Commitment engineering: Conceptual engineering without representations

Guido Löhr, Eindhoven University of Technology

*This paper has been accepted by Synthese, comments still welcome*

**Abstract:** It is largely assumed that conceptual engineering is essentially about revising, introducing, or eliminating representational devices, in particular the intension and extension of words and concepts. However, tying conceptual engineering too closely to representations is risky. Not everyone endorses the notion of representation as theoretically helpful or even real. Not everyone thinks that concepts or meanings should be understood in terms of the notion of representation. Does this mean that conceptual engineering is not interesting or relevant for these skeptics? In this paper, I motivate and propose a non-representationalist construal of conceptual engineering.I argue that conceptual engineers can be understood as primarily engineering linguistic entitlements and commitments rather than representational devices. Note that this account is non-representationalist, but explicitly *not* anti-representationalist. Representations may play a significant role when it comes to justifying and completing commitment engineering projects.

**Keywords**: Conceptual engineering, commitments, concepts, non-representationalism**,** normativity, pragmatism, commitments, concepts, meanings, non-representationalism

**1 Introduction**

Most accounts of conceptual engineering construe conceptual engineering as an essentially representationalist enterprise. It is largely assumed that what it is to engage in conceptual engineering is to introduce, eliminate or revise the intensions and extensions of words (Cappelen, 2018, Deutsch, 2020), concepts (Haslanger, 2000; Plunkett, 2015) and speaker meanings (Pinder, 2020; Deutsch, 2020). Another branch of the literature argues that conceptual engineers improve our bodies of representations used in categorization (Machery, 2018; Haslanger, 2006). However, representationalist approaches to conceptual engineering face well-known challenges: How can changes of representational content be implemented in light of semantic externalism (Deutsch, 2020; Jorem, 2020; Cappelen, 2018; Koch, 2021)? How can we change representational devices without changing the topic (Sawyer, 2020, Cappelen, 2018, Haslanger, 2000)? Finally, representation engineers may face more general worries regarding the explanatory value of representations, as well their potential indeterminacy and naturalization.

However, the practice of conceptual engineering is not essentially tied to representations.[[1]](#footnote-1) If the label “conceptual engineering” is supposed to *prescribe* what conceptual engineering is about, we can hold a non-representationalist theory of concepts following e.g., inferentialists like Rorty (1979) or Brandom (1994) or philosophers who take concepts to be abilities as opposed to objects (cf., Glock, 2010). If the label tracks an independently existing practice inside and outside of philosophy, we can describe this practice, again, in non-representationalist terms. This is what I will attempt in this paper: I argue that paradigm cases of conceptual engineering can fruitfully be construed as a practice that aims at changing our linguistic commitments and entitlements. However, note that this proposal is meant to loosen but not to cut all ties to representations. The aim is to provide an alternative framework for thinking about conceptual engineering for those who are skeptical about the notion of representation.

In the second section, I briefly elaborate on what I mean by “representationalist theories of conceptual engineering”. I take this summary to describe the current mainstream in the debate from a highly general point of view. In the third section, I briefly introduce some challenges for representation engineers. I do *not* take these challenges to be decisive against representationalist accounts. The purpose is merely to flag some worries that motivate the non-representationalist framework. In the fourth section, I introduce the notions of linguistic commitment and entitlement that are relevant for this paper. I argue that the relevant notion of commitment denotes normative relations that are grounded in the shared goal of the respective joint action. I then argue that conceptual engineering can be construed as the engineering of linguistic commitments. In the final two sections, I show how the present proposal can avoid problems of representationalist theories of conceptual engineering and how it relates to competing accounts.

**2 Representation engineering**

Let me introduce the representational framework of conceptual engineering by means of a concrete example. Imagine a fictional journalist called Piers who condemns only obvious discriminatory behavior based on skin color as racist and is largely oblivious about more structural or less obvious instances of discrimination. It is not that Piers lacks the concept of structural racism. Instead, he thinks that structural racism either does not exist or that it should not be called “racism”. Piers’ opponent, the activist Munroe, applies the expression ‘racist’ to a lot more cases of injustice. Piers feels uncomfortable with Munroe’s behavior and Munroe is appalled by Piers’. Still, both are trying to convince each other to change their behavior. Since both are public figures, they do so on national television. I take this to be a paradigm case of non-academic public conceptual engineering that we want to describe and explain.

Note that the story about Piers and Munroe does not involve the notion of representation. It is, however, easy to see how it could creep into a description of it, especially if one is trained in or influenced by the analytic tradition. To explain Piers’ behavior, we may assume that he mentally represents a set of conditions based on which he calls out cases of racism. These conditions are satisfied, for example, if a person is refused a job due to their race or skin color. Because more subtle instances of racism do not satisfy Piers’ conditions of racism, he will be oblivious about the wrongness of such cases or at least will not easily condemn them with the negatively charged label ‘racist’. Assuming Piers really has these representations, all that Munroe needs to do to change Piers’ behavior is change the conditions based on which he classifies and condemns events and actions as racist. At least this is largely assumed by a representationalist picture of conceptual engineering.

According to representationalist accounts of conceptual engineering, conceptual engineers evaluate, introduce, eliminate and revise representational devices, in particular the intension and extension of words, concepts, or meanings. Representations can minimally be construed as *objects* that have semantic properties, e.g., truth or satisfaction conditions, intensions or extensions. If an entity has such semantic properties, then it is a representation. If this object is a mental entity, then it is a mental representation. If this mental representation can be attributed to a person, as opposed to, say, the brain of a person, then it is a personal representation. If conceptual engineers aim to generate a change of behavior, then, on the representationalist picture, to engage in conceptual engineering is to engineer personal-level mental representations assuming that these representations are immediately accessible and under the voluntary control of the person.

