Conceptual Engineering and the Dynamics of Linguistic Intervention

Adam F. Gibbons
Department of Philosophy, Lingnan University
Hong Kong Catastrophic Risk Centre

Abstract. The Implementation Problem for conceptual engineering is, roughly, the problem conceptual engineers face when attempting to bring about the conceptual change they support. An important aspect of this problem concerns the extent to which attempting to implement concepts can lead to unintended negative consequences. Not only can conceptual engineers fail to implement their proposals, but their interventions can produce outcomes directly counter to their goals. It is therefore important to think carefully about the prospect of attempted implementation leading to unintended negative consequences: what sort of negative consequences can conceptual engineers expect? Are some forms of conceptual engineering more likely than others to lead to such consequences? And is conceptual engineering still viable even given the risk of such consequences? This paper addresses such questions. I begin by outlining different forms of conceptual engineering (Section 1), before examining how they can produce unintended negative consequences (Section 2). I then discuss some implications of the fact that attempted implementation can produce unintended negative consequences, suggesting that, among other things, some forms of conceptual engineering are less viable than others (Section 3). I conclude, though, by considering some of the ways in which conceptual engineering is nonetheless a worthwhile pursuit (Section 4).

Keywords: concepts; conceptual engineering; the implementation problem; linguistic intervention; unintended consequences

Introduction

Conceptual engineering is the activity of assessing and improving our concepts.[1] Not content with merely describing the concepts we currently have, conceptual engineers strive to identify flaws in these concepts, ameliorating them whenever possible, or perhaps even replacing them with entirely new concepts.

As one might expect, ambitious goals of this sort encounter serious practical difficulties. Foremost among these difficulties is what has come to be known as the *Implementation Problem* (Cappelen and Plunkett, 2020; Deutsch, 2020; Jorem, 2021; Gibbons, 2022; Löhr and Michel, 2023; Koch, forthcoming). This, roughly put, is the problem conceptual engineers face when attempting to bring about the conceptual change they endorse.[2] In order to change concepts in the way envisaged, many (though not all) proposals in conceptual engineering require large groups of agents to adopt them. But this is no simple task, and even if strictly philosophical considerations militate heavily in their favor, there is no guarantee that conceptual engineers will see their proposals implemented.

An important aspect of the Implementation Problem facing conceptual engineers concerns the extent to which their efforts will result in outcomes they neither intend nor desire. To implement their proposals, conceptual engineers are often required to intervene into large and complex systems constituted by concept-users with diverse aims, competences, and values. These *linguistic interventions*, much like other forms of large-scale intervention, can lead to unintended consequences that run directly counter to their ambitions.[3] Such consequences can of course impede their efforts, inconvenient obstacles hindering their progress. More worryingly, when sufficiently severe unintended

---

[1] For helpful overviews of many important issues related to conceptual engineering, see Cappelen (2018), Cappelen and Plunkett (2020), Isaac (2021), Koch (2021), Isaac, Koch, and Nefdt (2022), and Koch, Löhr, and Pinder (2023).

[2] Neufeld (forthcoming) refers to this problem as the *Feasibility Question* rather than the Implementation Problem. (See also Machery (2021)). I prefer "implementation problem" given its previous usage in the literature, but nothing hangs on this terminological point.

[3] See Löhr (2022) for earlier work focusing on ethical issues related to linguistic interventions and conceptual engineering. See also Sterken (2020).

consequences arise, conceptual engineers can find themselves in a worse situation than that in which they began, their ultimate goals now even further from their grasp. From a practical vantagepoint, then, thinking more carefully about the prospect of such unintended consequences is important for conceptual engineers: what sort of consequences can they expect? Are some forms of conceptual engineering more susceptible to unintended negative consequences than others? And how should one practice conceptual engineering, if at all, given that one's efforts may be met with consequences one doesn't intend?

In this paper I attempt to make some headway in answering these questions. I begin in Section 1 by first outlining several different forms of conceptual engineering, together with the kind of linguistic interventions each requires to implement the proposals in question. Next, in Section 2, I examine several types of *failure modes* of conceptual engineering—that is, ways in which efforts in conceptual engineering can lead to undesired and unintended consequences. In Section 3 I discuss some of the implications of the fact that attempted implementation can often produce consequences directly at odds with the intent of conceptual engineers. Among other things, I suggest that some forms of conceptual engineering are more likely to lead to unintended negative consequences than others, and that such forms of conceptual engineering are thereby less likely to succeed than other forms.[4] However, in Section 4, I outline some ways in which conceptual engineering is still worth pursuing, even given the prospect of unintended negative consequences.

Before moving on, some clarifications are in order. First, in what follows I focus on cases of conceptual engineering involving *social* or *political* aims. This is simply because, as a political philosopher, these are the cases with which I am most familiar. However, conceptual engineering without overt social or political aims (such as many efforts in academic or scientific contexts) can also

[4] Of course, similar claims have been defended by other philosophers (Chalmers, 2020; Koslow, 2022). However, as I show in Section 3, my analysis diverges from theirs in certain notable respects.

experience unintended consequences of various kinds, and the focus on social and political cases herein should not be taken to imply otherwise.[5] Second, and relatedly, this focus on social and political examples will inevitably lead me to discuss some heavily contested concepts—concepts such as VIOLENCE, HARM, and others—where not only do people typically feel very strongly one way or another, but there is often pressure to make one's first-order commitments about the correct way to engineer such concepts clear, lest one appear somehow unprincipled or otherwise untrustworthy.[6] Nevertheless, I do not take a stand on such issues in this paper, appearances be damned. The analysis that follows is intended to be purely *descriptive*. Claims to the effect that certain projects in conceptual engineering have faced (or will face) unintended negative consequences should be taken to imply neither opposition to nor support of these projects. The prospect of one's efforts leading to unintended consequences directly counter to one's aspirations is something people on all sides of these various disputes share.

