Abstract

This paper discusses the logical space of alternative conceptual engineering projects, with a specific focus on (1) the processes, (2) the targets and goals, and (3) the methods of such projects. We present an overview of how these three aspects interact in the contemporary literature and discuss those alternative projects that have yet to be explored based on our suggested typology. We show how choices about each element in a conceptual engineering project constrain the possibilities for the others, thereby giving rise to distinct groupings of possible projects under the banner of conceptual engineering. We conclude with a critical reflection on the potential ethical issues that arise as a result of effectively putting conceptual engineering into practice.

1 | INTRODUCTION

Conceptual engineering is a branch of philosophy concerned with the process of assessing and improving our concepts. It is motivated by the fact that, sometimes, our conceptual schema must be ameliorated to attain certain beneficial consequences, which may be social, theoretical, political, or otherwise. Examples of conceptual engineering can be found across the disciplines of science, law, and policymaking, for instance. As an example, when the International Astronomical Union voted to approve a new definition of "planet" in 2006, they replaced the corresponding concept to provide a more principled taxonomy of celestial bodies that reflected recent observational discoveries. Like Mr. Jourdain discovering, in The Middle-Class Gentleman by Molière, that he had been speaking prose for more than forty years without knowing anything about it, conceptual engineering might thus be something "we have been doing all along without realizing it" (Thomasson, 2021, p. 2). However, contemporary philosophical discussions of conceptual engineering theorize it to be a distinctive method, and philosophers explicitly provide methodological guidance about its practice. Conceptual engineering has now become a burgeoning field in analytic philosophy.¹

The term "conceptual engineering" was independently coined in Carnap scholarship (Carus, 2007; Creath, 1991) and metaphilosophy (Blackburn, 1999; Floridi, 2011).² Recently, these two strands of research have grown together...
(Brun, 2016) and have expanded further into what is known as “ameliorative analysis,” which developed in parallel within social philosophy (Haslanger, 2012).

There are two main aspects to current research in conceptual engineering. First, the theory of conceptual engineering reflects on conceptual engineering qua method and addresses its foundational issues (e.g., Burgess et al., 2020; Cappelen, 2018; Isaac & Koch, 2022). For instance, theorists interested in these issues ask questions such as: should conceptual engineering operate on concepts, words, or some other representational device? How does conceptual engineering fit into an overall theory of mind and language? What is the relationship between conceptual engineering and other philosophical methods? How, if at all, can the proposals made by conceptual engineers be implemented? How radically can a concept be revised before simply becoming another, different concept?

Second, the practice of conceptual engineering targets concepts that require improvement and advocates for specific ameliorations. For instance, deeming the concept of truth to be inconsistent, Scharp (2013) has proposed replacing it with two surrogate concepts, ascending truth and descending truth, to do natural language semantics for expressly rich languages. Similarly, Haslanger (2000) has proposed that, to combat social injustice, we should revise the concept of woman to signify, roughly, a person systematically subordinated based on perceived or imagined female bodily features.

Drawing on these two aspects, this article aims to provide a road map by which conceptual engineering can be put into practice (see Appendix for a synopsis). We begin by discussing the process of conceptual engineering and asking what “engineering” means (Sect. 2). Next, we present the various objects that can be the target of an engineering process and distinguish them from the goals one may pursue in such a project (Sect. 3). With these three parameters—process, target objects, and goals—in place, we outline the methods that are available to conceptual engineers and show how choices made with respect to each element constrain the possibilities for the next elements, thereby carving out distinct kinds of possible projects within the general field of conceptual engineering (Sect. 4). We conclude with a reflection on the potential ethical issues that may result from putting conceptual engineering into practice and how they may be overcome (Sect. 5).

2 THE ENGINEERING PROCESS

The recent excitement over conceptual engineering has been fomented by a number of factors. Presumably, chief among them is its “engineering” label. This term helps to draw a sharp contrast with the more traditional philosophical method of conceptual analysis. According to the dominant narrative, conceptual analysis is descriptive—simply describing the concepts we have and use—whereas conceptual engineering is prescriptive—prescribing the concepts we ought to have and how we ought to use them to suit specific aims (e.g., Nado, 2021b). This simplistic narrative has one significant consequence: it puts the ‘engineering’ classification front and center.

The ‘engineering’ label critically contributes to framing philosophy as a problem-solving enterprise. Problem-solving approaches, as we typically comprehend them, aim to create artifacts that serve human purposes, and to introduce them into our environment to fix antecedent problematic devices and detrimental behaviors. Thus, owing to the label, philosophy gains a “translational” (Machery, 2021, p. 2), use-inspired orientation: through conceptual engineering, philosophy aims to deliver workable solutions to tangible problems (cf. Blackburn, 1999, pp. 1–2).