To illustrate, imagine the meaning of ‘racism’ was determined by Piers’ mind who associates it with the complex concept DISCRIMINATION BASED ON SKIN COLOR. If Munroe wants to change the set of entities that the label ‘racism’ correctly applies to, she must convince Piers that it would be better to associate STRUCTURAL DISCRIMINATION BASED ON RACE with the term ‘racism’. If, however, the meaning of the word ‘racism’ was externally determined, e.g., by an external natural or social kind (cf., Löhr, 2021 for externalism applied to social kinds) or the use in the linguistic community, then Munroe has to change the world, e.g., this natural or social kind or the linguistic use (cf. Deutsch, 2020; Cappelen, 2018; Koch, 2021a). The potential change in concept or meaning, however, will not necessarily change the way Piers actually applies his concept of racism to the world as he might still only apply it to prototypical cases. What changes are the cases that the term or concept is *correctly* applied to. However, if Piers changes his meaning representations, he will at least be less defensive if Munroe directs his attention to more subtle instances of racism.

To really change Piers’ behavior, on the representationalist picture, Munroe has to change his “categorization device” (Löhr, 2020) or “conception” of racism, i.e., the set of beliefs, stereotypes, theories, exemplars or prototypes based on which he actually applies the word ‘racism’ or the concept of racism. She could, for example, show him movies that present him with less obvious cases of racism hoping that this will change Piers’ prototypes and exemplars. However, changing his behavior alone will not change the actual meaning of his terms, again, assuming the content of our words is not completely determined by actual individual use (they may still be determined by the mainstream or expert use). Munroe might get Piers to apply his words in a more inclusive way that may have positive consequences for discriminated people, but this change would now be based on a *wrong* application of words and concepts. Therefore, ideally, Munroe will change both the meaning of a term *and* our ways of applying this term and concept (cf. Koch, 2020).

Based on the distinction between representations that are applied (our words, meanings, concepts) and the set of representations based on which we decide whether these representations should be applied (what I called a “categorization device” or “conception”), we can now propose an abstract but still rather comprehensive overview of the different representationalist views of conceptual engineering present in the current literature: According to Edouard Machery (2018) or Sally Haslanger (2006), all we do when engineering concepts is change our categorization devices, i.e., the set of conditions based on which we judge whether a concept like ‘racism’ applies. According to Hermann Cappelen (2018), Sally Haslanger (2000) or Mark Pinder (2020), conceptual engineers change the standing or speaker meaning (construed as representations) of our expressions by intentionally replacing our concepts or meanings by trying to change the grounds based on which content is determined.

**3 Problems with representationalism**

There are mainly three kinds of challenges associated with representationalist approaches to conceptual engineering that at least weakly motivate the hope for alternative frameworks. One set of challenges concerns representationalism in general. For example, many philosophers have raised very fundamental worries regarding putting representations at the center of philosophical inquiry especially in the traditions of phenomenology (e.g., Heidegger, 1927/2010; Wittgenstein, 1953/2009; Merlaeu-Ponty, 1945) or American pragmatism (e.g., Rorty, 1979/2009; Brandom; 1994; Price, 2011). This disagreement is deep and complex and reviewing it will go beyond the scope of this paper. One especially relevant issue, however, concerns skepticism about whether language is essentially for representing the world rather than for acting in it. Most current theories of conceptual engineering clearly focus on language and thought as means to representing the world. This might be false.

The second set of challenges raised against representations in general and their role in an explanation of behavior in particular can be derived from the more recent philosophy of cognitive science. This literature is often inspired by American pragmatists but can be distinguished from it in scope and argumentation. In this literature, philosophers have raised concerns about the explanatory value of representations and the possibility of naturalizing their content (e.g., Stich, 1983; Ramsey, 2007; Egan, 2010; Hutto & Myin, 2012). Note that this tradition of representation skeptics is primarily responding to contemporary cognitive science and should be taken especially seriously by politically motivated conceptual engineers who are interested in changing our psychology, i.e., in changing behavior and thoughts. If representations cannot live up to their explanatory expectations or if there is no way of naturalizing their content, then this might put any representationalist approach to conceptual engineering in jeopardy.

The third set of problems of representationalism concerns conceptual engineering more specifically. These problems arise for current conceptual engineers arguably *because* they rely on representations as their primary target. Here is a selection (see Cappelen, 2018 for a discussion of these problems): First, what can it plausibly mean to *engineer* representations? If representations are essentially individuated by their content, then changing their content will simply replace them with another representation. Second, it has been argued that replacing our representations trivializes conceptual engineering, especially assuming that all we can really change are *speaker meanings* as opposed to the *semantic meaning* of a word (e.g., Deutsch, 2020, but see Koch, 2021b). Third, how will philosophers ever be able to change the representations of other people, which is especially a problem for conceptual engineers who adhere to semantic externalism? Finally, how can we change our representational devices without changing the topic given that the topic of a conversation is normally tied to representational content?

The latter class of more specific problems of representation engineering is familiar to anyone in the conceptual engineering debate and I do not deny that there have been very promising attempts to address them, perhaps even sufficiently (e.g., Simion & Kelp, 2019; Koch, 2021a on the issue raised by semantic externalism; or Thomasson, 2020 and Cappelen, 2018 on topics). However, it is also important to emphasize that at least some of these problems are arguably tied to the notion of representation (see section 6 for more details). We may only have to worry about topic continuity if we individuate topics in terms of representational content. Changing people’s behavior and thoughts only seems close to impossible (Cappelen, 2018) if we take this to mean that we have to change their representations. Nevertheless, the more general problems raised by phenomenologists, American pragmatists and contemporary philosophers of cognitive science should not be under-estimated. If there are fundamental problems with the notion of representation, we should not put all our hope into a representationalist account of conceptual engineering.

