are interested in providing a theory of co-reference that can

³⁷ Haslanger (2012, 251).

make sense of a thinker's ability to retain the same concept over radical changes in how they conceive of the referent of that concept. For example, one may retain the same subjective concept of WRONGNESS despite cycling through belief in any number of different metaethical and normative theories. This ability to retain the same concept over time and change is important for individuals' abilities to respond to new evidence without reinventing their conceptual schema from scratch with every substantive new piece of information.

Schroeter & Schroeter endorse the rough schema for co-reference:

Connectedness model: two token elements of thought express the same concept only if they are connected to each other in the right way.³⁸

A helpful analogy for this model (from Schroeter & Schroeter themselves) is to consider a stage theoretic model of personal identity: On such a view, a person is made up of different 3D time slices. What makes one set of time slices the same person and what makes other spatiotemporal slices not the same person does not involve some intrinsic feature of each person-stage, but rather arises out of the relationship of the person-stages to each other. Similarly, then, the Connectedness Model of co-reference says that two token instances of concept use are tokens of the same concept just in case they are connected to each other - either directly or via other tokens - in the right sort of way.

The challenge here is in fleshing out "the right sort of way" for ensuring concept-identity over time. Roughly, the idea is that a concept stays the same over time if it remains part of the same "representational tradition". Being part of the same "representational tradition" requires something like the following: Between any two pairwise tokenings of the concept, a rational subject would see these tokenings as being tokenings that have the *de jure* sameness of topic. So, for example, if I am thinking through some modus ponens inference which tokens concept *c* in premise one and in the conclusion, I must see these instances of *c* as being about the same thing (whether or not I have gained new information about *c* or its referent between premise 1 and the conclusion). This will also be true intersubjectively, if I am in conversation with you and we (rationally) use some term throughout that we take to be referring to the same concept, despite, perhaps, disagreeing about certain facts about the concept or its reference. A "representational tradition" is the recursive function that applies to the series of these interactions (both inter and intra-subjective) over time.

So far so good. However, we still don't have a theory about the interpretation function. Their answer appeals to reference judgments under epistemically ideal conditions:

³⁸ Schroeter & Schroeter (2014), 8.

Ideal Accessibility (IA): The correct semantic interpretation of the referential concept expressed by a term 'X' must make the subject's ideal, empirically-informed beliefs about which things are X in any possible world come out true.³⁹

While IA as stated is a necessary condition on reference determination, and thus compatible with a wide range of views, Schroeter & Schroeter are clear elsewhere that they want something like IA to be a *truth-maker* for outputs of the interpretation function.⁴⁰

If you're an ethicist, you may find this structure of explanation familiar from ideal advisor theories of reasons and well-being.⁴¹ On such theories, what reasons we have are determined by what some version of ourselves, properly idealized, would advise us to desire. These idealized desires must trace back to our actual desires as their basis. Similarly, in Schroeter & Schroeter's theory, we need some basis in our actual selves to explain our idealized judgments of reference. Furthermore, we need some account of *how* we are idealized, and such an account must not generate circularity.

With respect to the idealization base, Schroeter & Schroeter have already provided a sketch of an answer: The representational tradition's actual judgments about whether to place given entities within or outside of the scope of the concept. For the idealization process, Schroeter & Schroeter turn to a proposal from David Chalmers according to which the idealized subject is omniscient with respect to (a) the microphysical facts, (b) the indexical facts, and (c) a "that's all" clause. Call these the *scrutability base*. This can't be a wholly general theory of the interpretation function, since it would result in circularity for reference to the microphysical entities and indexical facts. But we can set that aside since an exhaustive metasemantic theory isn't what we are after here.

With this sketch, we can turn to see how to couple the Connectedness Model with Normative-Metasemantic Entanglement. We have two options here, depending on one's metaethical commitments. On the one hand, on a naturalist conception of the normative facts, the normative facts would be contained in/entailed by the scrutability base, so the normative facts could already play a role in the interpretation function. On the other hand, on a non-naturalist view, we would have independent reasons to add the basic normative facts to the scrutability base.⁴² Either way, then, we are in a position to accommodate NME.