1. Conceptual Engineering and Linguistic Intervention

Philosophers disagree about what it is we're doing when we engage in conceptual engineering, offering different accounts of the nature of concepts and what it means to engineer them.[7] In this paper I remain neutral about what the correct theory of conceptual engineering is. However, I assume that regardless of which theory one endorses, conceptual engineering in practice nearly always involves *linguistic interventions*.[8] Regardless of whether the final target of one's efforts is a word-meaning pair, a

[5] How exactly these unintended consequences arise and how they impact such forms of conceptual engineering are matters I leave for future investigation.
[6] See Silver and Shaw (2022) for some discussion of how failing to take sides on certain controversial issues can lead one to be viewed negatively by others.
[7] For some examples of this large and growing literature, see Plunkett and Sundell (2013), Simion and Kelp (2019), Chalmers (2020), Thomasson (2020), Löhr (2021), Nado (2021), Riggs (2021), Jorem (2022), Jorem and Löhr (2022), and Isaac (2023).
[8] See Chalmers (2020) and Koslow (2022) for similar remarks.

classification procedure, a mental representation, or something else entirely, one will invariably find concomitant proposals about words and how to use them.

This will prove important moving forward. The central contention of this paper, broadly put, is that projects in conceptual engineering can result in unintended consequences that are directly counter to the aims of the conceptual engineer in question. But there are different forms of conceptual engineering, each involving different kinds of linguistic intervention. Some of these forms, as we shall see, lead to unintended negative consequences more often than others. Before examining some of the ways in which projects in conceptual engineering can produce such consequences, then, it is necessary to first outline these different forms of conceptual engineering.

First, consider *conceptual amelioration*, which involves taking a pre-existing concept and engineering it to remove its perceived defects, provide it with new benefits, or both. For example, consider efforts to engineer the concept WOMAN in more trans-inclusive ways (Jenkins, 2016). Conceptual engineers pursuing such projects think that the existing concept of WOMAN unjustifiably excludes transwomen, and they propose to remedy this by engineering the concept such that it no longer excludes them.

Conceptual amelioration itself comes in different forms, depending on whether it involves *same-word* linguistic engineering or *different-word* linguistic engineering (Chalmers, 2020: 9-10). For instance, conceptual amelioration is often *competitively homonymous*, where a novel conceptual proposal will compete with the existing concept, the same word being used to express both concepts (Ibid, 10). Continuing with our previous example, trans-inclusive concepts of WOMAN are typically intended to replace previous concepts rather than exist alongside them. A degree of competition will therefore arise, since each concept is expressed with the same word.

Not all forms of homonymous conceptual engineering necessitate competition between different conceptual proposals, though, because conceptual engineers need not intend for their

proposals to replace existing concepts, even if they are expressed with the same word. Call such instances of conceptual engineering *noncompetitively homonymous*. For example, the concept of DOGWHISTLE that can be found in political contexts (referring, roughly, to a speech act which conventionally communicates one thing while also being used to tacitly communicate something else) is not intended to compete with the traditional concept of DOGWHISTLE, though both are expressed with the same word.[9] These concepts exist alongside each other with relative ease, with no conflict between them and no great difficulty understanding which concept is being used at any given time.

A final form of conceptual amelioration involves taking a concept and re-engineering it into "multiple strands", using different words to express each (Chalmers, 2020: 10). Since this form of conceptual amelioration involves different-word linguistic engineering, call it *heteronymous conceptual engineering*.[10] As an example, lexicalizing a distinct concept such as SEXUAL HARASSMENT constituted an advance over previous concepts of harassment *simpliciter* which, being insufficiently fine-grained, failed to pick out sexual harassment as a distinct kind of harassment.[11]

Two final forms of conceptual engineering differ from conceptual amelioration insofar as neither involves trying to improve or replace existing concepts. On the one hand, *de novo conceptual engineering* involves the creation of completely novel concepts, expressed by neologisms and the like (Chalmers, 2020: 6-9). For example, we might express the concept of a political arrangement that makes use of lotteries as selection mechanisms with the word "lottocracy" rather than using more roundabout expressions such as "democracy with lotteries rather than elections".[12] On the other hand, if a concept is deemed to be sufficiently defective, conceptual engineers might advocate for *conceptual*

[9] On dogwhistles, see Saul (2018).
[10] Cf. Chalmers (2020: 9-12).
[11] See Fricker (2007) for relevant discussion.
[12] See Guerrero (2014) for a classic outline and defense of lottocracy.

elimination rather than amelioration. Joshua Habgood-Coote, for example, has recently argued that we should abandon the concept FAKE NEWS because "fake news", the expression we use to express the concept, has no stable public meaning, doesn't provide us with the means to express anything we couldn't already express, and can serve as a vehicle for propaganda (Habgood-Coote, 2019).[13]

Though the details naturally vary from case to case, these five forms of conceptual engineering—competitively homonymous, noncompetitively homonymous, heteronymous, de novo, and elimination—each require some sort of linguistic intervention for the proposal to be implemented (or, in the case of conceptual elimination, for the target concept to be eliminated).

Both forms of homonymous conceptual engineering will require an existing expression to be used in novel, unfamiliar ways, with competitively homonymous conceptual engineering involving overt conflict with established patterns of usage. The latter may also involve criticizing or admonishing others for using the relevant expressions in ways inconsistent with the conceptual proposal in question, urging others to use the expressions in ways aligned with the proposal, and so on. Heteronymous conceptual engineering modifies an existing expression in some fashion, with the newly precisified expression being used in ways the unmodified expression was not (say, being used to refer to phenomena not previously picked out by the expression in question). De novo conceptual engineering requires the introduction of an entirely novel expression into a linguistic system, perhaps coupled with deliberate efforts to explain the concept and how to use it. Finally, conceptual elimination may require, *inter alia*, deliberately refraining from using certain expressions, advising others to refrain from using them, and so on.[14]

[13] For criticism of his arguments, see Pepp, Michaelson, and Sterken (2019) and Brown (2019). For a response, see Habgood-Coote (2022).

[14] One might also include concept *preservation* as a form of conceptual engineering, together with its own distinctive linguistic interventions (i.e., striving to use certain expressions even as their wider use declines, exhorting or otherwise incentivizing others to use them, and so on). See Lindauer (2020) for relevant discussion.