To date, several different positions have been adopted vis-à-vis the ‘engineering’ label. Among those who consider conceptual engineering to be concerned with engineering concepts or other representational devices, we can distinguish two approaches. The first takes the label figuratively. In this view, the engineering process corresponds roughly to any ameliorative changes to the previous state of our representational devices. The advantage of this figurative reading of the label is that it makes ‘conceptual engineering’ a convenient umbrella term: it can cover, and thus regiment and make sense of, a wide variety of similar practices across philosophy and beyond (including, for instance, conceptual changes in science) that would otherwise remain disconnected. Thereby, the notion of conceptual engineering fills a hermeneutical gap in our methodological reflections and may even contribute to reframing...
extant ameliorative practices (Appiah, forthcoming). On the other hand, it is difficult to see why conceptual engineering deserves the attention it attracts if it is merely a new label for something that we have been doing all along, and managing pretty well without any need for a novel classification (cf. Deutsch, 2020a, 2020b, 2021). In addition, some have questioned whether any insightful generalization can be drawn from such an all-encompassing coinage (cf. Isaac, 2021a).

By contrast, a second approach takes the ‘engineering’ label literally. On this reading, the engineering process must be staged, with some evaluation criteria for ameliorating the previous state of our representational devices. In this regard, the reading can successfully accommodate the idea that conceptual engineers target representational objects qua devices—as tools serving certain functions or purposes—to increase their functional efficacy (Nado, 2021b). An additional benefit of the literal approach is that it makes “conceptual engineering” a term for a new methodological program for future applications, instead of a mere rebranding strategy.

Several existing accounts can serve to reconstruct the engineering process in the literal sense. The most basic of these can be called “componential” accounts, because they focus on the static identification and description of the basic components of the engineering process and do not consider the relationships between these components (Cappelen, 2017, 2018; Cappelen & Plunkett, 2020; Eklund, 2021). Typically, between two and four components of the process are identified, under different terminologies, the most common of which are assessment and improvement.

The assessment component consists of assessing the functional quality of a given representational device, or the degree to which it successfully fulfills its function. For instance, such an assessment took place when the IAU’s Planet Definition Committee “was charged with considering […] whether the current naming procedures for planets and minor planets have exacerbated the problem of defining a planet and whether revisions are needed” (Ekers, 2006, p. 4).

The improvement component, in turn, aims to improve the functional quality of the representational device under consideration. Typically, this is done by fixing deficiencies in how it fulfills its function. For instance, having established empirically that CONSPIRACY THEORY is a thick concept—comprising both a descriptive and an evaluative dimension—Napolitano and Reuter (2021) propose a change to our linguistic practice, which aims to improve its functioning with respect to that concept. To neutralize the negative connotations associated with the concept in those theoretical contexts when only the descriptive content is relevant, they suggest that the lexical form "conspiracy theory" be replaced with "conspiracy explanation."

Another commonly identified component of conceptual engineering is implementation, which deals with implementing the prescribed improvements on the ground, to secure their uptake in their target user groups. This is exemplified by the International Union for the Conservation of Nature, which regularly introduces new categories for species threatened by extinction in the Red List of Threatened Species, with the goal of biodiversity needs being factored into policy decision-making processes.

A final, often-overlooked, component is description, which delineates and regiments the target object of an engineering project. For instance, social philosophers draw on the sex–gender distinction to understand the difference between conceptions of disability as naturalized or socially constructed, and, furthermore, to challenge the prevailing naturalization of disability (Tremain, 2021). Neglected as it might be, the descriptive component is a crucial preparatory step for identifying the target of a conceptual engineering project (see endnote 4). We shall elaborate on this in Section 5.

In summary, then, the literature identifies four basic components of the conceptual engineering process: assessment, improvement, implementation, and description.

Against this backdrop, one might articulate a more sophisticated, “relational” conception of the engineering process, which investigates the relationships between the basic components identified above (Burgess & Plunkett, 2020). This account considers two basic relationships: unions and intersections. A simple understanding of conceptual engineering in terms of the union of its basic components is evidently too inclusive, because anything that would count as just one of these components would count as an instance of conceptual engineering. This would entail, for instance, that merely assessing some concept, with no attempt to later resolve its deficiencies, qualifies
as conceptual engineering. On the other hand, understanding conceptual engineering as the intersection of its basic components is too restrictive: that intersection is certainly very small, if not empty. Few of the philosophers who are considered paradigm conceptual engineers carry out all four tasks, including describing and assessing a concept, suggesting an improvement thereof and then also trying to implement the proposed change in the community.

As a consequence, the relational account instead rests on the thesis that conceptual engineering consists of the union of the intersections between its basic components. This means that, for instance, pairs such as description and assessment, assessment and improvement, or improvement and implementation each by themselves amount to partial instances of conceptual engineering. Ultimately, "prototypical/paradigmatic projects in conceptual engineering are exercises in conceptual [improvement], guided by conceptual [assessment], in the service of conceptual implementation" (Burgess & Plunkett, 2020, pp. 286, 291). This characterization leads to the fullest accounts of the engineering process available to date, which may be called "dynamic".