This doesn't ensure that the normative facts would influence the outputs of the interpretation function in the right sort of way, since this depends on what idealized speakers would select as the reference for a wide variety of concepts. However, this is enough - for at this point, we can just point to the way conceptual engineering is practiced

³⁹ Schroeter & Schroeter (2019), 200.

⁴⁰ See, e.g. Schroeter & Schroeter (2014), 13-14.

⁴¹ See, for example, Railton (1986).

⁴² See Chalmers (2012), p.264-267 for discussion.

(by theorists and by ordinary speakers, deferentially or not) as evidence that the normative facts *do* make a difference. If an agent changes how she assesses the intension and extension of a term in light of normative argumentation, and no further idealization would undermine this change, then it follows that normative facts *do* play a role in the interpretation function. There is no circularity here: The criteria for conceptual engineering motivated NME, but the motivations for this version of the Connectedness Model are independent from the motivations for NME.

5. Normative-Metasemantic Entanglement and The Three Criteria

I've now discussed two independently plausible metasemantic theories which can capture *Normative-Metasemantic Entanglement*. Now it's time to wrap everything up into a bow by showing how views with this structure can meet the three criteria on a theory of conceptual engineering. (Unfortunately, we will have to ruin the elegant wrapping job by introducing a few remaining questions, complications, and upshots, in sections 6 and 7.) Here, again, are the three features we want a theory of conceptual engineering to explain:

Methodological. The methodology of conceptual engineers and ameliorators involves rational argumentation which appeals to normative (epistemic, moral, and socio-political) premises, and can target concepts, conceptions, and the function of concepts.

Metasemantic. Conceptual engineering is only a viable method if we can explain what has gone wrong with the MCC; this requires a metasemantics for conceptual engineering that motivates rejection of one of the premises of the argument.

Semantic. A viable theory of conceptual engineering must allow for the felicitousness of Ameliorative Reports, as well as the right kinds of reasons which lead people to endorse an ameliorative proposal.

Let's take each in turn.

Begin with *Methodology*. It should be clear how Normative-Metasemantic Entanglement can explain why ameliorative projects proceed in part via normative premises: The normative facts in question play a role in determining the outputs of the interpretation function, so appealing to them is directly relevant to understanding the referent(s) of the concept in question. Of course, such appeals are not without their limits: normative features are not, even on the versions of NME discussed above, the sole determinants of the referent of a concept. As Parfit says, "rivers could not be sonnets",⁴³ and that would be true even if it would be really wonderful if they could. However, we can interpret conceptual engineering as the attempt to get a better sense of the normative aspect of reference-fixing for some given concept or set of concepts.

⁴³ Parfit (2011 Vol.2:325-7).

Next consider *Metasemantic*. We haven't yet seen how the proponent of NME can respond to the MCC. However, we do have a template in a theory of conceptual engineering that I'll call *Ameliorative Realism*, most prominently defended by Sarah Sawyer (2020).⁴⁴ Sawyer, following some remarks from Sally Haslanger, aims to defend a more conservative reading of conceptual engineering, according to which it really is just a variation of traditional conceptual analysis. On Sawyer's view, conceptual engineering involves ameliorating our *conceptions*, rather than the concepts (or their content) themselves.

Sawyer, and ameliorative realists more generally, respond to the MCC by denying premise 1:

1. Conceptual amelioration involves altering either the intension of a semantic unit, the extension of a semantic unit, or both.

Conceptual amelioration, for the ameliorative realist, is a misnomer. The intension and extension of concepts remains the same after amelioration as it was before; only the *conception* has changed. So there is no fear of topic change, and no need to exert metasemantic control.

Ameliorative realism, as defended by Sawyer, represents a very restricted notion of conceptual engineering for two reasons. First, it limits the normative claims in amelioration to those with joint-carving and non-normative pattern implications. It can't obviously explain the amelioration of MARRIAGE, for example, or many other such socially constructed categories. Second, it limits the amelioration to conceptions, not allowing for amelioration in intensional or extensional content of a concept. This violates the last clause of the *Methodological* condition.

