



How to engineer a concept

Vera Flocke¹

Accepted: 2 September 2020 / Published online: 21 October 2020
© Springer Nature B.V. 2020

Abstract One dimension of cognitive success concerns *getting it right*: having many true beliefs and no false ones. Another dimension of cognitive success concerns *using the right concepts*. For example, using a concept of a person that systematically excludes people of certain demographics from its extension is a sort of cognitive deficiency. This view, if correct, tasks inquirers with critically examining the concepts they are using and perhaps replacing those concepts with new and better ones. This task is often referred to as “conceptual engineering”. However, so far it is unclear what exactly happens in cases of conceptual engineering. How does language change when we engineer a concept? This article offers an answer. I propose a view on which, when speakers assess the truth of propositions, they often rely on assumptions with regard to what is required for their truth. For example, when speakers assess whether unborn fetuses are people, they rely on assumptions with regard to what is required to be a person. Based on this idea, I develop a model of conceptual engineering according to which speakers “engineer concepts” when they change how they assess the truth of propositions. For example, speakers engineer the concept of a person when they change how they assess the truth of the proposition that unborn fetuses are people.

Keywords Conceptual engineering · Social reality · Social construction · Non-factual disagreements

✉ Vera Flocke
vflocke@iu.edu

¹ Bloomington, USA

1 Introduction

One dimension of cognitive success concerns *getting it right*: having many true beliefs, and no false ones. Another dimension of cognitive success concerns *relevance*. For example, knowing the number of hairs on my cat's body presumably is not a case of cognitive success—you are wasting your cognitive capacities—but when a doctor knows the symptoms of many kinds of illness, that is cognitive success. And yet another dimension of cognitive success concerns *using the right concepts*. For example, using a concept of a person that systematically excludes people of certain demographics from its extension is a sort of cognitive deficiency.

The idea that the concepts we use in our thought and inquiries can be better or worse has sparked several debates in recent years.¹ Some philosophers argue that certain concepts that are currently in use should be improved or replaced by better ones. For example, Haslanger (2000) argues that the gender concepts that are currently in use should be improved. Other philosophers argue that certain debates are best understood as concerning which concepts we should use. For example, Thomasson (2017) argues that metaphysical debates about the nature of personhood are best understood as concerning which concept of a person we should use. More radically, Cappelen (2018) argues that a big proportion of philosophical debates overall are cases of conceptual engineering. Yet other philosophers debate the dimensions in which concepts can be better or worse.

This article takes up certain *metasemantic* questions relating to conceptual engineering. Semantics is the study of what words mean; and metasemantics is the study of what makes it the case that words mean what they mean. For example, social semantic externalism—the thesis that the meanings of words are fixed by broad use patterns in the speech community—is a metasemantic theory. I am here interested in what exactly happens in cases of conceptual engineering. How does language change when we engineer a concept? This article offers an answer.

My main idea is that, when speakers assess the truth of propositions, they often rely on assumptions about what it takes to be an object of a certain kind. For example, when speakers assess whether an unborn fetus is a person, they rely on assumptions about what it takes to be a person. More generally, when speakers assess whether *F*s are *G*s, they often rely on assumptions about what it takes for an *F* to be *G*. When we “engineer concepts”, we modify these assumptions. That is, we change some aspect of how we assess the truth of propositions.

This answer is not the only and not the whole answer to the question of what happens in processes of conceptual engineering. The representation of the world through language is a complex process, and there may be many ways of improving this process. My aim here is to describe one way in which we can change and hopefully improve our representational devices. But there are other ways

The main difference between the sort of process that I describe here and other processes that have been described under the heading of “conceptual engineering” is that it is entirely within your control. You can change how you assess whether *F*s

¹ See Burgess and Plunkett (2013a, b) and Burgess et al. (2019) for an overview and discussion.

are Gs right here and right now. All you need to do is to change certain behavioral dispositions; specifically, dispositions that govern how you assess the truth of propositions. Of course, changing a behavioral disposition may not always be easy, as anyone who has ever tried to change sleep patterns, eating habits, or an exercise regime can attest. But it is something that is in principle within your control. In contrast, the kinds of processes that other authors describe under the heading of “conceptual engineering” usually have to do with changing the referents of terms (see Cappelen 2018, Part V). Assuming (some variant of) semantic externalism, changing the referent of a term is not within your control. This article thus offers a silver lining to people wanting to engineer a concept. There is a way by which you can change and hopefully improve your representational devices that is entirely within your control.