A direct precursor for dynamic accounts may be identified in the procedural reconstructions of Carnap’s method for explicating concepts (Carnap, 1950). The dynamic accounts order the relations between the four basic components as a staged process that is both non-linear and iterative (Andow, 2020; Appiah, forthcoming; Chalmers, 2020; Greenough, forthcoming). Notwithstanding slight variations in the sequence of stages between the different proposals, the general pattern is as follows:

1. Description
2. Evaluation
3. Improvement
4. Implementation.

Depending on the outcome of each stage, the stages need not be traversed in a linear manner: movement back and forth between stages is possible. The process may also be open-ended and iterative. One expected outcome of these dynamic accounts is that, once fully developed, they could render the entire engineering process controllable and scrutable. They might also help to regiment historical instances of like-minded projects (Appiah, forthcoming). However, one might rightfully doubt whether such an algorithmic treatment is applicable to philosophical issues.

3 | GOALS AND TARGETS

Now that we have clarified the process, let us turn to target objects: what is it that conceptual engineers are engineering? This question has recently received a fair amount of attention (Haslanger, 2020a, 2020b; Isaac, 2021a, c; Koch, 2021a; Machery, 2017; Richard, 2020; Sawyer, 2018, 2020a, 2020b, 2021; Scharp, 2013), in part because the notion of a concept is itself a candidate for re-engineering (Chalmers, 2020; Isaac, 2020; Scharp, 2020; Thomasson, 2020) and partly because a number of authors actively resist the idea that conceptual engineering has much to do with concepts at all (Cappelen, 2018; Flocke, 2020; Nado, 2020, 2021a; Pinder, 2021; Thomasson, 2021; see Nefdt, 2021 and Sawyer, 2020a for criticism). In what follows, we develop a pluralistic response to this issue. A key ingredient of our favored pluralism is the proper delineation of the goals and targets of conceptual engineering projects.

Conceptual engineering is inherently purpose oriented. Rarely do conceptual engineers suggest representational changes for their own sake. Instead, they most often consider such changes to be a means toward achieving a further purpose (Koch, 2021a; Nado, 2021a; Pinder, 2020b; Riggs, 2019). For instance, that purpose might be setting a consistent basis for truth-conditional natural language semantics (Scharp, 2013), or fostering emancipation from oppressive structures (Haslanger, 2000). However, when we speak of the goals of conceptual engineering, we do not have such general purposes in mind. Rather, we mean the specific cognitive or linguistic difference that conceptual engineers aim to make with their proposed changes, and the relation between that difference and their overall
purposes. For illustrative purposes, suppose that a conceptual engineer proposes to change the truth-conditional content of "woman" because she takes this change to contribute to the advancement of social justice. For her project, furthering social justice is the purpose, whereas changing the truth-conditional content of 'woman' is the goal.

Once we abstract away from the content of concrete engineering projects—like their concern with woman or truth—we can distinguish between different kinds of goals. In other words, we may tease apart the various kinds of cognitive or linguistic differences for which the proposals of conceptual engineers might aim. In what follows, we shall argue that there is a variety of divergent, yet prima facie equally worthwhile goals—or kinds thereof—each requiring the engineering of different target objects. Therefore, as we see it, conceptual engineering is a less unified enterprise than the recent literature, preoccupied with the search for a universal account, has suggested.

A first, important goal for conceptual engineers is to change how we mentally classify some of the objects we encounter, for instance, how we classify people into genders or races. These mental classifications play an important role in how we behave and interact with each other. More accurate or pertinent ways of classifying our surroundings can produce important improvements, for instance, by fostering relevant generalizations. Although what drives such mental classification is not uncontroversial, many philosophers believe that we classify objects by applying concepts to them. If this is right, then we may change how we classify objects by changing the application conditions of concepts. So, where the conceptual engineer's goal is to change how people mentally sort the objects in their environment, it seems that the application conditions of their concepts are a plausible target of their engineering endeavors.

Conceptual engineers might also wish to change the conscious or unconscious inferences that we draw about the things we encounter. Where such inferences are problematic, we might want to interfere with them, by severing or altering the psychological connections that produce them. For instance, many claim that the problem with social stereotypes is not the fact that social categories have the wrong criteria of inclusion, but, rather, that they trigger further problematic inferences—for instance, making inferences on the basis of stereotypical gender roles that a secretary is female or that a pilot is male. Whereas some of these inferences might be underwritten by our ordinary beliefs about the world, others might be sourced in beliefs that are special for being retrieved independently from context and with particular speed and automaticity. Psychologists and philosophers of psychology often think of such beliefs as the content of one's concepts (Machery, 2009, 2017). This type of conceptual content can, but need not, coincide with what we earlier described as a concept's application conditions (cf. Isaac, 2020, 2021a, b, c; Koch, 2021a; Machery, 2017, 2021). Thus, where the conceptual engineer's goal is to alter the unconscious and automatic inferences that people are likely to draw about what they encounter, psychologically construed conceptual content is a plausible target of their endeavors.

Let us now consider goals relating to language. Conceptual engineers might reasonably hope to change the truth conditions of sentences containing certain expressions. Suppose that there is some language in which sentences that ascribe married status to couples that do not consist of one man and one woman are analytically false. Then, it might be a reasonable goal for a conceptual engineer to create a situation where such sentences can, at least in principle, be true. If this is correct, the semantic value of expressions—that is, the contribution they make to the truth values of sentences—is a further plausible target for conceptual engineering (Flocke, 2020). Given that the relationship between semantic values and people's psychological makeup is unclear, we cannot simply assume that, by addressing the above targets, semantic values will be addressed ipso facto (Cappelen, 2018; Koch, 2021a). Thus, if a conceptual engineer's goal is to change those sentences that count as (analytically) true in a given language, the semantic value of expressions ought to be their target.