2. The Unintended Consequences of Linguistic Intervention

The sort of linguistic interventions that conceptual engineers rely upon to see their proposals implemented can lead to unintended consequences. But what sort of consequences can they expect? In this section I examine several types of failure modes related to (1) miscomprehension, miscommunication, and misuse; (2) conceptual exploitation; and (3) conceptual expansion and credibility deficits. Many of these failure modes have been discussed elsewhere in the literature on conceptual engineering, though not all together. Accordingly, this section is an attempt to provide a unified analysis of practical problems facing conceptual engineers that have heretofore been analyzed separately.

2.1. Miscomprehension, Miscommunication, and Misuse

Successful conceptual engineering often requires the cooperation of large numbers of people. A conceptual engineer might put forth a proposal with a certain goal in mind, and the best-case scenario involves their target audience not only sharing their ultimate goal, but also agreeing that the proposal they defend is an effective way to pursue that goal. But in order to agree with and subsequently adopt a proposal, one must first understand it.[15]

First, an inability to understand a novel proposal can prevent people from accepting it, even if they otherwise support the goal (or goals) it subserves. Rather than comprehending the proposal, some people experience frustration or confusion in response to linguistic interventions involving nonstandard uses of the relevant expressions (Stokoe, 2018: 44-5). Instead of achieving one's aims, linguistic interventions may just lead to widespread bewilderment—an unintended consequence, though admittedly one of which conceptual engineers are likely already aware. Such an outcome

[15] Note that I am not claiming that people must understand conceptual proposals in order to accept them. After all, people may simply defer to the conceptual proposals of others. But sincere agreement with the proposals (and subsequent acceptance) plausibly requires an understanding of them. For related discussion of semantic deference, see Pollock (2019).

appears more likely for homonymous conceptual engineering (both competitive and noncompetitive) than for other forms of conceptual engineering, with the existence of entrenched homonyms increasing the likelihood that the proposal in question leads to confusion (Chalmers, 2020: 12; Koslow, 2022: 94). But it is also likely a problem for other forms of conceptual engineering. The fine-grained distinctions characteristic of heteronymous conceptual engineering are easier to grasp for some than others, as are the entirely novel concepts found with de novo conceptual engineering. Perhaps even the exhortations to refrain from using certain concepts distinctive of conceptual elimination can be confusing for many. In short, people will not always be able to understand the proposals of conceptual engineers, and their reactions, rather than agreement and acceptance, may range from confusion to annoyance.

Another closely related problem arises when one attempts to engage in homonymous conceptual engineering. In much the same way that introducing novel concepts into a system with previously entrenched homonyms can sow confusion, this practice can also lead to miscommunication (Cappelen, 2018: 97-107; Traldi, 2024: 33). Consider a simple example. Suppose a political philosopher, Sam, wants to package her political views using an ameliorated concept of NEOLIBERALISM—roughly, she favors free markets, liberalism, constitutional democracy, and so on.[16] She is aware that she is dealing with a complicated and contested concept while still thinking that since she can trace the historical lineage of her views to paradigmatic neoliberal scholars, NEOLIBERALISM is the most appropriate concept to use. And this may not be an unreasonable thing to conclude. However, the scholarly credentials of her views cannot change the fact that the expression 'neoliberalism' is a "controversial, incoherent, and crisis-ridden term", one that is used in many ways by many people (Venugopal, 2015). In addition to the possibility that she might simply

[16] Presumably one would want to make further precisifications, not least in relation to other concepts such as LIBERALISM and DEMOCRACY which are themselves often heavily contested. I set aside such complications here. On neoliberalism more generally, see Vallier (2022).

confuse many people by attempting to ameliorate an expression already in circulation, her efforts might inadvertently lead to much miscommunication, with people assuming she endorses the views of others who use the expression in different ways, subsequently (mis)informing others about the content of her views, and so on.

Half-truths and outright misconceptions about her views may eventually disseminate throughout her target audience, rather than a clear understanding of her views prevailing; and these half-truths and misconceptions may shape how people use the concept moving forward, with people using it in ways that pull apart from her intentions. Assuming she wanted her views to be widely understood (or perhaps even to influence public policy), she might have frustrated her goals by choosing to *ameliorate* a pre-existing concept rather than simply *using* pre-existing concepts to communicate the content of her views. Of course, one's views can be misunderstood even if one diligently restricts oneself to using pre-existing concepts. Still, attempts to ameliorate already entrenched concepts will often exacerbate such problems.

2.2. Conceptual Exploitation

The problems outlined in the previous section arise simply because people possess different abilities to reason with (and correctly use) novel concepts. But it is important to note that such problems would often be even more severe if the target audience in question were prone to politically motivated reasoning, partisanship, and the like.[17] Political concepts (such as NEOLIBERALISM, CAPITALISM, SOCIALISM, COMMUNISM, and so on) often possess lexical effects and prior associations that can bias some people against views linked to them, such that people's ability to fairly engage with novel conceptual proposals involving these concepts is compromised.[18] In such conditions, people may not

[17] On politically motivated reasoning, see Lodge and Taber (2013).
[18] On lexical effects, see Cappelen (2018: 122-34).

competently engage with conceptual proposals even if they are in principle capable of doing so, their unwillingness to correctly use the novel concept yet another manifestation of their political irrationality.

Such political biases pose impediments to successful implementation in other ways. A conceptual engineer might introduce a concept with a certain goal in mind—perhaps they want to draw attention to a certain phenomenon that previously lacked a lexicalized concept—only to find that an uncooperative audience *deliberately* uses the concept incorrectly to further their own ends.[19]

Consider a concept such as DISINFORMATION, which is often introduced specifically to pick out *intentionally* false or misleading information (Benkler, Faris, and Roberts, 2018: 6; Brown, 2021: 3). It is important to know whether false or misleading information has been produced intentionally or not, since, among other things, efforts to curb the intentional distribution of false or misleading information require quite different methods from efforts to curb the unintentional distribution of false or misleading information. Accordingly, DISINFORMATION constitutes a relatively useful piece of heteronymous conceptual engineering.