However, endorsing Normative-Metasemantic Entanglement allows us to provide the same kind of response to the MCC without the limitations. On the NME view, conceptual engineers use normative premises in their arguments for amelioration as a way of defending a particular theory of the referent of some concept (or set of concepts). Given NME, the normative facts can help to guide the reference of a concept whether or not that concept can play some other non-normative, joint-carving role (though it often will, since epistemic normativity is presumably connected to joint-carving distinctions). And NME will allow for this to occur at the level of semantic/conceptual content as well as at the level of conceptual functions, not merely at the level of conceptions. So it can also capture this aspect of the *Methodological* condition. The response to the MCC, then, is to reject premise 1: The MCC relies on a misinterpretation of what conceptual engineers are up to.

⁴⁴ But see also Haslanger (2012, 2017).

To see why the rejection of premise 1 is acceptable, let's remind ourselves of what motivated it in the first place: Premise 1 was an attempt to properly characterize what conceptual engineers seem to be up to when they engage in conceptual amelioration. It's true that the version of ameliorative realism + NME that I've proposed here does not capture *all* of the things that conceptual engineers report that they are doing. However, what is important, I think, is that it captures the core of the practice: Providing mixed normative/descriptive arguments in favor of an improvement in the content of our terms/concepts. A slight revision in how we *describe* what conceptual engineers are up to should be acceptable to them, given that we don't undercut the methodology itself.⁴⁵

Finally, consider the *Semantic* data that needs to be explained; we need a theory of conceptual engineering that can make sense of the felicitousness of Ameliorative Reports and the discourse in which they arise. Ameliorative Reports suggest that speakers take themselves to be using the same concept both prior to and after some accepted ameliorative proposal. And the felicitousness of such reports suggests that they are not mistaken. A metasemantic view that incorporates NME can explain this straightforwardly: The concept has *not* changed between t_0 (before the amelioration was accepted) and t_1 , after. This makes perfect sense of (and will generally render true) Ameliorative Reports. It's worth noting the relationship between Ameliorative Reports and the MCC. The prevalence of Ameliorative Reports actually provides independent, if defeasible, evidence against premise 1 of the MCC. Imagine Miriam reads Clark & Chalmers' (1998) paper on the extended mind, and changes her mind about what constitutes a belief on the basis of normative considerations. She utters "I used to think that beliefs had to be inside the head, but I was convinced by Clark & Chalmers that things that count as BELIEF can also apply to information I have logged in a notebook or in my cell phone." Miriam is speaking precisely as though the concept did not change; rather, her belief about the reference of that concept changed. This is exactly what a metasemantic theory with NME predicts.

6. Shared Content and Metanormative Commitments

I have argued that conceptual engineers must commit to Normative-Metasemantic Entanglement in order to defend the variety and method of ameliorative projects as well as the metasemantic possibility of conceptual amelioration. I have also sketched two independently plausible metasemantic views that endorse NME. This would be the part of the paper where it would be nice to give a clear and precise statement of the minimal metaethical commitments required for deeply normative metasemantics of the sort that ameliorative projects presuppose. Unfortunately, I can't give a univocal answer here for

⁴⁵ Thanks to an anonymous referee for pressing me to clarify this point.

principled reasons: An answer depends on where we fall on certain background commitments in semantic and metasemantic theory. Let me explain.

It's an open question what the essential explanatory goals of a semantic theory are, but here are three widely accepted (non-exclusive) candidates:

Communication A theory of semantic content should be able to explain the possibility - and indeed, widespreadness - of successful communication between competent speakers and between the same thinker using the same concept over time.⁴⁶

Disagreement. Relatedly, such a theory should be able to explain disagreement over some proposition between agents or within the same agent over time, and differentiate between genuine disagreement and merely talking past each other.

Psychological Generalizations. A theory of semantic content should explain and predict agential behavior, ideally in a way that generalizes over and across psychologically typical agents.