There are other plausible language-related conceptual engineering goals that go beyond altering truth conditions. Since the rise of semantic externalism, it has become widely accepted that an expression's literal meaning does not necessarily coincide with what a speaker who uses that expression intends to convey. The latter type of meaning is often referred to as the speaker's meaning or reference and is contrasted with a term's semantic meaning or reference (Grice, 1957; Kripke, 1977). It matters a great deal what speakers who actually use a term are trying to tell us and how they are interpreted, and, consequently, what a speaker means has an important role in communication. At the
same time, prominent examples furnished by Kripke and others demonstrate that speaker meanings are irreducible to semantic meanings and vice versa. So, if the goal is to change what people typically mean by particular utterances in communicative contexts, and how other people interpret them, the appropriate targets for conceptual engineers might turn out to be speaker meanings, rather than semantic meanings (Pinder, 2021). 13

Lastly, neither semantic meanings nor speaker meanings are necessarily normative—or, at least, whether they should be considered as such is controversial and contested. Even if they are not normative, our linguistic practices are undoubtedly governed by moral norms, social norms, and conventions (Nimtz, 2021; Thomasson, 2021). It is morally permitted to say certain things, but not others, and we have all manner of social expectations about which bits of language our fellows are likely or unlikely to use, about how they will interpret our utterances, and what we will look like in the eyes of others if we use certain expressions. We suggest that shifting the norms relating to our use of language can also be a worthwhile goal of conceptual engineering. Among other things, linguistic norms determine whether a certain speech act is permissible or not, which is likely to affect the public discourse. So, if the goal of conceptual engineering is to shift linguistic norms, the things that ground these norms—for instance, how people use language and how they expect others to use it—become legitimate targets of conceptual engineering as well (cf. Jorem & Löhr, 2022; Löhr, 2021).

This brief survey of the different goals and targets of conceptual engineering is not meant to be—and probably isn’t—exhaustive. Nonetheless, we hope that it has illustrated the vastness of the terrain of possible engineering projects. Before moving on to the next section, three general comments are in order.

First, each of the goals listed here can be pursued with a specific focus on one of many particular groups. For instance, certain conceptual engineering projects will plausibly count as successful if a group of specialized scientists changes their classification procedures or their inferential patterns, while other projects will require that such changes be adopted by society at large. Similarly, where the goal is to change linguistic norms or conventions, a conceptual engineer might hope that these changes are adopted either in society at large, or by small subgroups thereof. This account of the goals and targets of conceptual engineering is intended to be consistent with such group specificity, even if drawing out the details might require further work for which there is no scope here.

Second, the listed goals and targets are not mutually exclusive. Quite the opposite: we suspect that they complement each other and therefore often appear together (Isaac, 2021a; Koch, 2021a). In general, it seems likely that conceptual engineers will aim for changes at both the individual and the collective levels. We suspect that few conceptual engineers would wish to change linguistic norms but resist changes in people’s mental classification of the objects around them. In addition, achieving some of the goals listed here likely implies or requires the achievement of others. For instance, it is unlikely that changing the semantic meaning of a term is possible without first changing the term’s speaker meaning as it is used by people (Pinder, 2021). We do not mean to deny this. But the interconnections between the items on the above list are simply too complex to be articulated here.

Nonetheless, we believe, third, that the above taxonomy shows there to be no single unified thing that philosophers mean when they discuss conceptual engineering. 14 Conceptual engineering projects adopt radically different goals: to change how we mentally classify the objects around us, what we communicate to each other in specific contexts, or how our words contribute to the truth conditions of sentences in which they feature. If we are right, these disparate goals are best achieved by means of engineering different kinds of target objects. This shows that it is misguided to seek a unified account of conceptual engineering’s target objects. It also shows that, prior to engaging in a conceptual engineering project, practitioners must clearly specify their goals and targets.

### 4 | METHODS

There are a plethora of methodological options for how one might pursue conceptual engineering. This is especially true given the division of labor described thus far. In Section 2, we showed how most engineering projects involve description, assessment, improvement, and implementation (but not necessarily all together), each of which allows

---

13

14
for multiple methodological approaches. Furthermore, as shown in Section 3, different targets warrant distinct methods. For instance, different methods are required for assessing or identifying defective concepts than for assessing or identifying defective lexical items. The universe of methodological possibilities can split at each junction at which a choice is presented, creating a multitude of potential forms of investigation. In this section, we shall briefly discuss selected prominent examples of conceptual engineering methods, each of which is used on a different target at each respective stage.\textsuperscript{15}

The first example relates to the assessment stage. In the final chapter of Philosophy within its Proper Bounds, Machery (2017) proposes a framework for "naturalized conceptual analysis." In this framework, concepts are taken to be psychological entities whose analysis requires the use of experimental methods and delivers empirical propositions about the mind, thereby showing how people conceptualize things.\textsuperscript{16} Machery argues that analyzing a range of cases according to this method can serve to uncover conceptual deficiencies that justify improvements. He then uses this approach to assess the lay concept of innateness.