In this paper, I argue that despite the objections to representationalism, this does not mean that conceptual engineering is a hopeless endeavor. I propose a novel way of thinking about conceptual engineering that does not rely on representations. I argue that conceptual engineers do not primarily change our representational devices but our *linguistic commitments* and *entitlements*. Even if representation engineering is difficult if not impossible (Cappelen, 2018) and may not necessarily lead to any changes in linguistic and non-linguistic behavior (e.g., Deutsch, 2020), changing our entitlements clearly is possible (we do it all the time). However, any proposal to change our commitments requires justification (cf. Sterken, 2020). Here is where I take representations to become potentially relevant. One way of *justifying* a change of commitment is by proposing that the *standing meaning* of the term (construed as a representation) is incompatible with our current commitments and entitlements.

**4 Commitments in joint action**

Unlike representations, normative commitments and entitlements are *relations*, as opposed to mental or abstract *objects*.[[2]](#footnote-2) They are relations that normatively constrain the set of actions we are entitled to (e.g., Gilbert, 2013; Searle, 2010). For example, imagine Piers promises Munroe to bake a cake for her and Munroe accepts his promise. Piers now occupies the first place in a commitment relation of which Munroe occupies the second. Piers is, therefore, *committed* (to Munroe) to bake the cake. He would wrong Munroe if he failed to do so, e.g., if he simply forgot (Darwall, 2013; May, 2015). Piers may not feel like baking a cake for Munroe anymore, but he still has a *pro tanto* reason for doing so given his normative commitment directed at Munroe. Furthermore, if Piers breaches his commitment to Munroe, Munroe now has reason to criticize Piers (Brandom, 1994; Gilbert, 2006; Geurts, 2019), which might put further normative pressure on Piers’ future actions. If he breaks his commitments to Munroe more than once, she will eventually feel entitled to end her relationship with him and report his behavior to others.

The notion of commitment that is used here should not be conflated with different *non-normative* notions that are also associated with the term (cf., Tebben, 2019). For example, one can commit to a person or object, meaning that one has made a decision or generated an intention. One might also be committed to *going* to the gym regularly, meaning that one is dedicated or highly motivated to going to the gym.[[3]](#footnote-3) Both uses of ‘commitment’ express certain actions or psychological states as opposed to normative relations. Actions and psychological states like beliefs, motivations or intentions may have normative consequences (they may generate normative commitments or obligations) but are themselves not normative constraints. One might publicly “commit” to marry another person by means of a certain speech act and this action might generate a normative commitment to marry, such that one is vulnerable to criticism if one breaches the commitment. Still, a promise or assertion is itself not a normative commitment. They are actions with normative consequences.

A number of different kinds of normative constraints on behavior have been identified in the literature. A familiar kind of normative commitment are *moral obligations*. Moral obligations are grounded in moral norms and give us normative reasons to act that are largely independent of our inclinations (Searle, 2010; Gilbert, 2013). For example, I may be morally obligated to save a drowning child whether or not I would like to do so. I might be morally obligated not to use slurs even if I am inclined to do so. Another familiar kind of normative constraint has been called “rational commitment” (Shpall, 2014). Rational commitments are grounded in psychological states and norms of rationality. A person who makes a decision or generates an intention to act seems rationally committed to act as decided, such that she would make a mistake if she simply forgot to act as intended (Bratman, 1987; Gilbert, 2006). Similarly, based on Descartes’ beliefs, one might say that he was rationally committed to endorse ontological dualism. Anything else would be irrational.

In this paper, I focus on an arguably different kind of constraint on our actions and attitudes. It is largely assumed that when acting with others we owe them something – we have obligations and commitments to them (Bratman, 2013; Gilbert, 2013; Tuomela, 2013; Alonso, 2009; Searle, 2010; Roth, 2017). Such commitments are usually taken to derive – in some way or other – from the joint intentions or shared goals that normatively guide joint action (e.g., Gilbert, 2013). If we want to go to the cinema together, there are certain actions that do and certain actions that do not contribute to this goal. Arriving in the middle of the movie, for example, does not contribute to the goal of enjoying a fun and relaxing movie night together. If I am significantly late, you would be entitled to criticize me by referring not necessarily to any general moral rule but simply our agreement and our shared goal.[[4]](#footnote-4)

How can we change our commitments when cooperating with others? One way of course is to change our goals. If we no longer jointly intend to go to the cinema together, we are also no longer committed to act accordingly. I am then no longer committed to be on time and wait for you in front of the cinema. However, changing a goal is not the only way one can change what is most cooperative for a given joint action. What we are socially committed and entitled to do in a joint action also depends on the attitudes of the participants in the joint action, e.g., what we believe to be true and *how* we would like to reach the shared goals. The attitudes of each participant exert genuine normative pressure on our actions. For example, there are many ways one can fulfill the shared goal of taking a walk together (Gilbert, 2013). If I know that you do not like certain kinds of jokes, I might not be entitled to make them even if taking a walk together would still be possible if I did. Similarly, if I convince you that racism is a systemic problem in Europe, you are no longer entitled to criticize me for making certain inferences that follow from this new common ground.

A third way of changing our cooperative commitments, i.e., to change how to best reach a certain shared goal, is to change our social norms. Imagine most members of our group are mostly on time and take this to be what everyone in the group ought to do (cf., Bicchieri, 2016). If we now decided to meet to discuss your paper tomorrow, it would be most cooperative, i.e., most conducive to our shared goal, if you respected this rule to be on time. Diverging from the rule disrupts our cooperative potential. If you do not show up on time, you owe me an explanation. Thus, it is not necessarily the social norm itself that generates the normative force (I do not want to commit here to any view on the normativity of social norms). The coordinating potential of such a norm is already enough to change what action is and what action is not most conducive to a shared goal. In other words, a reason why we are often committed to respect the social norms in our group is that, unless agreed otherwise, relying on our social norms is the best way of coordinating our individual actions.