My discussion will be structured as follows. In the next section, I describe a few cases of conceptual engineering (Sect. 2). Next, I present questions that a metasemantic theory of conceptual engineering must answer (Sect. 3). I go on to present my positive view. I first offer an informal statement of my account of what is going on in cases of conceptual engineering (Sect. 4), and then provide a formal semantics for the view (Sect. 5), before explaining how the view meets the challenges (Sect. 6). I go on to compare my approach with alternative theories of the metasemantics of conceptual engineering section). and conclude with a summary (Sect. 8).

2 Conceptual engineering

It is controversial whether there are any cases of conceptual engineering. (More on that below.) I will here describe three cases that are plausible candidates, and explain what makes them plausible candidates. I will then go on to discuss some of the challenges that people who are of the view that these are cases of conceptual engineering (“conceptual engineers”, for short) must confront.

Is Waterboarding Torture?

Different institutions employ different definitions of ‘torture’. The United Nations defines ‘torture’ as any act inflicting severe suffering, physical or mental, to obtain information or to punish.² Under this definition, waterboarding clearly is a form of torture. The former U.S. Justice Department, on the other hand, defines ‘torture’ as any such act inflicting pain rising to the level of death, organ failure, or the permanent impairment of a significant body function.³ On this definition, only the most extreme forms of waterboarding are forms of torture. Suppose now that speaker A says “waterboarding is torture” and speaker B says “waterboarding is not torture”. In certain contexts, speaker A can be understood as proposing or endorsing the

² See United Nations, General Assembly (1984).

³ See U.S. Department of Justice, Office of the Assistant Attorney General (2002).

United Nation's definition of torture, while speaker B can be understood as proposing or endorsing the former U.S. Justice Department's definition.⁴

Are Fetuses Persons?

'Fetal personhood' refers to the notion that unborn fetuses have the rights of people, which includes protection against harms such as assault and homicide. In general, proponents of fetal personhood oppose certain rights of pregnant women, such as the right to have an abortion or the right to use certain legal drugs which might endanger the fetus' life or well-being. Opponents of fetal personhood, however, generally are advocates for the unrestricted rights of pregnant women. Suppose now that speaker A says "unborn fetuses are people" and speaker B says "unborn fetuses are not people". In certain contexts, speaker A can be understood as proposing or endorsing a concept of personhood that includes unborn fetuses, while speaker B can be understood as proposing or endorsing a concept of personhood that does not include unborn fetuses.⁵

Is Sex Binary?

Fausto-Sterling (2000) argues that there are more than two sexes among humans.⁶ On her view, the sexes are distinguished in at least five different dimensions: chromosomal sex, gonadal sex, hormonal sex, genital sex, and brain sex.⁷ None of these dimensions of sex is binary. E.g., some people have XXY chromosomes and hence a third chromosomal sex. Furthermore, the different dimensions of sex do not pattern uniformly. Some people who score 'male' in one of the dimensions nevertheless score 'female' in other dimensions. For example, some people with a male chromosomal sex nevertheless possess female external genitals because of a testosterone insensitivity. However, Byrne (2018a) argues (for a weakened version of the thesis) that sex is binary. On Byrne's view one's sex depends on the relative size of the reproductive cells that one produces: females produce large gametes, while males produce small ones. By this criterion, most people can be classified as either male or female, which leads Byrne to conclude that sex is binary.^{8,9} Fausto-Sterling and Byrne can be understood as proposing alternative concepts of the human sexes.

⁴ For discussions of this case, see Chalmers (2011), Sundell (2011) and Plunkett and Sundell (2013, p. 19).

⁵ The New York Times' Editorial Board (Dec. 28, 2018) has provided a opinion piece that critically discusses some of the recent legal and social developments concerning fetal personhood. See Thomasson (2017) for a philosophical discussion of fetal personhood.

⁶ See Fausto-Sterling (2018) for a popular summary of Faust-Sterling's position.

⁷ To explain the less obvious of these notions, 'gonadal sex' refers to the sort of gametes that one produces (egg or sperm cells). 'Brain sex' refers to the way in which brain cells stimulate the levels and patterns of sex hormones.

⁸ More precisely, Byrne concludes that "The existence of some unclear cases shows that it would be incautious to announce that sex (in humans) is binary. By the same token, it is equally incautious to announce that it isn't—let alone that this is an established biological fact".

⁹ Clarke (2018) provides a response to Byrne, to which Byrne (2018b) supplies a brief rejoinder.