Drawing on his previous work with colleagues (Griffiths et al., 2009), Machery begins by analyzing innate traits in terms of three features: (1) fixity, (2) typicality, and (3) teleology. Briefly, (1) pertains to the unchanging or change-resistant nature of a trait in the face of environmental factors, (2) states how these traits are connected and what it is to be a certain kind of organism, and (3) consists of the relationship between innate traits and proper functioning development. Based on a survey methodology, Griffiths et al. (2009) then tested the respective weight of each feature. Lay people were presented with fictitious cases of birdsong that displayed all possible combinations of features (1)–(3), and were then asked to judge whether the birdsong was innate to its species. The results indicated that the lay concept of innateness is deficient in that it licenses empirically invalid inferences. That is, thinkers using that concept reasoned from true premises to false conclusions. "Traits that have some features associated with the concept of innateness," Machery writes, "are judged to be innate and, thereby, to have further traits associated with innateness," while traits (1)–(3) are in fact "plausibly largely independent from one another [...] By bundling together the three notions of typicality, fixity, and functionality, the concept of innateness is thus a source of unreliable inferences about traits" (Machery, 2021, pp. 10–11). If we take this to be true, there are empirical reasons to refine the lay concept of innateness so that it better aligns with the scientific concept. This example shows how experimental methods such as surveys and psychological experiments can be used at the assessment stage of a conceptual engineering project.\textsuperscript{17}

To discuss the improvement stage, we shall consider a different case study, namely the seminal work of Scharp (2007, 2013) on the concept of truth. Scharp argues that logical paradoxes such as the liar sentence ("This sentence is false") are attributable to an inconsistency in our ordinary concept of truth. He suggests replacing it with two new concepts, arguing that, in so doing, we may avoid the logical paradoxes generated by the ordinary concept. In Scharp's project of engineering the concept of truth, the assessment stage involves showing how the constitutive principles of our ordinary truth concept—roughly, these principles mandate that, from \(<p>\) infer \(<\neg \neg p>\) is true>, and from \(<\neg \neg p>\) infer \(<p>\)—give rise to the infamous paradoxes. But we shall not let this stage detain us here. Rather, we are interested in Scharp's method for improving the allegedly inconsistent concept, which he accomplishes by replacing one lexical item by two newly introduced ones.

The two concepts Scharp introduces to replace TRUTH are ASCENDING TRUTH and DESCENDING TRUTH.\textsuperscript{18} These replacement concepts have the following constitutive principles:

1. From \(<p>\) infer \(<\neg \neg p>\) is ascending true>.
2. From \(<\neg \neg p>\) is ascending true> and \(<\neg \neg p>\) is safe> infer \(<p>\).
3. From \(<\neg \neg p>\) is descending true> infer \(<p>\).
4. From \(<p>\) and \(<\neg \neg p>\) is safe> infer \(<\neg \neg p>\) is descending true>. (Scharp, 2007, p. 616).

ASCENDING TRUTH and DESCENDING TRUTH are each supposed to obey one of the constitutive principles of the inconsistent concept of TRUTH. Given that applying (1) or (3), or their converses, to most sentences of
the language does not result in contradiction, most sentences are "safe." In these sentences, "ascending truth and descending truth are identical" (Scharp, 2007, p. 616). But the application of either (1) and (3) or (2) and (4) to other sentences—most notably the liar sentence—does lead to contradiction. These sentences are unsafe.

The full axiomatization of Scharp’s concepts is impossible to reproduce here. For our purposes, it matters only that using ASCENDING TRUTH and DESCENDING TRUTH instead of TRUTH makes it impossible to derive the liar paradox and other logical paradoxes related to truth. In other words, instead of solving the paradoxes directly, for instance, by denying premises from which they follow, Scharp provides an axiomatization of truth-like replacement concepts that prevent the derivation of the liar paradox. It should be noted that the scope of Scharp’s engineering project is restricted to logicians or philosophers of logic for whom the paradoxes are most pronounced—Scharp explicitly restricts his project’s ambition to natural language semantics for sufficiently rich languages. Nevertheless, his work serves as a prominent example of how a conceptual engineering’s improvement stage can benefit from the method of logical analysis.

Lastly, we shall consider a recent methodological proposal that Nimtz (2021) has proffered at the implementation stage. Nimtz argues that, in their endeavors, conceptual engineers can exploit the normative pressure generated by social norms. Roughly, the idea departs from the recognition that word meanings are closely related to our linguistic behavior and that we can change a word’s usage by manipulating the social norms that surround its use (cf. Bicchieri, 2016; Bicchieri & Mercier, 2014). Following Bicchieri (2006), Nimtz understands social norms as rules of behavior to which people conform because they believe that (a) most people in their reference network actually conform to them (empirical expectation) and (b) most people in their reference network believe that they ought to conform to them (normative expectation). If social norms do indeed govern linguistic behavior, we can change linguistic behavior either by convincing people to shift their reference network, or by convincing members of the current reference network to change their linguistic behavior. Drawing on examples from Bicchieri and others, Nimtz argues that this general mechanism provides us with an effective, feasible, and specific means by which to change word meanings (at least given internalist premises, which bind linguistic behavior and meaning rather tightly). In other words, his proposal shows how methods and tools from behavioral research on changes in social norms can be exploited at the implementation stage of conceptual engineering projects.