Each condition pushes us in the direction of semantic content which is *public*, shared across competent speakers, over time. There are also powerful arguments that in many cases these conditions should be met even between speakers of different languages, thus requiring shared mental content.⁴⁷

If these arguments are compelling, we will want a metasemantic theory that ensures a large amount of shared semantic (mental) content across individuals with a variety of backgrounds, beliefs, ideologies, and knowledge. Coupled with NME, this gives us:

Broad Intersubjectivity: The same fundamental normative facts are true for all individuals regardless of their particular spatio-temporal location, background, beliefs, ideologies, and knowledge.

Broad Intersubjectivity is compatible with a number of metaethical views, from robust non-naturalism, to naturalist realism, or even to a broadly enough construed constructivism. There may even be an expressivist or inferentialist reading of it.⁴⁸

Nonetheless, there is slack in the explanatory goals of a semantic theory, and a number of ways theorists have tried to meet them without endorsing such a broad commitment to shared content.⁴⁹ The details here would take us very far afield, but these arguments are varied in their targets as well as their scope. Nonetheless, such alternatives tend to cut around the margins of shared semantic content, rather than deny it altogether.⁵⁰

⁴⁶ "Speakers, more often than not, succeed in mutual comprehension despite a diversity of beliefs, intentions, interests, goals, audiences, conversational contexts, and perceptual inputs." Cappelen & Lepore (2006), section 2.

⁴⁷ Edwards (2014).

⁴⁸ See Price (2013), Barker (2020).

⁴⁹ For example, Begby (2013), Belleri (2017).

⁵⁰ The considerations in favor of shared content are much more numerous than I have listed in the text, and that is part of the reason why there are very few wholesale rejections of it. See Cappelen & Lepore (2006), section 5, for a sustained defense of the importance of capturing shared content.

How this restricts *Broad Intersubjectivity* will depend on the details of exactly how this goes. As a rule of thumb, what can be said is that the greater the commitment to shared content across individuals, *especially* with respect to individuals with varying normative commitments, the broader the commitment to intersubjective normative facts for a theory that endorses NME.

What a bare commitment to NME entails for metaethics, then, depends on other questions about the degree of publicity in semantic (mental) content. But we can say more concrete things about each of the specific views discussed above - Williams' Normative Reference Magnets and the Schroeter & Schroeter inspired Connectedness + Normative Idealization model. Williams' route to Normative Reference Magnets rests crucially on moral twin earth intuitions of disagreement: When the citizens of Utilitas and citizens of Kantopia make varied claims about which actions exhibit WRONGNESS, we take them to be in genuine disagreement about what actions are wrong, not merely talking past each other.⁵¹ I don't want to rehash the debates about moral twin earth here, but Williams sees these cases as supporting *Referential Stability* for the basic normative concepts:

Referential Stability (Schematic) Necessarily, if an agent has a concept W that plays role R, then W denotes P.⁵²

If Referential Stability holds, any possible thinker with any thin normative concepts refers to the same (normative) properties. This suggests that the normative properties are fixed stance-independently, as the robust non-naturalist realist holds. Williams' view appears to commit us to much more than Broad Intersubjectivity, since he appears to be committed to stance-independent normative facts playing an essential role in the interpretation function.⁵³

Turn now to the Connectedness + Normative Idealization model. Recall that on this model, the output of the interpretation function for some concept is determined by users of the term's properly idealized judgments over that concept's "representational tradition". This idealization will contain full information about the distribution of the use of the term over time, as well as the normative facts (directly and indirectly). What, then, must these normative facts be like to play this role in the interpretation function? A key feature of the Connectedness model is that co-reference of some concept C is explanatorily prior to the facts about the reference of C.⁵⁴ If MARRIAGE in 1950 is part of the same representational

⁵¹ The original statement of moral twin earth is Horgan & Timmons (1992). Williams (2019), Sect. 4.2.

⁵² Williams (2019, 78). But Williams does slightly weaken this schematic, see sect. 4.4. Notice also the similarity here with Eklund's idea of *referential normativity*. See Eklund (2017), Sect. 3.1.

⁵³ Williams seems to be committed to this independently of anything I've said in this paper, though as far as I know this hasn't been noted in the literature.