Still, there are potential costs to the introduction of such concepts. Its introduction into a system of concept-users does not only allow good-faith, competent people to reliably pick out intentionally false or misleading information; it also provides a tool for self-interested political actors who hope to discredit others, regardless of whether its application is accurate or not.[20] Instead of being pressed into the service for which it was intended, it can instead be used as a sort of linguistic cudgel to harm one's political or ideological opponents. In short, concepts like DISINFORMATION (and MISINFORMATION, MALINFORMATION, and the like) can be used as intended, but they can

[19] Two comments are in order. First, by foregrounding this possibility, I do not mean to suggest that conceptual engineers themselves are incapable of such self-interested behavior. (See Shields (2021) for relevant discussion.) Second, I also do not mean to suggest that one cannot *correctly* use such concepts to further one's own ends. However, in what follows I primarily have in mind cases where people deliberately misapply the relevant concept.
[20] See Gibbons (2023) for relevant discussion.

also be *exploited* (Shields, 2021). And the greater the extent to which such concepts are exploited, the more their utility in political discourse diminishes, since people aware of the blatant exploitation may take applications of them to be insincere or inaccurate.

As with the prospect of conceptual engineering leading to widespread miscommunication, it is important to note that previously existing concepts can also be exploited by people looking to further their own ends. Such problems are not unique to the proposals of conceptual engineers. Neither are they unique to heteronymous conceptual engineering, as the above example might suggest. Any conceptual proposal involving the introduction of a novel concept could, in principle, encounter such problems. Still, conceptual engineers should at least proceed with caution, lest they provide powerful new conceptual tools to those who are willing to exploit them. As the preceding example illustrates, their efforts to create new concepts allowing us to more finely distinguish between important phenomena can instead have the unintended result of making it *more difficult* to distinguish between them, while at the same time making it easier for insincere actors to smear their opponents.

2.3. Concept Expansion and Credibility Deficits

The greater the extent to which concepts are known (or merely believed) to be exploited, the more cynical people will become about the ways in which they are being applied. And when this occurs, it becomes more difficult for sincere people aiming to apply the concepts accurately to do so credibly. The misconduct of some concept-users harms the communicative capabilities of others.

However, this form of diminished credibility can arise for reasons unrelated to conceptual exploitation. To see how, consider recent discussions of a phenomenon that has come to be known as concept creep (Haslam, 2016). Very roughly, this refers to a notable trend where concepts such as HARM, VIOLENCE, TRAUMA (and more) are increasingly being applied both to new phenomena they were not previously applied to, as well as to less extreme forms of phenomena that more naturally

fall under the scope of the relevant concept. For example, certain experiences not typically thought of as traumatic—such as having certain speakers visit one's campus—may increasingly be seen as contributing to or causing trauma. For another, certain forms of speech may be thought of as constituting (and not merely causing) violence, even though typical concepts of VIOLENCE involve the use of physical force to injure or kill others.

Even setting aside conceptual exploitation of the sort discussed earlier, there are many reasons why conceptual engineers and others might want to expand concepts in this manner, with the reasons often depending on the concept in question. An expanded concept of VIOLENCE might allow one to foreground the perceived severity of certain acts that fall short of paradigmatic physical violence.[21] An expanded concept of HARM might better enable regulation of certain behaviors that one thinks ought to be regulated. And so on for efforts to expand other concepts.

Predictably, though, efforts to expand concepts in this way can backfire in ways conceptual engineers do not intend. People who expand concepts like TRAUMA and VIOLENCE operate in an environment populated by others who largely possess less expansive concepts. For most others, the experience of having a speaker one disagrees with on campus is not traumatic, even if it is aggravating or upsetting.[22] Likewise, for most others, speech alone cannot constitute violence, even if it can be hurtful or offensive. In effect, there are large *conceptual mismatches* between people who seek to expand concepts like TRAUMA, VIOLENCE, and HARM (or who already use expanded versions of those concepts), and people who do not seek to expand such concepts and do not want to use them.[23]

---

[21] For instance, see Tirrell (2012) on *linguistic* violence.

[22] This is not to suggest that something like this could never be traumatic even by the standards of more typical, less expansive concepts of TRAUMA. Much would depend on the precise details of any given case.

[23] Some of the latter group may even want to *contract* the relevant concepts, such that, for example, actions or events commonly thought of as violent no longer fall under the scope of the concept, experiences commonly thought of as traumatic are no longer viewed as traumatic, and the like. But such efforts would face similar problems to efforts to expand those concepts. Conceptual contraction, much like conceptual expansion, can lead to unintended and undesired consequences.

Such conceptual mismatches can frustrate the goals of those who seek to expand certain concepts. For instance, people may view the application of TRAUMA to events they believe are relatively mild as exaggerated, or perhaps even flippant and unserious.[24] Perhaps more worryingly, novel ways of using the relevant expressions are sometimes viewed as an abuse of language.[25] Such reactions can, in turn, diminish the perceived credibility of those putting forth the relevant conceptual proposals. They may experience subsequent difficulties having their claims taken seriously, even if not involving expanded, novel conceptual proposals. Applications of the relevant concepts that otherwise would have been acceptable to others may be dismissed, given what others see as their unreliable track record in applying these concepts.[26] Overall, their conceptual proposals, widely viewed in a negative light, diminish their credibility more generally.

It is no coincidence that the above examples involve competitively homonymous conceptual engineering. Such proposals seem especially prone to eliciting reactions of the sort outlined above. Many people, whether reasonably or not, are predisposed to be skeptical of novel proposals that aim to displace existing concepts (Cappelen, 2018: 5). With that said, dismissive attitudes towards the proposals of conceptual engineers, together with ongoing skepticism towards their claims and proposals more generally (whether they involve conceptual expansion or not), can arise even when conceptual engineers do not pursue competitively homonymous proposals. Accordingly, heteronymous conceptual engineering, de novo conceptual engineering, and perhaps even conceptual elimination may lead to conceptual engineers experiencing diminished credibility. As a result, far from

[24] Cf. Löhr (2022: 843).

[25] As an example of this sort of response, consider Suzanna Danuta Walters' comments in the aftermath of the scandal involving Rebecca Tuvel's article 'In Defense of Transracialism' (Tuvel, 2017). Walters wrote that "the idea that any article in a specialized feminist journal causes harm, and even violence…is a grave misuse of the term "harm"" (Walters, 2017). Commenting on the same scandal, José Luis Bermúdez claimed that "the concept of harm has been twisted beyond all recognition" (Bermúdez, 2017).