Given that it was not possible to present an exhaustive survey here, this section must settle for being illustrative. The purpose-directed nature of conceptual engineering suggests that methodological pluralism is the best overall strategy. If certain socially constructed concepts can be successfully improved by the methods of linguistics, or moral concepts by formal semantic analysis, then we should apply those tools to those tasks. Similarly, if experimental techniques can describe what people take to be the functions of certain concepts, then use those tools for that purpose. From an engineering perspective, the only constraints on methodological choices seem to be feasibility and expertise. However, from a systematic perspective, we might want to couple certain methods with the particular goals and targets to which they are most naturally suited. This constraint can be best described negatively. If you think that conceptual engineering is about concepts qua cognitive devices, you had best not leave out the tools of cognitive science. Similarly, if you think that the field should be concerned with linguistic meanings, then linguistic analysis had best not be ignored. But this remains a relatively weak constraint, given that any theoretical choice still permits the possible use of additional tools notwithstanding this "natural coupling" constraint.

5 | ETHICAL ISSUES

Generally speaking, the reception of conceptual engineering among philosophers has been positive. Nevertheless, certain criticisms have been voiced, mostly in relation to the actual or possible practice of conceptual engineering. In this final section, we address and offer some tentative responses to those criticisms that grant the feasibility of conceptual engineering but raise ethical issues about its potential misuse.

How might conceptual engineering be misused? Engineers’ motives and ends are one source of concern. Cappelen (2018, p. 159) alleges that conceptual engineering projects “could be used for evil as well as for good—and
it’s extremely hard to tell them apart.” For instance, conceptual engineers could aim to precipitate, or even enforce uptake of, proposals that serve their own agenda or interests to the detriment of others, exercising what amounts to a form of “conceptual domination” (Shields, 2021). These malevolent uses stem directly from the meta-normative neutrality of conceptual engineering (see endnote 10). While they do not disqualify conceptual engineering as a method, such misuses contradict its basic “welfarist” rationale (Crisp, 2022), which is to ameliorate concepts to support the attainment of some beneficial consequences for their users. Thus, in response to this criticism, one might encourage the development of guidelines for the fair use of conceptual engineering—including, for instance, the explicit statement of an engineer’s ends and the disclosure of any conflicts of interest.

A second concern is that, despite good intentions, conceptual engineering projects can also have harmful consequences (Cappelen, 2018; Marques, 2020). For instance, they can degrade a representational device that, although sub-optimal, was still reasonably functional before the intervention. The first worry can be mitigated by a careful execution of the description and assessment phases in the engineering process, as set out in Section 2, and the decision to interrupt the project if no reasonable need for improvement is detected. The second worry could in part be controlled by adding a test phase before the implementation stage. That phase would be designed precisely to assess the proposed improvements in terms of their cost, feasibility, added value, risks, and side-effects. Moreover, we submit that conceptual engineers should be held responsible and accountable for their projects.

Another tangible, distinctively political worry concerns the actual implementation of conceptual engineering projects (Queloz & Bieber, 2021). When these projects target collective-level objects such as public meanings or shared lexica, they might be actively propagated by bodies such as technocratic governments or expert committees. Such an institutionalization might, in turn, lead to concentrations of power that abuse the less powerful through mass persuasion. To mitigate the fear of such an anti-liberal and anti-democratic situation, the risk of such abuses might be decreased by limiting the power to control collective representational repertoires to very specific contexts (e.g., education) and very restricted knowledge niches. Typically, these contexts and niches will require well-defined concepts and will involve a hierarchical structure of knowledge transmission with a strong deference to experts. The legitimacy of such expert-led projects in conceptual engineering—which are no longer akin to mass persuasion—would have to be established on a case-by-case basis. In particular, this would have to be done by an assessment, in the test phase, of how the proposed change to a given representational repertoire might affect people’s lives.

The final worry we shall consider, which also concerns the actual implementation of conceptual engineering projects, operates at the individual level and is distinctively ethical (Kitsik, forthcoming). When projects in conceptual engineering attempt to alter people’s cognitive makeup, their execution may lead to the manipulation of persons through the intimate intrusion into private territory. There are two main facets to this concern. The first relates to individual agency: conceptual engineering might undermine people’s conditions for autonomy by subverting their cognitive capacity for speech, thought, and behavior. The second relates to the engineers’ fallibility: the users of a concept targeted by an engineering project are, presumably, in a better position to know what is best for them than would-be engineers. To mitigate these risks of ethical transgressions, we submit that projects in conceptual engineering should comply with two basic requirements. First, they should engage with and empower reflective and deliberative decision-making on the part of concept users by being explicit, visible, and transparent. Second, they should be inclusive and participatory, building on inputs from their target user groups. Thereby, conceptual engineering will be not only an empirically based, interdisciplinary methodological enterprise (Sect. 4), it will also, as its name suggests, serve as a program for philosophy’s engagement with the real world (Sect. 2).