⁵⁴ As Schroeter & Schroeter (2014, 13) point out, this is also true of many versions of externalism going all the way back to Kripke.

tradition as MARRIAGE today, these are the same concept. Once we've established co-reference in this way, and barring semantic forking, we need our interpretation function to generate the same output for all speakers within the representational tradition, whatever their particular idiosyncratic beliefs may be.

The normative facts, or at least those relevant to the output of the interpretation of some representational tradition, must be shared by all individuals across the tradition, despite all of their subjective differences. Given that some representational traditions will be extremely wide and long standing, this will entail a quite universal conception of the normative facts. However, unlike Williams' view, it does not require a particular metaphysical conception of those facts. The facts neither need to be necessary, nor do they need to be radically stance-independent or non-natural. We only need them to be universal enough to explain how the normative facts are compatible with co-reference across any individual who shares any representational tradition. This looks much more like the minimal, constructivist compatible, *Broad Intersubjectivity*. Depending on one's other commitments, the fact that this view is more ecumenical than Williams' may be seen as a defeasible reason in its favor.

7. An Objection, a Response, and an Implication

Before concluding, I want to make two more brief remarks about the view I've defended here.

First, the claim of normative-metaseantic entanglement that I've defended above results in a view whereby normative factors can help to determine which of multiple eligible referents a concept has. One might worry that this has the result of making inherently harmful or oppressive concepts impossible - a result that might be nice but is otherwise implausible. However, this would be to misunderstand the strength of normative influence in metaseantic determination relations. We can see this by analogy with reference magnetism as traditionally construed. Reference magnetism does not make reference to wildly gerrymandered properties impossible: It merely serves as a guide once various eligible referents are already determined by basic metaseantic principles. Any metaseantic theory which includes reference magnets must also contain other components, and these other components may by themselves rule out any reasonably joint-carving referent.⁵⁵ Similarly, any metaseantic theory which includes normative entanglement (such as the two discussed above) must also contain other components, and these other components may rule out any reasonably normatively good referent or function.

⁵⁵ See Sider (2013), 3.2, Dunaway (2020), 110.

Second, notice that the theory defended here allows for stepwise amelioration. Suppose traffic engineers announced tomorrow that they had determined the ideal set of traffic laws and traffic signal coordination to minimize commute times and safety. However, the set is wildly counterintuitive, and involves several new signals that drivers have no familiarity with. In such a case, the best path forward would be to institute the changes gradually, and perhaps even in some cases not at all (if the payoff is small and the complication cost is large). We need to separate the question of what the *ideal* system is from the question of *what is the best improvement given our current non-ideal starting point*. So too will be the case with conceptual amelioration. Put in terms of normative reference magnets, it may be that the normatively ideal referent of a concept may at some earlier time not be an eligible referent, though it may become one at some later time (just as with the traffic rules, once earlier ones have been implemented).

Interestingly, this possibility of stepwise amelioration represents another advantage of the theory defended here over other ameliorative realists, such as Sawyer, or Derek Ball's temporal externalism.⁵⁶ On Ball's view, for example, any future ameliorative meaning reverberates back throughout all historical uses of the term. Now suppose we have some concept C that undergoes two separate ameliorations over a period of time, a1 and a2. Once the second amelioration happens, it becomes the case that a2 is and always has been the semantic content of C. This means that any amelioration that is not *final* has no role to play in the meaning of C *at any time*. This idea, that there can be no stepwise amelioration, is an awkward implication of Ball's view. Sawyer's view also has this implication, but for the simpler reason that for her, there is no genuine amelioration: Only our beliefs are ameliorated. Unlike those views, a proponent of NME can also allow for stepwise amelioration just as in the traffic rule case.

8. Conclusion

The aim of this paper has been twofold. First, I tried to draw attention to a number of related issues at the intersection of conceptual engineering, metasemantics, and questions of normativity. Whether some version of the accounts of Normative-Metaethical Entanglement that I defend in this paper are ultimately accepted, I think it is important that more people working in the conceptual engineering literature see that the normative critiques, in their various forms, at the heart of most conceptual engineering projects, means that there will be deep entanglement, at least in the questions (but probably also in the answers) that arise in developing a plausible theory of conceptual engineering, how it is possible, and what its success conditions are.