[26] See Case (2019) for relevant discussion.

achieving their intended goal, conceptual engineers might make it even more difficult for themselves to bring about the sort of change they desire.

3. The Consequences of Unintended Consequences

Attempting to implement conceptual proposals often requires linguistic interventions into complex systems, and these interventions can produce consequences directly counter to the aspirations of conceptual engineers. What are some of the implications of this for the practice of conceptual engineering?

First, a clear methodological implication is that conceptual engineers who are interested in implementing their conceptual proposals need to look at more than just the distinctively philosophical considerations weighing in favor (or against) any given conceptual proposal. In addition, they need to consider the likelihood that the proposal in question will lead to unintended negative consequences.

If taken seriously, this would alter the practice of conceptual engineering in at least two noteworthy ways. On the one hand, it suggests a clear role for empirical research alongside more traditional philosophical analysis of the comparative merits of competing conceptual proposals. After all, ascertaining the degree to which a conceptual proposal is likely to elicit unintended negative consequences is an empirical question, one to be addressed with empirical research. I expand upon this point in the following section. For now, it suffices to note that philosophically trained conceptual engineers interested in addressing such matters need to either undertake the relevant empirical research themselves or collaborate with others who are willing to do so. On the other hand, taking account of possible failure modes can impact our judgements about which conceptual proposals out of a range of alternatives are *all-things-considered* best. The distinctively philosophical merits of a proposal, such as whether it satisfies certain normative criteria, can be greater than those of a competing alternative, and yet the alternative can be all-things-considered preferable if the former is

overly susceptible to certain failure modes. In short, conceptual engineers may need to settle for philosophically sub-optimal proposals that have a better chance of achieving their goals than philosophically superior alternatives that are less likely to be successfully implemented without creating unintended negative consequences.

A closely related second implication concerns the degree to which the possibility of unintended negative consequences can alter the overall expected costs (and expected benefits) of attempting to implement conceptual proposals. Several philosophers have argued, roughly, that one should attempt to implement a conceptual proposal only if the expected benefits of doing so exceed the expected costs (Andow, 2021; Gibbons, 2022; Löhr, 2022). If this is correct, and if the emergence of unintended negative consequences renders the expected costs of attempted implementation greater than the expected benefits, then one ought not to attempt implementation. In short, the likelihood that one might frustrate (rather than further) one's goals by staging a linguistic intervention can militate against attempted implementation altogether.

One might question this conclusion for at least two reasons. First, one might think that the risk of harm is sufficiently low whenever those attempting to implement some proposal are motivated by good intentions (Catapang-Podosky, 2022: 8). Second, one might think that in some cases the expected costs of conceptual inaction—that is, failing to attempt to implement a conceptual proposal whatsoever—are greater than the expected costs of attempted implementation, even given the likelihood of attempted implementation leading to unintended negative consequences. So, even if attempting implementation threatens higher costs than benefits, it can still be preferable to not attempting implementation whatsoever when the former is the least bad option.

Regarding the first point, it is important to note that all of the failure modes outlined earlier can emerge even if those attempting to implement conceptual proposals have good intentions: the possibility of miscomprehension and miscommunication arises not because of any ill intent on the

part of conceptual engineers, but because some people will not fully understand the proposal in question; conceptual exploitation, though often related to the ill intent of those who wish to use novel concepts for their own gain, need not be the result of ill intent on the part of the conceptual engineers responsible for the concept; and conceptual mismatches between conceptual engineers and their target audiences can ultimately lead to diminished credibility regardless of the intent of the former. Overall, then, unintended negative consequences can emerge as a result of linguistic interventions even if those intervening are well-intentioned.

Regarding the second, this point ought to be conceded. It is possible that, in some cases, an unwillingness to attempt implementation whatsoever can lead to worse outcomes than attempting to implement some proposal, even if the latter option would lead to unintended negative consequences. Predictably, much here depends on the details of each case. With that said, while the prospect of unintended negative consequences should not be taken to militate entirely against attempting implementation in every case, it should make conceptual engineers and conceptual activists more broadly much more *cautious* about their efforts. Linguistic interventions can not only fail, but they can seriously backfire, making it harder to achieve one's goals. It is worth considering such risks and, if possible, taking steps to mitigate them before attempting to implement conceptual proposals.

Another implication of the unintended negative consequences of linguistic interventions concerns the degree to which different forms of conceptual engineering are adversely impacted. Although all forms of conceptual engineering outlined earlier can lead to unintended negative consequences, it appears that homonymous conceptual engineering is especially likely to produce such outcomes. Homonymous conceptual engineering is especially likely to lead to widespread confusion and miscommunication, for instance, and efforts to expand (or contract) concepts already in circulation are especially likely to lead to diminished credibility.

Other philosophers have also noted such features of homonymous conceptual engineering. For instance, David Chalmers writes that "[homonymous] conceptual engineering can lead to confusion with all the multiple meanings floating around" (Chalmers, 2020: 12). Elsewhere, in a fascinating paper drawing upon research in historical linguistics and diachronic semantics, Allison Koslow discusses "the fact that languages have a well-documented tendency to avoid homonymy, in particular confusing homonymy", noting the constraints this places on the feasibility of conceptual engineering (Koslow, 2022: 11). The account offered in this paper partly explains the difficulties of homonymous conceptual engineering such philosophers have discussed. Indeed, the account offered in this paper suggests that, if anything, the difficulties of homonymous conceptual engineering have previously been understated. After all, such difficulties arise not solely due to the confusion they can create, but also because of the role homonymous conceptual proposals play in people downgrading the credibility they assign to conceptual engineers, as well as the possibility that these proposals can be exploited (though, of course, they share this feature with other forms of conceptual engineering).