ACKNOWLEDGMENTS
Thanks to Delia Belleri, Mirela Fuš, Hans-Johann Glock’s Theoretical Philosophy Colloquium, Sigurd Jorem, Teresa Marques, Mark Pinder, Joey Pollock, and the Seminarkolloquium of the University of Zurich. Swiss National Science Foundation, Grant number: P4P4PG_199180.
ENDNOTES

1. Consult the PhilPapers entry "Conceptual Engineering," edited by Steffen Koch, for a comprehensive and constantly updated reference list: https://philpapers.org/browse/conceptual-engineering

2. AlthoughBrandon (2001, p. 587) also uses the term, in his case, it refers to the naturalized kind of conceptual analysis promoted by Fodor, Dretske, and Millikan. Floridi (2019) now prefers to speak of "conceptual design."

3. See Cappelen and Plunkett (2020: sect. 2) for a list of the central challenges for the theory of conceptual engineering.

4. The descriptive vs. prescriptive narrative is simplistic in two ways. First, conceptual engineering also includes a descriptive dimension in its preliminary phase, when one maps the conceptual territory (Glock, forthcoming; Jackson, forthcoming; see also below). Second, conceptual analysis also includes a prescriptive dimension, because when one describes a content, one regiments it at the same time (Machery, 2021, p. 217).

5. Introduction, revision, replacement, and elimination of one or several representational devices are widely considered the four main ameliorative strategies available to conceptual engineers when desiring to change a representational status quo (e.g., Cappelen & Plunkett, 2020, pp. 10–11).

6. Such omissions may be attributable to the fact that description is not seen as explanatorily central to and distinctive of conceptual engineering (Burgess & Plunkett, 2020, p. 293; see also endnote 4 above).


8. Importantly, according to dynamic accounts, a project may count as an instance of conceptual engineering, even if a process is not necessarily completed. As Chalmers (2020: 4) notes, "[each step] alone is at least a distinctive and important mode of conceptual engineering and conceptual engineering is most powerful when all [four] modes are combined."

9. We are not the first to advocate for pluralism. See especially Belleri (2021) and Fuś (2021).

10. Dutilh Novaes (2020) adds that conceptual engineering exhibits "meta-normative neutrality": the general framework of conceptual engineering is compatible with whatever purpose one might want to achieve by it. Establishing these purposes, in turn, requires a normative input that is external to the framework (cf. Haslanger, 2005).

11. Nado (2021a) argues that classification procedures need not be so closely linked to concepts and their application conditions. In her view, classification procedures can, and ought to, be the direct target of conceptual engineering. This illustrates a wider point: it is not always uncontroversial how best to theorize the relevant target objects of conceptual engineering, even where there is agreement about the respective goals.

12. Of course, the practical relevance of truth is a contested issue (Isaac, 2021c). But see Koch (2021a) for some reasons to believe that truth does matter in the context of conceptual engineering.

13. See Ludlow (2014) and Jorem (2021) for related proposals.

14. This being granted, philosophers may still disagree over which of these truly deserves the label of "conceptual engineering" (Isaac, 2021a, b).

15. For the sake of brevity, we do not discuss the descriptive step, as it is a common undertaking in many areas of inquiry within and outside conceptual engineering (see endnote 6).

16. Machery's view considers a concept to be "a subset of people's belief-like states" about a particular individual, class, event-type, or substance, which are retrieved from our long-term memory in a manner characterized by a homeostatic cluster of three properties: speed, automaticity, and context-independence (cf. Sect. 3).

17. For other proposals discussing the potential role and use of experimental methods at various stages of the engineering process in a Carnapian framework, see Koch (2019), Nado (2021c) Pinder (2017), Schupbach (2017), and Shepherd and Justus (2015).

18. Technically, these are not "ways of being true" but alethic properties with different statuses similar to modalities such as "possibility" or "necessity."

19. A reference network consists of "the range of people whom we care about when making particular decisions" (Bicchieri, 2016, p. 14).

20. See Thomasson (2021) for a related proposal.
21 Other problems discussed in the literature concern the feasibility of conceptual engineering. These are expressed either in concrete terms that depend on contingent psychological, social, and historical facts (Fischer, 2020; Koslow, 2022; Machery, 2021), or in more abstract terms related to metasemantic or metaphysical considerations (Andow, 2020; Cappelen, 2018; Deutsch, 2020a,b, 2021; Jorem, 2021; Koch, 2021a,b; Pinder, 2020a, 2021; Pollock, 2021; Riggs, 2019). These kinds of problems are omitted in what follows.

22 See Podosky (2022) for a similar worry in the conversational context.

23 For instance, in certain cases, trying to fix alleged deficiencies in scientific concepts—such as vagueness, imprecision, or unclarity—might be detrimental to scientific inquiry itself (Machery, forthcoming).