⁵⁶ Ball (2020, forthcoming).

Works Cited:

- Abreu, P. (2023). Metalinguistic negotiation, speaker error, and charity. *Topoi*, 1-16.
- Ball, D. N. (2020b). Revisionary analysis without meaning change (or, could women be analytically oppressed?). *Conceptual engineering and conceptual ethics*.
- Ball, D.N. (forthcoming). *Definition and Dispute*. Oxford University Press.
- Begby, E. (2013). Semantic minimalism and the “miracle of communication”. *Philosophical studies*, 165, 957-973.
- Belleri, D. (2017). Verbalism and metalinguistic negotiation in ontological disputes. *Philosophical Studies*, 174(9), 2211-2226.
- Benoist, J. (2021). Looking Back on Proposals on the Biological Aspects of Race. UNESCO, Moscow, 1964. Interview by Jean-Luc Bonniol. *Appartenances & Altérités*, (1).
- Brake, E. (2011). *Minimizing marriage: Marriage, morality, and the law*. Oxford University Press.
- Burgess, A., & Plunkett, D. (2013). Conceptual ethics I. *Philosophy Compass*, 8(12), 1091-1101.
- Burgess, A., & Plunkett, D. (2020). On the relation between conceptual engineering and conceptual ethics. *Ratio*, 33(4), 281-294.
- Burgess, A., Cappelen, H., & Plunkett, D. (2020). *Conceptual engineering and conceptual ethics* (p. 474). Oxford University Press.
- Cappelen, H. & Lepore, E. (2006). “Shared Content”, in Lepore, E., & Smith, B. C. (2006). *The Oxford handbook to the philosophy of language*.
- Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford University Press.
- Card, C. (1996). Against marriage and motherhood. *Hypatia*, 11(3), 1-23.
- Cantalamesa, E. A. (2021). Disability studies, conceptual engineering, and conceptual activism. *Inquiry*, 64(1-2), 46-75.
- Chalmers, D. J. (2012). *Constructing the world*. OUP Oxford.
- Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, 58(1), 7-19.
- Creath, R. (1990). *Dear Carnap, dear Van: The Quine-Carnap correspondence and related work: Edited and with an introduction by Richard Creath*. University of California Press.

- Davidson, D. (1973). Radical interpretation. *Dialectica*, 313-328.
- Davidson, D. (1994). Radical interpretation interpreted. *Philosophical perspectives*, 8, 121-128.
- Deutsch, M. (2020). Speaker's reference, stipulation, and a dilemma for conceptual engineers. *Philosophical studies*, 177(12), 3935-3957.
- Dunaway, B. (2020). *Reality and morality*. Oxford University Press.
- Dutilh Novaes, C. (2020). Carnap meets Foucault: Conceptual engineering and genealogical investigations. *Inquiry*, 1-27.
- Eder, A. M. A. (2021). Explicating the concept of epistemic rationality. *Synthese*, 199(1-2), 4975-5000.
- Edwards, K. (2014). Keeping (direct) reference in mind. *Noûs*, 48(2), 342-367.
- Eklund, M. (2023). Possible Limits of Conceptual Engineering: Magnetism, Fixed Points and Inescapability. *Argumenta*.
- Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 47, 187-225.
- Haslanger, S. (2005). What are we talking about? The semantics and politics of social kinds. *Hypatia*, 20(4), 10-26.
- Haslanger, Sally (2006). What good are our intuitions: Philosophical analysis and social kinds. *Aristotelian Society Supplementary Volume* 80 (1):89-118.
- Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.
- Haslanger, S. (2017, June). I—Culture and critique. In *Aristotelian Society Supplementary Volume* (Vol. 91, No. 1, pp. 149-173). Oxford University Press.
- Haslanger, S. (2020). Going on, not in the same way. *Conceptual engineering and conceptual ethics*, 230-260.
- Horgan, T., & Timmons, M. (1992). Troubles on moral twin earth: Moral queerness revived. *Synthese*, 92, 221-260.
- Isaac, M. G., Koch, S., & Nefdt, R. (2022). Conceptual engineering: A road map to practice. *Philosophy Compass*, 17(10), e12879.
- Knoll, V. (2020). Verbal disputes and topic continuity. *Inquiry*, 1-22.