However, my account of the practical problems created by unintended negative consequences also notably diverges in certain ways from previous accounts. First, consider some of Chalmers' comments regarding the potential upshots of engaging in homonymous conceptual engineering relative to alternatives such as heteronymous or de novo conceptual engineering. While correctly noting that it is often more difficult to get people to use new concepts, he also suggests that "old words are entrenched and often have an associated prestige" (Chalmers, 2020: 11). Tying a new concept to an old word with prestige of this kind can, in principle, help one implement one's proposal. However, we have seen that the prior associations of entrenched words—particularly words expressing political concepts—can in fact bias people against novel conceptual proposals linked to them. In such cases, expressions have something akin to a bad reputation rather than anything we might characterize as prestige. At the very least, potential upshots of homonymous conceptual are

heavily offset by the fact that entrenched expressions can come with baggage that makes it more difficult to implement one's conceptual proposal. Of course, homonymous conceptual engineering also encounters *other* problems, at least one of which—its tendency to lead to confusion—is acknowledged by Chalmers. Overall, then, it is plausible that in many cases the upshots of homonymous conceptual engineering are overshadowed by its downsides.

Second, Koslow offers a measured defense of conceptual engineering strategies that involve the gradual *broadening* or *narrowing* of some expression over time (Koslow, 2022: 19). It is uncontroversial that expressions in natural language often broaden (or narrow) gradually over time. For example, the expression "aunt" used to refer exclusively to one's father's sister, before broadening over time to include either parent's sister (Ibid, 19). For another, the expression "accident" used to refer to any sort of happenstance before narrowing over time to refer only to "unfortunate happenstances" (Ibid, 18). Perhaps, the suggestion goes, conceptual proposals that involve gradual broadening or narrowing of this sort are more feasible than others. However, conceptual engineers opting for such strategies ought to first consider the risks of doing so, for we have already seen that efforts to expand or contract concepts can not only fail to achieve their intended purpose, but even diminish their credibility more broadly.[27] Once again, homonymous conceptual engineering seems more seriously affected by such issues than other forms of conceptual engineering, though we earlier noted that any sort of conceptual engineering could result in the diminished credibility of the conceptual engineer in question, depending on the details of the case.

Generally speaking, then, although all forms of conceptual engineering can result in unintended negative consequences, homonymous conceptual engineering (and perhaps especially competitively homonymous conceptual engineering) is plausibly the form of conceptual engineering

[27] It should be noted that Koslow also recognizes the limitations of such strategies, writing that "few engineering proposals with a revolutionary flavor will involve broadening or narrowing in this sense (Koslow, 2022: 19).

most likely to result in such consequences. All else being equal, then, conceptual engineers ought to consider other forms of conceptual engineering to achieve their goals, whenever this is possible.

4. Conceptual Engineering Despite Unintended Negative Consequences

We have reached some moderately pessimistic conclusions about efforts to implement conceptual proposals given the risk of unintended negative consequences. Since attempting to implement conceptual proposals can result in outcomes directly counter to the goals of conceptual engineers, the expected costs of attempted implementation may just exceed the expected benefits (though, as we also noted, that we should opt for conceptual inaction is not entailed by this fact). In addition, the problems one can encounter when attempting to implement proposals are especially pronounced for homonymous forms of conceptual engineering, suggesting that, all else being equal, one should eschew such forms in favor of others if possible. However, in this concluding section, I briefly discuss some of the ways in which the practice of conceptual engineering is still worthwhile, even given the risk of unintended negative consequences.

To begin, as Steffen Koch persuasively argues in a recent paper, conceptual engineering need not always involve efforts to implement conceptual proposals (Koch, forthcoming). Conceptual engineering is standardly thought of as being comprised by three separate though related activities: the assessment and evaluation of existing concepts; the design of novel concepts; and the implementation of novel conceptual proposals (Cappelen and Plunkett, 2020; Chalmers, 2020). Even if the latter were rendered entirely infeasible by the unfortunate possibility that efforts to implement conceptual proposals can seriously backfire, this does not entail the infeasibility of the other activities that conceptual engineers engage in. As Koch writes, "[assessing] and designing concepts are intellectual activities that can be pursued in separation from the activism required for implementation" (Koch,

forthcoming: 8). Consequently, even if *implementation* can be threatened by the prospect of unintended negative consequences, the other core activities of conceptual engineering are plausibly less so.

Of course, these other activities may themselves lead to unintended negative consequences. The assessment and evaluation of concepts, for instance, involves the possibility of assessing and evaluating concepts *incorrectly*, inclining us towards proposals that we would be better off rejecting—an unintended negative consequence of sorts. However, if there are risks involved with such activities, they can be much more narrowly contained relative to efforts to implement concepts. "One might argue", Koch explains, "for a particular revisionary definition while refraining from advertising its use to one's broader community" (Ibid, 8). This notably differs from efforts to implement conceptual proposals which, in order to achieve their intended effect, must be advertised in a suitably broad manner. Additionally, there are risks associated with *failing* to assess and evaluate our concepts. Most obviously, an unwillingness to critically reflect on one's concepts—what Hermann Cappelen calls *representational complacency*—can leave one stuck with harmful and defective concepts when it would be better to seek out alternatives.[28]

One might grant that core activities of conceptual engineering are viable while still reasonably wondering what should be done about projects in conceptual engineering where implementation is sought. One of the central lessons of the previous section is that such projects are indeed threatened by the prospect of unintended negative consequences. But though conceptual engineers should proceed more cautiously when attempting to implement their proposal on a large scale, successful implementation is not precluded entirely.

First, some proposals may be easily and safely implementable, depending on the size of one's target audience, the degree to which their values are aligned with those of the conceptual engineer in question, and so on. In a recent paper, Gibbons writes that "ameliorative projects vary in scale. In

[28] On representational complacency, see Cappelen (2018: 5-7). See also Belleri (2023) for related discussion.

some cases, conceptual engineers will be satisfied once a relatively small community of concept-users adopts the ameliorative proposals, while in others much larger numbers will be involved" (Gibbons, 2022: 3). In a similar vein, Pinder notes that different projects in conceptual engineering have "different success conditions (Pinder, 2022: 334). The success conditions of some projects require the sort of large-scale linguistic interventions discussed throughout this paper, but this is not a feature shared by all projects in conceptual engineering. And though more modest projects may also lead to unintended negative consequences, they are plausibly more tractable than much more ambitious projects.