24 For instance, if Haslanger’s (2000: 46) famous slogan “to bring about a day when there are no more women” were to be misinterpreted as a call for femicides.

25 With this in mind, compare, for instance, the novel ways of thinking and acting promoted by UNESCO’s “One Planet, One Ocean” initiative (Casati, forthcoming), the restricted view of torture advanced in the Memorandum on the Standards of Conduct for Interrogation by members of the US Department of Justice in 2002 (Shields, 2021), the depathologization of homosexuality by the American Psychiatric Association in the second edition of the Diagnostic and Statistical Manual of Mental Disorders in 1973 (Cooper, forthcoming), or the definition of obesity as a disease by the World Obesity Federation (Lalumera, forthcoming).

26 For instance, on a psychological approach to concepts for conceptual engineering, one may attempt to re-delineate the content of, say, a prototype structure—that is, its constitutive features, their weights, or their dependency relationships—by changing the relevant parameters of the environmental or discourse environment with which it frequently co-occurs and thus “refurbish” people’s minds (Fischer, 2020; Isaac, 2021b; Machery, 2017).

REFERENCES


Casati, R. Rethinking the ocean: Conceptual engineering for the next 10000 years. In M. G. Isaac & K. Scharp (Eds.), New Perspectives on Conceptual Engineering (Vol. 3). Springer. (forthcoming).


Machery, É. Should we really engineer confused, unclear, or otherwise deficient concepts? The case of scientific concepts. In M. G. Isaac & K. Scharp (Eds.), New Perspectives on Conceptual Engineering (Vol. 3). Springer. (forthcoming).


AUTHOR BIOGRAPHIES

Manuel Gustavo Isaac is a Swiss NSF Research Fellow at the University of Zurich. He holds degrees in philosophy (BA, MA), logic (BA, MA), linguistics (MA), and historical epistemology (PhD) from Paris Sorbonne Universities. His work has been published in Erkenntnis, Ratio, Philosophy, Inquiry, Synthese, History and Philosophy of Logic, Logica Universalis, among other venues. Dr. Isaac has conducted as PI four postdoctoral research projects on conceptual engineering funded by prestigious fellowships and grants in Amsterdam (ILLC), Barcelona (LOGOS), St Andrews (ARCHÉ), and Zurich. He is the creator of the Conceptual Engineering YouTube Channel (https://www.youtube.com/c/ConceptualEngineering).

Steffen Koch is an assistant professor in theoretical philosophy at Bielefeld University, Germany. In 2021, he completed his PhD at Ruhr University Bochum, where he was a member of the Emmy Noether Independent Junior Research Group "Experimental Philosophy and the Method of Cases: Theoretical Foundations, Responses, and Alternatives (EXTRA)". He received a BA and an MA from Humboldt-University Berlin, and an MLitt from St.
Andrews and Stirling. His work has been published in journals like *Erkenntnis, Inquiry, Philosophical Studies, Philosophy Compass,* and *Synthese.* He is the editor of the PhilPapers entry “Conceptual Engineering.”

**Ryan M. Nefdt** is a senior lecturer in philosophy at the University of Cape Town. He completed his PhD at the University of St Andrews as a member of the Arché Research Center in Logic, Language and Epistemology. He has published on various topics including conceptual engineering, cognitive science, philosophy of science, and the philosophy of linguistics which have appeared in *Mind & Language, Philosophy Compass, Synthese, Linguistics & Philosophy, The Review of Philosophy & Psychology,* *Inquiry* and other journals. He is currently completing two monographs, one for Oxford University Press and other for Cambridge University Press.


### APPENDIX

**Table A1** Conceptual engineering canvass

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>PROCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goals</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Targets</strong></td>
<td>Delineating and regimenting a representational device as target object</td>
</tr>
<tr>
<td><strong>Mental classifications</strong></td>
<td>Concepts (semantic sense), intensions</td>
</tr>
<tr>
<td><strong>Default inferential patterns</strong></td>
<td>Concepts (psychological sense), conceptions</td>
</tr>
<tr>
<td><strong>Sentential truth conditions</strong></td>
<td>Semantic value of expressions</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>Speaker-meanings</td>
</tr>
<tr>
<td><strong>Linguistic norms</strong></td>
<td>Linguistic practices</td>
</tr>
</tbody>
</table>

**Remark.** Table A1 represents a provisional template framework for projects in conceptual engineering. It aims to offer a synopsis of the different components of conceptual engineering (that is, purposes, goals, target objects, process, and methods) we surveyed in this article and thus to provide would-be conceptual engineers with a “road map to practice” as they embark in their project. The column “parameters” covers the three parameters that specify the engineering process of a conceptual engineering project. Only the purposes, goals, and targets discussed in Section 3 have been listed in the corresponding cells. The column “process” and the two rows underneath refer to the engineering process as discussed in Section 2. The column “methods” correspond to Section 4 where only the methods listed in italics have been discussed. The remaining cells listing methods serve a purely indicative role whose justification falls beyond the limits of this article. Several of these methods can be placed in multiple cells. Finally, a few goals, targets, and even stages in the engineering process could be added to the table, which, as said, only includes those presented in the corresponding sections.