**5 Commitment engineering**

What does the possibility of changing our commitments in a joint action have to do with the practice of conceptual engineering? I argue that the dos and don’ts in a joint action also apply to or can be extended to the linguistic actions that are taken to contribute to a shared goal. Not all speech acts contribute equally well to a shared goal and there are often norms regulating what to say and how to say it, i.e., how to act linguistically in a given situation. This includes terminological choices. If we want to cook pasta and we need to add salt, I would *not* be entitled to ask you whether you can pass the sodium chloride even if you are entitled to infer that I want salt (given that sodium chloride is salt). The reason is that unless you take the unusually formal request to be amusing, the unusual terminological choice likely delays cooperation and does not seem to add anything to the joint activity. You would now be entitled to (you have a good reason to) criticize me for this overly formal style and defend this criticism based on the fact that it delays fulfilling our joint intention.

To return to the practice of conceptual engineering, I argue that conceptual engineers can be understood as the engineering not of representational devices (primarily at least), but of linguistic commitments and entitlements, i.e., the normative constraints pertaining to the use of words and sentences that are grounded in the shared goal of a joint action (that uses language). I propose that one way of accomplishing a change of commitments and entitlements that can be called conceptual engineering is to convince individuals in the group that changing their behavior and norms contributes better to a shared goal. Once we change our social norms and behaviors, we change what is most cooperative given a shared goal that we are mutually committed to in a joint action. From this mutual commitment in combination with our norms derive our commitments to act linguistically, i.e., to interact using words and sentences (engage in speech acts).

I take the simplest case of conceptual engineering to be the introduction of a term. For example, imagine that our group regularly observes and talks about a phenomenon, say, men taking up a lot of physical space on the train by spreading their legs and arms. A member of the group may now propose to introduce the term ‘manspreading’. If she can convince everyone that the term is helpful relative to a shared goal (say, of describing the world or to reach gender equality), it now seems that the group members are entitled to use the term in order to avoid the complex description of the phenomenon. If the term catches on, it would be uncooperative to keep using the more complex description. It may now also be cooperative to ask the speaker why she does not use the newly accepted terminology. The uncooperative speaker is now committed to present a good reason for her action. This reason, however, must relate to the shared goals of the group.

Another common instance of conceptual engineering is elimination. If we can convince other members of the group that a word does more harm than good, say the word ‘concept’ or ‘innateness’ (Machery, 2009; 2018), then we are no longer entitled to use this word. This likely has the consequence that we either no longer talk about certain topics or perhaps talk about the same topic in new ways that may be better relative to a certain goal. Importantly, however, just because a term may be objectively harmful, this does not mean that we are yet entitled to avoid it. It might still be more cooperative relative to a shared goal to keep using the term. Only once the change is adopted by the others, i.e., if it catches on, will it be most conducive to reaching a certain shared goal.[[5]](#footnote-5)

Furthermore, conceptual engineers evaluate whether our current entitlements and commitments actually serve a function or purpose in our joint action or whether they may even harm or stand in the way of reaching the shared goal (Simion & Kelp, 2020; Thomasson, 2020). Once they identify entitlements that are better replaced by other entitlements, the engineer will want to convince the other speakers to change their behavior accordingly. For example, the engineer might lobby for a change of entitlement, e.g., to apply the word ‘woman’ not to individuals with certain biological features, but to individuals who are oppressed based on certain features perceived as feminine. Once this change is adopted by the group, we are then entitled to infer from calling someone a woman that the speaker is committed to consider this person to be oppressed. Once we manage to change our linguistic entitlements without having to justify the new use, the engineer’s work has been done. If everyone in the community uses ‘woman’ in the new way, it would be uncooperative not to do so.

Importantly, other inferences are thereby blocked by the new social norm. We are no longer entitled to infer that the individual has certain biological properties if we are not entitled to such an application of the term. The reason is simply that it would be uncooperative to use the word ‘woman’ in a sense that contradicts the norm without justification. Moreover, this change of inferential commitment often comes with institutional changes, e.g., the entitlement to use certain restrooms or other restricted spaces. Changing our linguistic commitments may also have tremendous consequences for the way we represent and act in the world. For example, becoming socially entitled to call a person who engages in more subtle forms of discriminatory behavior a racist will generate further entitlements. It will entitle others to infer that this person has certain irrational and dangerous beliefs or a bad character. It will also socially entitle others to avoid this person and maybe even deny them access to certain groups or public services. Changing our social entitlements and commitments changes our social world as well as our representation of it (cf., Simion, 2018).

Another interesting and arguably less subversive kind of commitment engineering is to make an explicit commitment to a use of a word that one was entitled but not committed to. Thus, consider that we are often using the word ‘racism’ in a relatively broad or imprecise way. In many contexts, the speaker will be entitled to specify their use. If we are genuinely socially entitled to this specification, we are usually not committed to give any reasons (unless we are in certain social contexts with special reason-giving social norms, e.g., in an academic debate). If other people feel entitled to ask for reasons, this is fallible evidence that we were not entitled to the given specification. Of course, we may sometimes also be *committed* and not just entitled to specify what we mean by a term. This could be the case if the use of the term is too broad for the purpose of the joint action. For example, if there are too many sets of inferences one could be entitled to when using the term to actually make a useful contribution to the goal of the given context, we should specify what set of inferences we wish to commit ourselves to.

To illustrate this last point, Munroe and Piers might argue about whether a public person, say, Donald Trump is racist. To avoid miscommunication, they may explicitly commit to a specific use of ‘racist’. They might mutually agree that they socially commit to use ‘racist’ only for discriminatory behavior based on race or ethnicity, at least for the sake of the argument. They are entitled to this reduction of entitlements in this context, e.g., in a private discussion because it contributes to their shared goal. When they later talk to other people with whom they do not stand in this new commitment relation, they may no longer be committed to the narrow use even if they may still be entitled to it. Despite being a less subversive case of commitment engineering, explicit linguistic commitments of this kind are still an important part of communication in science and conceptual engineering in general.