Second, although implementation is often extremely difficult, it is not impossible. Some projects in conceptual engineering, even some rather ambitious projects, have been somewhat successful (Cantalamessa, 2021; Isern-Mas, 2023). Accordingly, conceptual engineers could empirically investigate the preconditions for success in conceptual engineering, whether as practiced by philosophers or by other groups. Such empirical investigation may include engaging in case studies of previously successful projects in conceptual engineering in order to determine what it was about them that was conducive to success, as well as more closely examining cases of failure in conceptual engineering. For example, an investigation of relevant cases may vindicate skepticism about homonymous conceptual engineering, showing that they very often backfire. But perhaps these cases studies could also show that such disadvantages can be overcome provided that the right *strategies* are deployed, the right *audience* selected, the right *type of concept* chosen, and so on. In short, engaging in case studies could provide conceptual engineers with better evidence about the feasibility of their proposals in advance of undertaking them, allowing them to avoid pursuing projects that could lead to undesirable consequences counter to their goals.[29]

[29] Other philosophers have also commented on the possible usefulness of case studies. For example, see Gibbons (2022: 19-20) and Löhr (2022: 845). See also Landes (forthcoming) for related discussion.

5. Conclusion

Conceptual engineers face a well-known *Implementation Problem*. An important aspect of this problem is the fact that efforts to implement conceptual proposals can not only fail, but also lead to unintended negative consequences directly counter to the goals of conceptual engineers. Indeed, these consequences might move their goals even further from their grasp. Instead of implementing their proposals, the linguistic interventions they use might just lead to widespread confusion and miscommunication. Instead of securing the benefits they expect from implementing their proposal, they might find the novel concept they put forth exploited by others pursuing their own narrow self-interests. Instead of acceptance, they may find their credibility diminished, their proposals dismissed as misguided.

Although such problems can arise for all forms of conceptual engineering, some forms fare worse in this regard. Specifically, homonymous conceptual engineering—conceptual engineering that involves same-word engineering—is especially likely to produce unintended negative consequences. All else being equal, homonymous forms of conceptual engineering should be eschewed in favor of other forms. With that said, it may be the case that, given the possibility of creating unintended negative consequences, the expected costs of attempting to implement a conceptual proposal exceed the expected benefits, regardless of the form of conceptual engineering in question. Conceptual engineers should thus approach their task with caution.

Still, it does not follow that we ought to abandon conceptual engineering entirely. On the one hand, other activities central to conceptual engineering—activities such as the assessment and evaluation of concepts, as well as the design of novel concepts—are much less threatened by the prospect of unintended negative consequences. Such activities are still worthwhile, even if implementation can face serious problems. On the other hand, even projects in conceptual

engineering that require implementation can still be viable. Some projects, for instance, do not require the sort of linguistic interventions most likely to elicit unintended negative consequences. But more generally, conceptual engineers could engage in case studies of both successful and unsuccessful examples of conceptual engineering to determine what it is that makes the former succeed where the latter fail. Conceptual engineering can be difficult—sometimes extraordinarily difficult—but it can still be valuable, whether implementation is sought or not.

References

Andow, J. (2021) 'Conceptual Engineering is Extremely Unlikely to Work: So What?', *Inquiry: An Interdisciplinary Journal of Philosophy* 64 (1-2): 212-26

Belleri, D. (2023) 'Conceptual Engineering Between Representational Skepticism and Complacency: Is There a Third Way?', *Topoi* https://doi.org/10.1007/s11245-023-09924-3

Benkler, Y., Faris, R., and Roberts, H. (2018) *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford, UK: Oxford University Press

Bermúdez, J.L. (2017) 'Defining 'Harm' in the Tuvel Affair', *Inside Higher Ed*

Brown, E. (2019) '"Fake News" and Conceptual Ethics', *Journal of Ethics and Social Philosophy* 16 (2): 144-54

Brown, E. (2021) 'Regulating the Spread of Online Misinformation', in M. Hannon and J. de Ridder (eds.) *The Routledge Handbook of Political Epistemology*. New York, NY: Routledge

Cantalamessa, E.A. (2021) 'Disability Studies, Conceptual Engineering, and Conceptual Activism', *Inquiry: An Interdisciplinary Journal of Philosophy* 64 (1-2): 46-75

Cappelen, H. (2018) *Fixing Language: An Essay on Conceptual Engineering*. Oxford, UK: Oxford University Press

Cappelen, H. and Plunkett, D. (2020) 'A Guided Tour of Conceptual Engineering and Conceptual Ethics', In H. Cappelen, D. Plunkett, & A. Burgess (eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 1-26). Oxford, UK: Oxford University Press

Case, S. (2019) 'The Boy Who Inflated the Concept of 'Wolf'', *Quillette*

Chalmers, D. (2020) 'What is Conceptual Engineering and What Should it Be?', *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/0020174X.2020.1817141.

Deutsch, M. (2020) 'Speaker's Reference, Stipulation, and a Dilemma for Conceptual Engineers', *Philosophical Studies* 177 (12): 3935–3957

Fischer, E. (2020) 'Conceptual Control: On the Feasibility of Conceptual Engineering', *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/0020174X.2020.1773309

Fricker, M. (2007) *Epistemic Injustice: The Power and Ethics of Knowing*. Oxford, UK: Oxford University Press

Gibbons, A.F. (2022) 'Rational Conceptual Conflict and the Implementation Problem', *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/002174X.2022.2037455

Gibbons, A.F. (2023) 'Bullshit in Politics Pays', *Episteme*. https://doi.org/10.1017/epi.2023.3

Guerrero A (2014) Against Elections: The Lottocratic Alternative. *Philosophy and Public Affairs* 42: 135-78

Habgood-Coote, J. (2019) 'Stop Talking About Fake News!', *Inquiry: An Interdisciplinary Journal of Philosophy*. 62 (9-10): 1033-65

Habgood-Coote, J. (2022) 'Fake News, Conceptual Engineering, and Linguistic Resistance: Reply to Pepp, Michaelson, and Sterken, and Brown', *Inquiry: An Interdisciplinary Journal of Philosophy* 65 (4): 488-516

Haslam, N. (2016) 'Concept Creep: Psychology's Expanding Concepts of Harm and Pathology', *Psychological Inquiry* 27 (1): 1-17. https://doi.org/10.1080/1047840X.2016.1082418

Isaac, M.G. (2021) 'What Should Conceptual Engineering Be All About?', *Philosophia* 5: 1-13