The disagreements in these three cases all are of legal relevance. If waterboarding is torture, it is not a legal interrogation technique; if an unborn fetus is a person, it cannot be legally aborted; and if sex is not binary, then marriage cannot plausibly be defined as a relation between a man and a woman, since this would unjustly preclude a large number of people from getting married. This fact—that all three cases are of legal relevance—highlights that the stakes for deciding one way rather than another are high. But not all cases of conceptual engineering are of political or legal relevance; it is merely a common feature of the three cases considered here.¹⁰

What makes the three cases look like cases of conceptual engineering is that they have a certain definitional dimension. Each of the three disagreement turns in some way on how we define a certain term (such as ‘torture’, ‘person’ or ‘man’ and ‘woman’). The two parties in the three disputes consistently use terms in different ways. A proponent of fetal personhood calls fetuses “people”, but an opponent of fetal personhood never calls a fetus a person. Similar for the other cases. To settle these disagreements, we must settle on a common definition of certain key terms.

Saying that a disagreement has a “definitional dimension” is not very clear and is not very precise. Metasemantic theories of conceptual engineering try to turn this vague impression into a clear and precise theory that explains what is going on in disagreements with such a “definitional dimension”.

3 Challenges for conceptual engineers

A metasemantic theory of conceptual engineering needs to meet several demands. First, and most obviously, there is a need to explain the nature of the process. What are concepts? What does it mean to propose an improvement of the concepts we use, or a replacement of the ones we use by alternative ones? What does it mean to engineer a concept? Surprisingly, the main theories of the metasemantics of conceptual engineering that are currently available do not invoke concepts at all; and the theory I present here will be no exception. But this makes questions about the nature of the process even more pressing. What could it mean to engineer a concept if there aren’t any concepts?

Second, a metasemantic theory of conceptual engineering must make plausible that conceptual engineering is at least in principle possible. But there is apparent reason for skepticism. According to an objection that descends from an argument initially due to Strawson (1963), there is not such thing as conceptual engineering. When speakers attempt to “engineer a concept”, none of their concepts change but they rather change the topic. For example, on this view, we can talk about *people*, or about *people*_{*}. But there is no such thing as the engineering of the concept of a person. When we shift from talking about *people* to talking about *people*_{*}, we merely shift the topic. I will call the challenge created by these skeptics THE STRAWSONIAN CHALLENGE.

¹⁰ The converse obviously does not hold either. Not every disagreement that is of political or legal relevance is a case in which speakers attempt to engineer a concept.

Third, conceptual engineers typically think that the engineering of concepts is an important and worthwhile project to undertake. Haslanger (2000) thinks that we ought to improve our gender concepts; Cappelen (2018, p. 72) thinks that we “must keep trying” to improve our concepts. But there is reason to think that this endeavor is in most cases bound to fail. Specifically, semantic externalism is the thesis that the semantic values of the terms we use does not depend on speakers’ mental states or intention. The basic thesis can be spelled out in different ways, but they all have the consequence that a single speaker’s intention to use a term with a particular meaning will have no immediate effect on the term’s meaning. At best, the speaker could try to change broad use patterns in the linguistic community at large and thereby over time and slowly produce a change in a term’s meaning. This is a slow and uncertain process, whose success is not in the control of a unique speaker, or a small group of speakers.

Semantic externalism arguably is the most popular contemporary metasemantic framework. But the view creates a challenge. If semantic externalism is true then it looks like engineering concepts is not in the power of speakers and therefore a pointless exercise. Maybe some speakers are *trying* to engineer a concept, but this project is bound to fail. But if ought implies can, and we cannot engineer concepts, then it is not the case that we ought to engineer concepts. I will call this THE EXTERNALIST CHALLENGE.

Fourth, conceptual engineers need to explain *why* we should try to engineer concepts. What is good about changing the concepts we use? In response to this challenge, conceptual engineers often emphasize the idea that conceptual engineering has a certain “worldly” dimension (see Cappelen 2018, Part IV). In some way, changing the concept of a woman influences who are the women in this world. But this is surely puzzling. How could our linguistic activities have any effect on who is or is not a woman? As the old saying goes, if you called a dog’s tail a leg, he would still have just four legs. Calling a tail a leg doesn’t make it one. I’ll call this challenge PUZZLING WORLDLINESS.

Summing up, here is a list of the challenges that a metasemantic theory of conceptual engineering needs to meet and a brief statement of what is needed to meet them.