**6 Relation to other accounts**

Note that the commitment engineering account is compatible with many representationalist accounts of conceptual engineering. Again, the purpose of this paper is not to show that these accounts are false, hopeless or uninteresting. The aim is merely to provide an alternative framework for how to think about conceptual engineering. In fact, I take it that there are many ways in which current representationalist frameworks of conceptual engineering can be incorporated into a commitment view. I take commitment engineering, therefore, to be a broader notion than representation engineering. In fact, one can understand representation engineering as one way of defending a proposal to change commitments and entitlements. The important point is that conceptual engineering construed as commitment engineering does not necessarily require representations if, e.g., it turns out that representation-talk is less useful than currently thought.

More concretely, I argued above that there are largely two kinds of mental representation that representation engineers are trying to engineer. First, there are those representations that we apply to things in the world (we apply the concept of table to tables, for example). Second, there are those mental representations that we rely on to *judge* whether a representation correctly applies (e.g., how we judge whether something is a table or not). I also argued that this distinction helps us classify the different approaches to conceptual engineering prominent in the literature. According to Cappelen (2018), Haslanger (2000) or Pinder (2020), conceptual engineers change the correctness or application conditions of words and concepts. According to Machery (2018) or Haslanger (2006), engineers change the set of conditions based on which we judge whether a concept like ‘racism’ applies. Both approaches, I argued, are compatible and both are important. We do not only want to change the way people actually apply representational devices. We want this change to lead to correct applications.

On the commitment engineering view, these different representationalist approaches to conceptual engineering can be understood as different ways of *justifying* and of *accomplishing* a change of commitments and entitlements. However, to say that representations may be important or useful for conceptual engineering does not mean that conceptual engineering is essentially about changing intensions and extensions of representations or about changing bodies of beliefs. It might be that because we often accomplish changes of commitments by referring to representations as means of *justification* that we are let to believe that conceptual engineering is *essentially* about changing representations. In fact, however, representation engineering is merely a means to an end. The end, I argued is a change of commitments and entitlements. For example, Sally Haslanger, arguably does not just want to change our representations but her linguistic entitlement to use the word ‘woman’ in a new way when communicating with her peers.

More concretely, Haslanger (2006) justifies her proposal that we should be entitled to apply ‘woman’ also to trans women by means of her claim that the actual standing meaning of this term (construed representationally) refers to a social as opposed to a biological kind (assuming semantic externalism). In other words, she relies on a meta-semantic claim about the actual meaning and correct use of the term to *justify* her demands about how we should be entitled to use the term in a given joint action (say, in academia). This, however, does not necessarily mean that using the term in this “correct” way is also most cooperative, i.e., that we are socially entitled or committed to it. For this to be the case another step is necessary. Only if everyone in the relevant group uses the term in this way, e.g., because they are convinced by Haslanger’s reasoning, does it become cooperative to follow this use. Only then do we become entitled and committed to do so.

Similarly, according to Machery (2018), conceptual engineering is essentially about changing our categorization devices, i.e., the body of information we rely on when judging whether a certain word applies. As mentioned above, changing our attitudes can lead to changes in commitments. For example, if Munroe manages to convince Piers that racism can be systemic and that the word ‘racism’ can pick out this systemic kind of racism, this changes the set of things Piers is now entitled to apply the term to. So even if Munroe merely changes Piers’ attitudes, she thereby also changes what actions are most conducive to reaching a certain shared goal, say the goal of discussing publicly about whether white people in the UK are racist simply by benefiting from the racist system set up by a white elite. Thus, Machery’s representation engineering can be viewed as one way of changing our commitments.

At this point, one may wonder about the relevance of shared goals in my proposed framework. I assumed that commitments derive from shared goals of a joint action. However, often conceptual engineers have different goals from the people whose concepts they want to change.[[6]](#footnote-6) For example, when Haslanger proposes to change the notion of woman, she does so because she takes the conceptual change to contribute better to her political goals. Of course, not everyone shares the same goals as Haslanger, for example. These people first need to be convinced of her goals, e.g., by means of a peaceful political conversation. To achieve such a change of goals, it is usually best to use language that everyone understands. Thus, Haslanger is not normally entitled to her new use of ‘woman’ when talking to people who are not already on board with her goals. It would then be better to rely on the old use, at least temporarily.

Note that Haslanger herself explicitly talks to other philosophers following certain shared goals and does not propose to change the use of the word when talking to people who either do not know the debates she is responding to or who do not share her political goals. I thus assume that conceptual engineers tend to change the linguistic entitlements and commitments of the community they are part of rather than the linguistic entitlements of members of other groups with other shared goals. For example, imagine two groups have fundamentally different goals. In such a case, Haslanger would only *not* be entitled to propose a change of commitments. Being an outsider, she would probably not even be heard or taken seriously. The same is the case of course for construing conceptual engineering as a representationalist project. The probability of a group changing the representations of another without engaging in manipulation and without exerting political pressure is rather small. I take it that if we think of conceptual engineering as a means to change linguistic entitlements and commitments, and language primarily as a means to facilitate cooperation toward a shared goal, engaging in conceptual engineering without sharing at least some goals is not only impossible but difficult to justify.

None of this is to say that we cannot use breaches of linguistic commitments to disrupt a certain discourse, i.e., to engage in activism. Political activists often rely on explicit breaches of cooperative commitments to initiate a change of norms and commitments (cf., Sterken, 2020). It is important, however, to acknowledge that such disruptions are powerful *because* they are breaches of linguistic commitments. We may even be justified in doing so if there are more general moral obligations or principles that support this breach of a local cooperative commitment (see more on this below). Thus, the notion of commitment engineering helps us understand the disruptive potential of conceptual engineering. Changing representations alone arguably lacks this disruptive potential.