Isaac, M.G. (2023) 'Which Concept of Concept for Conceptual Engineering?', *Erkenntnis* 88: 2145-69

Isaac, M.G., Koch, S., and Nefdt, R. (2022) 'Conceptual Engineering: A Road Map to Practice', *Philosophy Compass*. https://doi.org/10.1111/phc3.12879

Isern-Mas, C. (2023) 'Implementing Conceptual Engineering: Lessons from Social Movements', *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/0020174X.2023.2253274

Jenkins, K. (2016) 'Amelioration and Inclusion: Gender Identity and the Concept of *Woman*', *Ethics* 126 (2): 394-421

Jorem, S. (2021) 'Conceptual Engineering and the Implementation Problem', *Inquiry: An Interdisciplinary Journal of Philosophy*. 64 (1-2): 186-211

Jorem, S. (2022) 'The Good, The Bad, and the Insignificant—Assessing Concept Functions for Conceptual Engineering', *Synthese*. 200 (2): 1-20

Jorem, S. and Löhr, G. (2022) 'Inferentialist Conceptual Engineering', *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/0020174X.2022.2062045

Koch, S. (2021) 'Engineering What? On Concepts in Conceptual Engineering', *Synthese*. 199: 1955-75

Koch, S. (forthcoming) 'The Anti-Conceptual Engineering Argument and the Problem of Implementation', *American Philosophical Quarterly*

Koch, S., Löhr, G., and Pinder, M. (2023) 'Recent Work in the Theory of Conceptual Engineering', *Analysis*. https://doi.org/10.1093/analys/anad032

Koslow, A. (2022) 'Meaning Change and Changing Meaning', *Synthese* 200 (94): 1-26. https://doi.org/10.1007/s11229-022-03563-8

Landes, E. (forthcoming) 'How Language Teaches and Misleads: "Coronavirus" and "Social Distancing" as Case Studies', in M.G. Isaac, K. Scharp, and S. Koch (eds.) *New Perspectives on Conceptual Engineering*

Lindauer, M. (2020) 'Conceptual Engineering as Concept Preservation', *Ratio* 33 (3): 155-62

Löhr, G. (2021) 'Commitment Engineering: Conceptual Engineering Without Representations', *Synthese* 199 (5-6): 13035-52

Löhr, G. (2022) 'Linguistic Interventions and the Ethics of Conceptual Disruption', *Ethical Theory and Moral Practice*. https://doi.org/10.1007/s10677-022-10321-9

Löhr, G. and Michel, C. (2023) 'Conceptual Engineering, Predictive Processing, and A New Implementation Problem', *Mind & Language*. https://doi.org/10.1111/mila.12471

Lodge, M. and Taber, C.S. (2013) *The Rationalizing Voter*. Cambridge, UK: Cambridge University Press

Machery, E. (2021) 'A New Challenge to Conceptual Engineering', *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/0020174X.2021.1967190

Nado, J. (2021) 'Classification Procedures as the Targets of Conceptual Engineering', *Philosophy and Phenomenological Research*. 106 (1): 136-56

Neufeld, E. (forthcoming) 'Engineering Social Concepts: Lessons from the Science of Categorization', in S. Haslanger, K. Jones, F. Schroeter, and L. Schroeter (eds.) *Social Hierarchy*. Oxford, UK: Oxford University Press

Pollock, J. (2019) 'Conceptual Engineering and Semantic Deference', *Studia Philosophica Estonica* 12: 81-98

Pepp, J., Michaelson, E., and Sterken, R. (2019) 'Why We Should Keep Talking About Fake News', *Inquiry: An Interdisciplinary Journal of Philosophy* 65 (4): 471-87

Pinder, M. (2022) 'Is Haslanger's Ameliorative Project a Successful Conceptual Engineering Project?', *Synthese*. https://doi.org/10.1007/s11229-022-03803-x

Plunkett, D, and Sundell, T. (2013) 'Disagreement and the Semantics of Normative and Evaluative Terms', *Philosopher's Imprint* 13 (23): 1-37

Podosky, PM.C. (2022) 'Can Conceptual Engineering Actually Promote Social Justice?', *Synthese* 200 (160): 1-22. https://doi.org/10.1007/s11229-022-03469-5

Riggs, J. (2021) 'Deflating the Functional Turn in Conceptual Engineering', *Synthese* 199 (3-4): 11555-86

Saul, J. (2018) 'Dogwhistles, Political Manipulation, and Philosophy of Language', in D. Fogal, D.W. Harris, and M. Moss (eds.) *New Work on Speech Acts*. Oxford, UK: Oxford University Press

Silver, I. and Shaw, A. (2022) 'When and Why "Staying Out Of it" Backfires in Moral and Political Disagreements', *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001201

Simion, M. and Kelp, C. (2019) 'Conceptual Innovation, Function First', *Nous* 54 (4): 985-1002

Shields, M. (2021) 'Conceptual Domination', *Synthese* 199: 15043-67

Sterken, R. (2020) 'Linguistic Interventions and Transformative Communicative Disruptions', In H. Cappelen, D. Plunkett, & A. Burgess (eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 417-34). Oxford, UK: Oxford University Press

Stokoe, E. (2018) *Talk: The Science of Communication*. London, UK: Robinson

Thomasson, A. (2020) 'A Pragmatic Method for Normative Conceptual Work', In H. Cappelen, D. Plunkett, & A. Burgess (eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 435-58). Oxford, UK: Oxford University Press

Tirell, L. (2012) 'Genocidal Language Games', in I. Maitra and M. McGowan (eds.) *Speech and Harm: Controversies over Free Speech* (pp. 174-221). Oxford, UK: Oxford University Press

Traldi, O. (2024) *Political Beliefs: A Philosophical Introduction*. New York, NY: Routledge

Tuvel, R. (2017) 'In Defense of Transracialism', *Hypatia* 32 (2): 263-78

Vallier, K. (2022) 'Neoliberalism', *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/archives/win2022/entries/neoliberalism/

Venugopal, R. (2015) 'Neoliberalism as Concept', *Economy and Society* 44 (2): 165-87

Walters, S.D. (2017) 'Academe's Poisonous Call-Out Culture', *The Chronicle of Higher Education*