THE STRAWSONIAN CHALLENGE Explain how the engineering of a concept is different from merely changing the topic.

THE EXTERNALIST CHALLENGE Explain how we can engineer concepts even if semantic externalism is true.

PUZZLING WORLDLINESS Explain why we should engineer concepts. Specifically, explain the worldly dimension of conceptual engineering.

4 To count an F as a G

The key idea of the theory I want to propose is that speakers assess the truth of propositions guided by assumptions regarding what is required for their truth. Put in very general terms, when speakers assess whether *F*s are *G*s, they often rely on

assumptions regarding what it takes for an F to be a G . That is not to say that speakers *always* assess the truth of propositions in a rule-governed way. Sometimes speakers accept a proposition without assessing its truth in any way. Furthermore, if they do use rules, the relevant rules do not need to take the form of assumptions regarding what it takes to be an object of a certain kind. The relevant rule could alternatively defer to an authority (“believe what the teacher says”), or take some altogether different form (“believe every other proposition”). However, the view is that *often* speakers assess the truth of propositions in a rule-governed way, and *often* these rules are assumptions regarding what is required for the truth of a given proposition.¹¹

The view can be illustrated by mean of the three cases presented in the last section. It suggests the following diagnosis:

- When speakers assess whether waterboarding is torture, they often rely on assumptions regarding what is required for an interrogation technique to constitute a form of torture. The UN and the former U.S. Justice Department rely on different.
- When speakers assess whether an unborn fetus is a person, they often rely on assumptions regarding what it takes to be a person. Fetal rights activists and women’s rights activists rely on different assumptions.
- When speakers assess whether people who are disposed to produce comparatively small gametes (sperms) are male, they rely on assumptions regarding what it takes to be male. Byrne and Fausto-Sterling rely on different assumption.

An important aspect of this analysis of the three cases is that, in these cases, speakers assess the same propositions guided by *different* rules. For example, the former U.S. Justice Department and the UN assess whether waterboarding is torture guided by different assumptions about what it takes to be an act of torture. The disagreement between the UN and the former U.S. Justice Department concerns the conditions under which an interrogation counts as a form of torture. The two parties disagree on how the truth of the proposition that waterboarding is a form of torture should be assessed. In order to resolve this disagreement, at least one of the two parties has to adjust what they think is necessary and sufficient for an interrogation technique to count as a form of torture.

The claim that different rules of assessment are in principle available does not entail that these rules are all *correct*. Perhaps the UN definition of torture is in fact the correct definition of torture, and the former U.S. is making a mistake by relying on a different one. However, telling a representative of the former U.S. Justice Department that they are using the wrong definition of torture, by itself without any additional argument or evidence, is very unlikely to change their mind. The point here is that the disagreement concerns how to assess whether waterboarding is a form of torture. Whether or not this is a factual disagreement is inessential for the analysis.

¹¹ This view is inspired by my interpretation of Carnap. In Flocke (2020), I argue that Carnap thought that speakers assess each other’s utterances guides by rules that include syntactic rules, semantic rules and rules for the evaluation of empirical evidence.

The idea that speakers often assess the truth of propositions guided by rules, and that sometimes at least multiple rules are in principle available naturally leads to a certain view on the metasemantics of conceptual engineering. Conceptual engineering happens when speakers change how they assess the truth of propositions. The three cases are cases of attempted conceptual engineering because, in these cases, speakers are trying to influence other speakers on how they assess the truth of propositions. For example, the UN tries to influence the former U.S. Justice Department to change what it thinks is required for an interrogation technique to count as a form of torture. If the former U.S. Justice Department changes its definition of torture, that would mean that it effectively changes how it assesses the truth of the proposition that waterboarding is a form of torture.

5 Semantics

I will give a model-theoretic semantics for this view, which is based on the idea that propositions (what is said by declarative sentences) do not represent objective reality, but represent how things are as seen from an individual's standpoint. An individual's standpoint is shaped by the rules of assessment that the individual follows. For example, from the standpoint of fetal rights activists, unborn fetuses are persons, but from the standpoint of women's rights activists unborn fetuses are not persons.

Given this idea, propositions cannot in general be sets of worlds. Worlds are classically complete and decide for each proposition p whether p is true. But standpoints can be undecided. For example, an individual may be undecided regarding what's required for an interrogation technique to be form of torture. The formal representation of individual standpoints should reflect the possibility of such indecision. So, if propositions represent how things are from an individual's standpoint, propositions cannot in general be sets of worlds.