Some readers may also wonder how the present account relates to the project of conceptual ethics and so-called “metalinguistic negotiation”. I take conceptual ethics to assess and negotiate which concepts or words we *should* use. Often, we engage in conceptual ethics by means of metalinguistic negotiations, i.e., negotiations about how to use a word (cf., Burgess & Plunkett, 2013). Conceptual engineering is the *practice* of changing (introducing, eliminating and revising) our concepts. If we adopt the framework introduced here, we can understand conceptual ethics and metalinguistic negotiations as instances of negotiating and assessing linguistic commitments and entitlements and the process of conceptual engineering as a kind of commitment engineering. Again, none of this means that we may not refer to representations to convince others to change our commitments and entitlements. The point is that we do not need to frame metalinguistic negotiations and engineering in representationalist terms (which is acknowledged by Burgess & Plunkett, 2013).

Finally, there has been some discussion on the question of the function of concepts (e.g., Simion, 2018; Nado, 2019; Simion & Kelp, 2020; Haslanger, 2020; Thomasson, 2020). The idea is that conceptual engineers:

engage in reverse engineering to figure out, empirically, what function(s) the concepts have served and do serve [...] and do descriptive conceptual work in figuring out how they work [...]. But we must also do work in conceptual ethics to determine what functions our concepts are to serve, going forward, *given our shared purposes*. Finally, we combine that with empirical work, in doing constructive conceptual engineering: determining whether (given worldly constraints) certain modifications or precisifications would better enable the term to fulfill its function. (Thomasson, 2020; p. 454, my emphasis).

So, Nado, Simion, Kelp, Haslanger, Thomasson and other conceptual engineers assume that concepts have functions and that these functions play an important role in conceptual engineering. I endorse this idea and we can easily frame the same point in terms of commitments. Commitment engineering then involves identifying the function of a term and how it relates to our commitments and entitlements. If, for example, we are entitled to apply ‘marriage’ only to heterosexual couples, we may identify the role this entitlement plays in society, observe that it is harmful or unfair (assuming that our shared goal is to live in a fair society) and then decide to change these normative constraints in the ways I proposed above. As argued by Thomasson (forthcoming), this is not always easy and requires empirical work as well as negotiating constraints on our engineering ambitions, but it is clearly not impossible or even difficult in some cases – especially when the time is ripe (cf., Bicchieri, 2016). I thus share Kelp and Simion’s (2020) enthusiasm that, with the right framework, conceptual engineering can be rather easy or at least possible even if one may disagree with their representationalist construal of the function of concepts.

**7 Some problems avoided**

Construing conceptual engineering as the practice of engineering cooperative commitments, allows us to avoid at least some of the concerns some philosophers have had with representationalism. First, conceptual engineering is no longer primarily about changing our representations. We also no longer presuppose any strong relation between representations and behavior. If conceptual engineering is primarily about changing our commitments and entitlements in collective actions, we can leave open what role representations play in generating a change of behavior. This may be attractive both for representationalists and their opponents. Thus, even those opposed to representations can still think about conceptual engineering as a fruitful activity, namely as a kind of commitment engineering.

Moreover, avoiding representations helps us avoid challenges to conceptual engineering that are more specific to representation engineering (see section 3). Recall the argument that replacing our representations trivializes conceptual engineering, especially assuming that all we can really change are *speaker meanings* as opposed to the *semantic meaning* of a word (e.g., Deutsch, 2020, but see Jorem, 2020). However, changing our commitments and entitlements is not as easy as simply replacing concepts or speaker meanings. We have to change social norms of shared goals and we have to have many people on board with what we are proposing such that commitments actually change. Moreover, if we get people to change our social norms and behaviors with respect to certain applications of words and concepts, we can literally change the world. If people become socially entitled to apply ‘marriage’ to homosexual couples, this will have tremendous institutional and personal consequences.

Another problem that an account of commitment engineering avoids is the following: As argued above, we can be in control over conceptual engineering without the need for any “implementation” of a representational change. We are in control of conceptual engineering because we are somewhat in control of our social commitments and entitlements, e.g., by changing our social norms. One method of changing social norms is to convince enough people that using a word in a new way is preferable, which will hopefully change their behavior (which then changes our cooperative commitments and entitlements). Convincing others to change their actions is clearly possible, even if it is admittedly not always easy. However, it is also not as difficult as one might think, at least in cases where a change of entitlement is “ripe”, i.e., when the society is generally at a stage where it is open to the respective change.

The time is ripe for a commitment change when enough people clearly see the importance of changing our commitments and entitlements when conversing and debating. This is arguably the case for the term ‘mansplaining’ that was mentioned above. Many people immediately saw that the expression helps us communicate about a certain phenomenon. The time is currently not ripe for a large-scale change of the use of terms like ‘racist’. Many people resist the change of commitment also because they fail to see the use of such a change. Currently, activists (who I consider to be conceptual engineers) are pushing for a commitment change not just by convincing others of the use of such a term but by actually using the term in this new way. This breach of commitment is intentional and causes friction (when talking to more conservative audiences) but might actually contribute to more and more people adopting the new use thereby making it a new social norm. This breach of commitment is usually justified on moral as opposed to cooperative grounds.

Finally, consider how commitment engineers can respond to the problem of topic change (cf., Koch, forthcoming). How can we change our representational devices without changing the topic given that the topic of a conversation is normally tied to the sameness of content (Cappelen, 2018)? Again, I do not want to argue that there have not been very promising representationalist attempts to avoid this problem, e.g., by changing the way we individuate topics (e.g., Sawyer, 2020 or Cappelen, 2018 for attempts). I merely wish to point out that construing conceptual engineering as commitment engineering has the potential to avoid this problem elegantly. The solution is to individuate topics not in terms of representational content but in terms of a shared goal. For example, think of the question of whether free will is possible given what we know about the brain. A commitment engineer will argue that this formulation is not precise enough for counting as a shared goal or topic. For it to be a topic, it is not clear enough what we mean by ‘free will’, i.e., what actually follows from using the word ‘free will’ and when we are entitled to apply it. Proposing a more specific use of the word ‘free will’ is then not an attempt to change but to clarify what the topic is supposed to be. Clarifying the topic makes discussing the topic possible in the first place. Commitment engineers do not change the topic. They clarify and determine the topic or question under discussion (von Stutterheim & Klein, 1989).