- Koch, S. (2021). Engineering what? On concepts in conceptual engineering. *Synthese*, 199(1-2), 1955-1975.
- Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.
- Lewis, D. (1974). Radical interpretation. *Synthese*, 331-344.
- Marcus, R. B. (1961). Modalities and intensional languages. *Synthese*, 13(4), 303-322.
- Marques, T. (2017). What Metalinguistic Negotiations Can't Do. *Phenomenology and Mind*, (12), 40-48.
- Mölder, B. (2010). *Mind Ascribed. An elaboration and defence of interpretivism*. John Benjamins Publishing.
- Nado, J. (2021). Conceptual engineering, truth, and efficacy. *Synthese*, 198(Suppl 7), 1507-1527.
- Nimtz, C., (2024). Engineering concepts by engineering social norms: Solving the implementation challenge. *Inquiry*, 67(6), pp.1716-1743
- Parfit, D. (2011). *On what matters: Two-volume set*. OUP Oxford.
- Pearson, J. (2017). Carnap, explication, and social history. *Social Theory and Practice*, 741-774.
- Pinder, M. (2021). Conceptual engineering, metasemantic externalism and speaker-meaning. *Mind*, 130(517), 141-163.
- Plunkett, D. and Sundell, T., (2013). Disagreement and the Semantics of Normative and Evaluative Terms. *Philosopher's Imprint*, 13(23).
- Plunketta, D. and Sundell, T., (2021). Metalinguistic negotiation and speaker error. *Inquiry*, 64(1-2), pp.142-167.
- Plunkett, D. (2015). Which concepts should we use?: Metalinguistic negotiations and the methodology of philosophy. *Inquiry*, 58(7-8), 828-874.
- Putnam, H. (1975). The meaning of " meaning". *Philosophical Papers, Mind, Language, and Reality*, 2, pp.215-271.
- Railton, P. (1986). Facts and values. *Philosophical Topics*, 14(2), 5-31.
- Riggs, J. (2019). Conceptual engineers shouldn't worry about semantic externalism. *Inquiry*, 1-22.

Romero, A. (2012). When whales became mammals: the scientific journey of cetaceans from fish to mammals in the history of science. *New approaches to the study of marine mammals*, 3-30.

Sawyer, S. (2020). Truth and objectivity in conceptual engineering. *Inquiry*, 63(9-10), 1001-1022.

Sawyer, S. (forthcoming). Concept pluralism in conceptual engineering. *Inquiry*, 1-26.

Scharp, K. (2007). Replacing truth. *Inquiry*, 50(6), 606-621.

Schroeter, L., & Schroeter, F. (2014). Normative Concepts: A Connectedness Model. *Philosopher's Imprint*, 14(25).

Schroeter, L., & Schroeter, F. (2019). The Generalized Integration Challenge in Metaethics. *Nous*, 53(1), 192-223.

Schroeter, L., & Schroeter, F. (2020). Inscrutability and its Discontents. *Canadian Journal of Philosophy*, 50(5), 566-579.

Sider, T. (2013). *Writing the Book of the World*. OUP Oxford.

Simion, M., & Kelp, C. (2023). Linguistic innovation for gender terms. *Inquiry*, 1-21.

Thomasson, A. L. (2020). A pragmatic method for normative conceptual work. *Conceptual engineering and conceptual ethics*, 435-458.

Thomasson, A. (2021). Conceptual engineering: when do we need it? How can we do it?. *Inquiry*, 1-26.

Wakil, S. (2021). Objectivity and orgasm: the perils of imprecise definitions. *Synthese*, 199(1-2), 2315-2333.

Werner, P. J. (2020). Getting a moral thing into a thought: Metasemantics for non-naturalists. *Oxford Studies in Metaethics*, Vol.15 (ed. Russ Shafer-Landau).

Williams, J. R. G. (2020). *The metaphysics of representation*. Oxford University Press, USA.