An alternative to possible-worlds semantics is given by a framework known as "possibility semantics" (see Humberstone 1981; Holliday 2018). A possibility semantics replaces Kripke frames, in which the basic states are worlds, with possibility frames. The basic states in a possibility frame are not worlds but more general "possibilities". Unlike worlds, possibilities can have refinements. For example, while every world is such that either waterboarding is or is not torture, a possibility may leave this matter unsettled. If possibility p has one refinement p' at which waterboarding is torture and another refinement p'' at which waterboarding is not torture, then waterboarding neither is nor is not torture at p .¹² Worlds are possibilities that cannot be refined.

¹² ¹²Coppock (2018) offers an alternative "outlook-based" semantics. She distinguishes between worlds and *outlooks*, where an outlook is a refined of a world and settles certain things that aren't settled by a world. She uses this framework to give a semantics for subjective taste predicates. The main reason why I prefer a possibility semantics, as presented here, has to do with the metaphysics of worlds. I think that worlds are complete and cannot have refinements as Coppock's (2018) outlook-based semantics requires.

With the general notion of a possibility in hand, we can proceed to giving model-theoretic interpretations of sentences. A proposition is a set of possibilities, and a proposition P is true at a possibility p if and only if $p \in P$. A proposition P is false at possibility p iff P is not true at any of the refinements of p . So, a proposition is neither true nor false at a possibility p iff $p \notin P$ but, for some refinement p' of p , $p' \in P$.

An individual occupies a standpoint partly in virtue of his or her mental states. For example, an individual who assesses the truth of the proposition that waterboarding is torture by considering whether waterboarding is torture according to the UN definition occupies an individual standpoint at which all and only forms of interrogation are forms of torture that are classified as such by the UN definition. So, individuals who assess whether waterboarding is torture in different ways occupy somewhat different individual realities. The proposition that waterboarding is torture may be true at some of these standpoints and false at others.

The talk of “truth at a possibility” should not be mistaken for a commitment to a form of truth-relativism, as defended by MacFarlane (2014). Truth-relativists think that truth is a dyadic property that propositions have only relative to something else, such as a world or some other non-standard parameter (such as, for instance, a taste-parameter). On MacFarlane’s (2014) version of this view, values for the relevant parameters are fixed by *contexts of assessment*. On this view, the proposition that vegemite is tasty as assessed from your context of assessment but false as assessed from my context of assessment. But according to the view proposed here, truth may well be a monadic property that propositions have absolutely speaking, and not relative to this or that parameter. For example, the proposition that waterboarding is torture is true if and only if waterboarding is torture. The upshot of the view is just that waterboarding may be torture at my standpoint but not at yours. Standpoint-dependent realities are more fine-grained than objective reality. People who seem to live in the same physical universe may nevertheless occupy different individual realities. That does not entail that truth is a dyadic property.

6 Meeting the challenges

The view can be summarized as follows: Speakers, in general, assess the truth of propositions guided by rules. For example, they assess whether waterboarding is torture guided by assumptions about what it takes for an interrogation technique to be a form of torture. Conceptual engineering takes place when a speaker, or a group of speakers, changes how they assess the truth of a given proposition. For example, one engineers the concept of torture when one changes how one assesses whether waterboarding is a form of torture. How does this theory answer the challenges I outlined in Sect. 3?

Recall THE STRAWSONIAN CHALLENGE, according to which there is no such thing as conceptual engineering. Purported cases of conceptual engineering simply are cases in which a speaker or group of speakers changes the topic. The view developed here has a straightforward response. When speakers change how they assess the truth of a proposition, they do not change the topic. The proposition

they assess is the same before and after the change. Speakers just modify what they do in order to assess its truth.

Semantic externalists have no reason for thinking that conceptual engineering, so understood should be impossible. Semantic externalism is a view about the relation between words and meanings. A semantic externalist thinks that what words mean does not depend on the intentions, dispositions, or other mental states of individual speakers. The theory of conceptual engineering presented here is silent on the relation between words and meanings. The idea is that speakers can assess the truth of semantics contents in various ways. That can be so whether or not the meanings of words are fixed by factors external to the speaker. This is how I answer THE EXTERNALIST CHALLENGE.