How can a goal individuate a topic or question under discussion? Take again discussions about free will in philosophy. It is misleading to think that philosophers get together to talk about the concept of free will. Instead, they get together for example because they notice that lay people or other philosophers often use the expression ‘free will’ and that they take this expression to be relevant when applying the expression ‘moral responsibility’, for example. Now some philosophers might get together to investigate the use and intuitions regarding these expressions. This is a shared goal, and the topic is arguably not the concept of freedom but getting closer to the actual or preferred use (our actual or preferred commitments) of the expression ‘freedom’. To propose that when using ‘freedom’ we should be entitled to mean XYZ then is not to change the topic, but to contribute to the topic or question under discussion.

If a critic objected that the new set of entitlements and commitments is changing the topic, i.e., changing a goal in the cooperative activity, she is saying that the engineer was not entitled to their use of words. With the tools introduced in this paper, we can now see why she thinks this, or better, whether she is entitled to this criticism. For example, if the overall or higher order goal is to study the actual use of a concept, the critic is right. Narrowing down what we are entitled and committed to when using the expression “freedom” might not be conducive to this goal and might even harm communication with others about what it means to be free. Still, now these philosophers might embrace the objection and simply change the shared goal, or they had a different goal in mind in the first place. The group might then split into two fractions with different shared goals.

For instance, imagine the goal is not to analyze the actual use but, say, to propose a legal text that requires more precision and use of technical terms. Arguably then the introduction of new entitlements and commitments in this context contributes to the shared goal without changing the topic. A third possibility is that the critic was not entitled to object that the change of entitlements generates a change of topic. She is not entitled to this objection if there was no shared goal in the first place. The critic would have to specify what the original goal or topic was that all participants are, according to the critic, committed to pursue. I suspect that in many cases where concerns of topic change are brought forward, there was no such topic to begin with.[[7]](#footnote-7)

**Conclusion**

There are many objections to the current theorizing of conceptual engineering. However, these objections mainly respond to a representation-heavy understanding of this practice. I argued that the study and practice of conceptual engineering are not doomed even if these critics are right. I proposed an alternative way of thinking about the practice of conceptual engineering as an attempt to change our social entitlements and commitments, especially those pertaining to language. Commitment engineering is neither trivial nor impossible. What we have to do is change what is cooperative in a given joint action, e.g., by changing our goals, attitudes and social norms. This does not mean that this is easy. However, if the time is ripe and the arguments or reasons presented convincing, conceptual engineering might actually be easier than we might think.

**Acknowledgements**

I thank Steffen Koch, Sigurd Jorem, Joachim Horvath, Alex Wiegmann, Vincent Müller, Phillip Brey, Michael Klenk, Julia Hermann, Joel Anderson and the anonymous reviewers for helpful comments.

**Funding**

This work is part of the research programme Ethics of Socially Disruptive Technologies, which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031). It has also been funded by Ruhr University Research School PLUS, funded by Germany’s Excellence Initiative [DFG GSC 98/3].

**References**:

Alonso, F. M. (2009). Shared intention, reliance, and interpersonal obligations. *Ethics*, *119*(3), 444-475.

Austin, J. L. (1975). *How to do things with words*. Oxford university press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bratman, M. (1987). *Intention, plans, and practical reason* (Vol. 10). Cambridge, MA: Harvard University Press.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard university press.

Brandom, R. (2009). *Articulating reasons: An introduction to inferentialism*. Harvard University Press.

Burgess, A., & Plunkett, D. (2013). Conceptual ethics I. *Philosophy Compass*, *8*(12), 1091-1101.

Camp, E. (2015). Logical concepts and associative characterizations. In: Margolis, L. & Laurence, S. (eds). *The Conceptual Mind: New Directions in the Study of Concepts*. MIT Press. 591-621.

Cappelen, H. (2018). *Fixing Language: an essay on conceptual engineering*. Oxford: OUP.

Clark, H. H. (2006). Social actions, social commitments, in *Roots of Human Sociality: Culture, Cognition, and Interaction*, eds N. J. Enfield and S. C. Levinson (New York, NY: Berg), 126–150.

Darwall, S. (2013). *Morality, authority, and law: Essays in second-personal ethics I*. Oxford University Press.

Deutsch, M. (2020). Speaker’s reference, stipulation, and a dilemma for conceptual engineers*. Philosophical Studies*, **177,**3935–3957.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Geurts, B. (2019). Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*(1-2), 1-30.

Gilbert, M. (2013). *Joint commitment: How we make the social world*. Oxford University Press.

Glock, H. J. (2010). Concepts, abilities and propositions. *Grazer Philosophische Studien*, *81*, 115-134.

Gomez-Lavin J, Rachar M (2021). Why We Need a New Normativism about Collective Action. *Philosophical Quarterly*.

Gomez-Lavin J, Rachar M (2019). Normativity in joint action. *Mind & Language* 34(1):97–120.

Haslanger, S. (2000). Gender and race:(What) are they? (What) do we want them to be?. *Noûs*, *34*(1), 31-55.

Haslanger, S. (2006). I—Sally Haslanger: What Good are Our Intuitions?. In *Aristotelian Society Supplementary Volume* (Vol. 80, No. 1, pp. 89-118). Oxford, UK: Oxford University Press.

Haslanger, S. A. (2020). How not to change the subject. In T. Marques & A. Wikforss (Eds.), *Shifting Concepts*: *The Philosophy and Psychology of Conceptual Variability*. Oxford: Oxford University Press.

Hampton, J. A., & Jönsson, M. L. (2012). Typicality and Composition a Lity: the Logic of Combining Vague Concepts. In: Werning, W. Hinzen, and E. Machery (eds.). *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press.