One might think that certain versions of externalism still create a problem. In particular, one might think that speakers don't have a choice between alternative ways of assessing a proposition. There is a correct view on what is required for the truth of the proposition that waterboarding is torture. When speakers assess the truth of this proposition while making any alternative assumption on what has required for its truth, they are simply making a mistake. On this view, it is not up to speakers to choose how they assess the truth of propositions. However, there is no real conflict between this externalist view and my theory of conceptual engineering. Even if there is a correct way of assessing the truth of a specific proposition, speakers may still assess it in different ways, and they may change how they assess it. If they do, they have "engineered a concept". Some of the ways of assessing the truth of a proposition may be based on a false belief, but that does not mean that speakers never make use of them. Sometimes people break the rules.

Why do we engage in these sorts of disagreements? Why do speakers engage in debates where they try to influence how other speakers assess the truth of propositions? The answer to this question relates to topics studied by psychologists working on social cognition.¹³ The core assumption of social cognition research is that the behavior of an individual is not directly determined by the social situation, but rather by the individual's interpretation of this situation. Psychologists sometimes express this idea in grandiose-sounding terms, by saying that "individuals construct their own individual reality" (see Bless et al., 2004). But what is meant really is that an individual's respond to a stimulus is not directly determined by the stimulus, but depends on complex cognitive processes by which the individual interprets the stimulus. For example, suppose you overhear someone at a party saying "I succeed at everything I put my mind to". You might interpret this utterance either as a sign of arrogance, or as a sign of confidence. In the former case, you will presumably develop a negative attitude towards the speaker and avoid them in future interactions. In the latter case, you might develop a positive attitude and seek them out for further interactions. In the two cases, you have interpreted the same stimulus in different ways and thus "constructed two different individual realities".

¹³ For general introductions to social cognition research, see Fiske and Taylor (1984), Bless et al. (2004) and Kruglanski (1989).

Psychologists working on social cognition research typically focus on the interpretation of *social* stimuli, which arise when two or more people interact. But the same model can be used to explain how individuals interpret other stimuli that are not obviously social. For example, a speaker who finds all sweet things tasty interprets sweet taste in a particular way. From the perspective of this speaker, it is *as if* every sweet thing is in fact a tasty thing. Another speaker who does not like sweet things will interpret the same stimulus—sweet taste—differently. From the perspective of this speaker, sweet things are not tasty. These two speakers interpret the same stimulus (sweet taste) in different ways. These divergent interpretations predict their future behavior more so than the stimulus by itself. Someone who likes sweet taste will seek out sweet food, and someone who does not like sweet taste will avoid it.

When speakers change how they assess the truth of propositions, they effectively change how they interpret certain stimuli. For example, when a speaker changes how they assess the truth or the proposition that an unborn fetus is a person, they change who, from their perspective, counts as a person. When speakers debate how to engineer a concept, they can hence be understood as trying to influence one another's interpretation of certain stimuli. They can be seen as changing how they construct their own standpoint-dependent reality (keeping in mind that what's meant by the construction of reality here really is the interpretation of certain stimuli).

Changing one's individual reality is not the same as changing objective reality. For example, by acquiring a taste for sweet things you may change which things are tasty from your individual standpoint; but you do not thereby change the sugar content of anything. However, individual realities often are much more important for guiding agents' decisions and actions, as compared to objective reality. As Bless et al. (2004, p. 6) put it, "it is not the stimulus per se that influences our behavior but our perception of it; in other words, the way in which we mentally construct and represent reality." For example, knowing that a fetus is 4 months old does not by itself determine whether one is for or against allowing its abortion. What is needed in addition is a view on what's required for being a person. Debates on how to assess the truth of certain propositions can therefore have far-reaching consequences. These debates have the power to influence a broad range of socially relevant behavior, by way of influencing how agents interpret stimuli, and thereby construct their individual realities.

This view explains the PUZZLING WORLDLINESS of conceptual engineering by appealing to the concept of an individual reality. The key idea of psychologists working on social cognition research is that individuals interpret stimuli with which they are presented, and thereby construct their own individual reality. I have argued that cases of conceptual engineering are cases in which individuals change how they interpret certain phenomena, and thereby change how they construct their individual reality. For example, an individual who changes what they think is required for someone to be a woman changes their individual reality so that, after the change, certain people either newly count as women or else no longer count as women. This is how, in a sense, changing the concept of a woman changes who the women are. The view is compatible with thinking that, in addition to individual realities, there is

also an objective reality that is entirely unaffected by people's linguistic activities, including how they go about assessing the truth of propositions.

What evidence can we give someone to make them change their mind? To answer this question, we would need to find out what motivates agents to interpret phenomena in one way rather than another; and this question in turn is an empirical question. Psychologists have developed three different models (see Bless et al., 2004, pp. 3–5). The first views agents as *consistency seekers*, the second views agents as *lay scientists*, and the third views agents as *cognitive misers*. Consistency seekers are primarily motivated by avoiding contradictions and inconsistencies in their thinking. Lay scientists are primarily motivated by trying to align their individual reality with objective reality. Cognitive misers are incapable of processing all the information they are confronted with every day and use shortcuts and rules of thumb when interpreting phenomena. Agents may use different strategies depending on the context, and plausibly often rely on a combination of the three approaches. Depending on what kind of agent you are talking to, different kinds of argument will have the greater power to convince. Consistency seekers will change their mind when they become aware of an inconsistency in their thinking; but cognitive misers are more likely moved to change their mind when presented with a handy shortcut.

This concludes my presentation of the positive view. I have given a theory of the metasemantics of conceptual engineering, and have sketched how the theory responds to some of the most common concerns about conceptual engineering. Next, I will go on to compare the view with alternative theories of conceptual engineering, and draw out the advantages of my account.

7 Alternative analyses

Several theories have been proposed or can be read as theories of the metasemantics of conceptual engineering. Some theories assume the existence of concepts or concept-like entities, and propose that processes of conceptual engineering modify these entities.¹⁴ However, other theories with virtually the same explanatory power have been proposed that do not assume the existence of concepts. As these theories are more parsimonious, they are to be preferred. I will here discuss two versions: Plunkett and Sundell's (2013) "metalinguistic negotiation" view, and Cappelen's (2018) "austerity framework".

First, according to Plunkett and Sundell (2013), disagreements of the kind I described in Sect. 2 are forms of "metalinguistic negotiation": speakers negotiate how certain terms ought to be used (see Plunkett and Sundell 2013). On this view, speakers often express meta-linguistic proposals by seemingly non-linguistic claims. For example, when a women's right activist says "an unborn fetus is not a person", they make a meta-linguistic proposal for how the term 'fetus' ought to be used. However, this approach has at least two drawbacks. First, speakers don't seem

¹⁴ See, for instance, Haslanger (2000). See Cappelen (2018, Ch. 16) for a critical discussion.

concerned with *terms*. When speakers debate whether an unborn fetus is a person, they are interested in personhood, not in the term ‘person’. They would in most cases be happy to express their view using alternative terms, without using ‘person’ at all. Second, it seems that the disagreement between women’s rights activists and fetal rights activists can be expressed in any number of languages. Use of the English term ‘person’ is not necessary for expressing this disagreement.

The social constructionist theory developed in this paper avoids this problem of the metalinguistic negotiation view since it does not ascribe to speakers any concern with *terms*. For example, the key claim about the fetal personhood case is that speakers disagree on what it takes to be a person. This disagreement has nothing to do with terms and can be expressed in any number of languages.

Second, Cappelen’s (2018) “austerity framework” explains conceptual engineering rather differently. The framework is built on an externalist foundation. Cappelen suggests that what terms refer to in general does not depend on speakers’ intentions or other mental states. Furthermore, Cappelen also thinks that all semantic change should be explained on the model of reference change, and he suggests that, when we engineer a concept both the intension and the extension of terms change. These views taken together imply that speakers can never purposefully engineer a concept. The engineering of a concept is in general out of our control. Cappelen is fully aware of this consequence of the austerity framework, but he insists that “we will and should keep trying” (p. 72).

The main advantage of the social constructionist theory as compared to Cappelen’s austerity framework is that I describe a process which is in principle within our control. Of course, changing any sort of disposition, including dispositions to assess the truth of propositions in a particular way, can be difficult. But it is with good will and effort in principle possible. That does not mean that I am providing the true metasemantics of conceptual engineering, and other theories are false. Language can change in many different ways. Processes through which the referents of terms change and processes through which we change how we assess the truth of propositions may both aptly be called processes of conceptual engineering. However, a speaker tasked with evaluating and perhaps improving or replacing the representational devices they are using has a choice to make and needs to decide which project they will embark on. Conceptual engineering of the kind I describe in this paper is the more feasible project and therefore to be preferred.