Jorem, S. (2020). Conceptual engineering and the implementation problem. *Inquiry* 64:1-2, 186-211. <https://doi.org/10.1080/0020174X.2020.1809514>

Merleau-Ponty, M. (1945/1982). *Phenomenology of perception*. Routledge.

Koch, S. (forthcoming). Why conceptual engineers should not worry about topics. *Erkenntnis.*

Koch, S. (2021a). The externalist challenge to conceptual engineering. *Synthese*, 198:327–348.

Koch, S. (2021b). There is no dilemma for conceptual engineering. Reply to Max Deutsch. *Philosophical Studies*, 178(7), 2279-2291.

Koch (2020). Engineering what? On Concepts in Conceptual Engineering. *Synthese*. <https://doi.org/10.1007/s11229-020-02868-w>

Löhr, G. (2020). Concepts and categorization: do philosophers and psychologists theorize about different things?. *Synthese*, *197*(5), 2171-2191.

Löhr, G. (2021). Social constructionism, concept acquisition and the mismatch problem. *Synthese*198**,**2659–2673*.*

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Machery, E. (2018). *Philosophy within its proper bounds*. OUP.

Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford University Press.

May, S. (2015). Directed duties. *Philosophy Compass* 10, no. 8 (2015): 523-532.

Nado, J. (2019) Conceptual engineering, truth, and efficacy. *Synthese*. https://doi.org/10.1007/s11229-019-02096-x

Owens, D. (2016). Promises and conflicting obligations. *J. Ethics & Soc. Phil*., 11, 1.

Pinder, M. (2021). Conceptual engineering, metasemantic externalism and speaker-meaning. *Mind*, *130*(517), 141-163.

Plunkett, D. (2015). Which concepts should we use?: Metalinguistic negotiations and the methodology of philosophy. *Inquiry*, *58*(7-8), 828-874.

Price, H. (2011). *Naturalism without mirrors*. Oxford University Press.

Prinz, J. (2002). *Furnishing the mind*. MIT Press.

Rorty, R. (2009). *Philosophy and the Mirror of Nature* (Vol. 81). Princeton university press.

Roth, A. S. (2017). Interpersonal Obligation in Joint Action. In M. Jankovic & K. Ludwig, (eds.). *The Routledge Handbook of Collective Intentionality* (pp. 45-57). London: Routledge.

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

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Searle, J. (2010). *Making the social world: The structure of human civilization*. Oxford University Press.

Shapiro, L. (2018). Commitment Accounts of Assertion. In: Sanford, G (ed.). *The Oxford Handbook of Assertio*n. Oxford University Press.

Shpall, S. (2014) Moral and rational commitment. *Philosophy and Phenomenological Research*, 88(1), 146-172.

Simion, M. (2018). The ‘should’in conceptual engineering. *Inquiry*, *61*(8), 914-928.

Simion, M., & Kelp, C. (2020). Conceptual innovation, function first. *Noûs*, *54*(4), 985-1002.

Sterken, R. (2000). Linguistic Interventions and Transformative Communicative Disruption. In Herman Cappelen, David Plunkett & Alexis Burgess (eds.), *Conceptual Engineering and Conceptual Ethics*. Oxford: Oxford University Press. pp. 417-434.

Tebben, N. (2019). Knowledge requires commitment (instead of belief). *Philosophical Studies*, *176*(2), 321-338.

Thomasson, A. (2020). A Pragmatic Method for Conceptual Ethics. In: Cappelen, H., Plunkett, D. and Burgess, A. (eds.). *Conceptual Engineering and Conceptual Ethics*. Oxford University Press. 435-458.

Thomasson, A. (forthcoming). Conceptual Engineering: When do we need it? How can we do it? *Inquiry*.

von Stutterheim, C., & Klein, W. (1989). Referential movement in descriptive and narrative discourse. In R. Dietrich & C. F. Graumann (Eds.), Language processing in social context (pp. 39–76). Amsterdam: North Holland.

1. Of course, if we define “conceptual engineering” as engineering representational devices (as done by Cappellen 2018, p. 3), conceptual engineering is tied to representations by stipulation. But this is not how the term is largely used in the literature given that there are now several non-representational notions of conceptual engineering (e.g., Thomasson, 2020) and given that the topic is largely introduced in terms of examples, e.g., of Haslanger’s proposal to change the meaning of ‘woman’. [↑](#footnote-ref-1)
2. Note that I build here on the recent literature on interpersonal obligations and entitlements in social ontology (Alonso, 2009; Roth, 2017; Gilbert, 2013) rather than Brandom’s notion of commitments and entitlements. The relation between both notions cannot be analyzed here but there is at least considerable overlap. I take Brandom’s notion to be grounded, however, not in shared goals but in social practices and norms (e.g., Brandom, 1994, p. 161) or *reciprocal recognition* (Brandom, 2009). Such an approach is currently explored for example by Thomasson (forthcoming). [↑](#footnote-ref-2)
3. The use of the gerund is evidence that we are talking about a state comparable to that of a dedication to an action or object. We can be dedicated to going to the gym but it feels less natural to say that we feel dedicated to go to the gym. The term obligation on the other hand is used normally without the gerund. [↑](#footnote-ref-3)
4. See Gomez-Lavin & Rachar (2019, 2021) for empirical evidence for such commitments and entitlements. [↑](#footnote-ref-4)
5. This is why Machery (2009), for example, cannot get around using the very term he wishes to eliminate – the word ‘concept’. Elimination takes time and if the community does not adopt the change of the norm, then it is more cooperative to keep using the accepted term. Even if the term might not be ideal, not using the term might delay reaching the shared goal even more than using it. [↑](#footnote-ref-5)
6. I thank an anonymous reviewer for pointing this out to me. [↑](#footnote-ref-6)
7. Of course, a representationalist could make use of the same strategy of individuating topics in terms of shared goal (thanks to Steffen Koch for pointing this out). But this would be a concession to the commitment engineer as she would now no longer individuate topic in terms of shared content. [↑](#footnote-ref-7)