8 Conclusion

The idea that “using the right concepts” is an important dimension of cognitive success has recently gained traction among philosophers. However, so far it is unclear what happens in cases when speakers “engineer a concept”. This paper has developed a theory. I have developed a view on which speakers assess the truth of propositions guided by assumptions about what’s required for their truth. For example, speakers assess how many sexes exist among humans guided by assumptions regarding what’s required to be a male or female (or to have some other sex). Starting from this idea, I have argued that speakers “engineer a concept”

when they change how they assess the truth of propositions. The main advantage of this view as compared to Cappelen's (2018) "austerity framework" is that it describes a process that is in principle at least within speaker's control. A consequence of Cappelen's approach is, however, that speakers can never purposefully engineer a concept.

Acknowledgements I would like to thank an anonymous referee for helpful comments and my audience at the 2019 Chapel Hill Normativity Workshop for helpful discussions. Work on this article has profited from several visiting fellowships at the research project ConceptLab at the University of Oslo. Many thanks to the directors for inviting me.

References

- Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition. How individuals construct social reality*. New York: Psychology Press.
- Burgess, A., Cappelen, H., & Plunkett, D. (2019). *Conceptual ethics and conceptual engineering*. Oxford: Oxford University Press.
- Burgess, A., & Plunkett, D. (2013a). Conceptual ethics I. *Philosophy Compass*, 8(12), 1091–1101.
- Burgess, A., & Plunkett, D. (2013b). Conceptual ethics II. *Philosophy Compass*, 8(12), 1102–1110.
- Byrne, A. (2018a). Is sex binary? *Arc Digital*, 11/1/2018. Retrieved November 22, 2018, from <https://arcdigital.media/is-sex-binary-16bec97d161e?sk=37aad314c28136c9290b27d612df2c8>.
- Byrne, A. (2018b). Alex Byrne responds. *philosomama (blog)*, 11/14/2018. Retrieved November 22, 2018, from <http://philosomama.blogspot.com/2018/11/alex-byrne-responds.html>.
- Cappelen, H. (2018). *Fixing language. An essay on conceptual engineering*. Oxford: Oxford University Press.
- Chalmers, D. (2011). Verbal disputes. *Philosophical Review*, 120(4), 515–566.
- Clarke, E. (2018). On whether sex is binary. *philosomama (blog)*, 11/09/2018. Retrieved November 22, 2018, from <https://philosomama.blogspot.com/2018/11/on-whether-sex-is-binary.html>.
- Coppock, E. (2018). Outlook-based semantics. *Linguistics and Philosophy*, 41, 125–164.
- Fausto-Sterling, A. (2000). *Sexing the body: Gender politics and construction of sexuality*. New York: Basic Books.
- Fausto-Sterling, A. (2018). Why sex is not binary. *The New York Times*, 10/25/2018. Retrieved November 22, 2018, from <https://www.nytimes.com/2018/10/25/opinion/sex-biology-binary.html>.
- Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. New York: Random House.
- Flocke, V. (2020). Carnap's Noncognitivism about ontology. *Nouûs*, 54(3), 527–548.
- Haslanger, S. (2000). Gender and race: (what) are they? (what) do we want them to be? *Nouûs*, 34(1), 31–55.
- Holliday, W. H. (2018). Possibility frames and forcing for modal logic. <https://escholarship.org/uc/item/0tm6b30q>.
- Humberstone, L. (1981). From worlds to possibilities. *Journal of Philosophical Logic*, 10(3), 313–339.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge*. New York: Plenum Press.
- MacFarlane, J. (2014). *Assessment sensitivity. Relative truth and its applications*. Oxford: Oxford University Press.
- Plunkett, D., & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, 13(23), 1–37.
- Strawson, P. F. (1963). Carnap's views on conceptual systems vs natural languages in analytic philosophy. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 503–518). Chicago: Open Court Publishing.
- Sundell, T. (2011). Disagreement, error, and an alternative to reference magnetism. *Australasian Journal of Philosophy*, 90(4), 743–759.
- The New York Times' Editorial Board. (Dec. 28, 2018). A woman's rights. how states are extending liberties to the unborn—by taking them from women. *The New York Times*, <https://nyti.ms/2GGuBkM>.
- Thomasson, A. L. (2017). Metaphysics and conceptual negotiation. *Philosophical Issues*, 27, 364–382.

- United Nations, General Assembly. (1984). Convention against torture ad other cruel, inhuman or degrading treatment or punishment. United Nations Treaty Series 1465.
- U.S. Department of Justice, Office of the Assistant Attorney General. (2002). Memorandum for Alberto R. Gonzales, Counsel to the President, Re: Standard of Conduct for Interrogation under 18 U.S.C. xx 2340–2.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.