

Crafting Tools for Thought and Talk

Essays on the Theoretical Foundations of Conceptual
Engineering



Sigurd Jorem

Doctoral Dissertation in Philosophy

Faculty of Humanities

University of Oslo

October 2021

Primary supervisor

Prof. Herman W. Cappelen

The University of Hong Kong

Secondary supervisor

Prof. Amie L. Thomasson

Dartmouth College

Crafting Tools for Thought and Talk

Essays on the Theoretical Foundations of Conceptual Engineering

Dissertation submitted to the Department of philosophy, classics, history of art and ideas at the Faculty of Humanities for partial fulfillment of the requirements for the degree of Doctor of Philosophy in Humanities at the University of Oslo.

Contents

Acknowledgments	v
------------------------------	----------

Summary.....	vii
ENGLISH SUMMARY	vii
NORSK SAMMENDRAG	viii

Chapter 1

General Introduction.....	1
1. CRAFTING TOOLS FOR THOUGHT AND TALK.....	1
1.1 What is conceptual engineering?	2
1.2 Why should we engage in it?	4
2. THE RESURGENCE OF CONCEPTUAL ENGINEERING.....	9
2.1 From Kant to Carnap.....	10
2.2 Quine, ordinary language philosophy and conceptual engineering's hiatus.....	15
2.3 Thought experiment-based philosophy	20
2.4 The resurgence of conceptual engineering	23
3. THE ESSAYS	25
4. METHOD.....	30

Chapter 2

Conceptual Engineering and the Implementation Problem	33
1. INTRODUCTION	33
2. DESIGN AND IMPLEMENTATION.....	36
3. STANDING MEANING	40
4. MEANING MODULATION	49
5. SPEAKER-MEANING	52
6. SPEAK A DIFFERENT LANGUAGE	57
7. CONCLUSION.....	61

Chapter 3

The Good, the Bad and the Insignificant – Assessing Concept

Functions for Conceptual Engineering	63
1. INTRODUCTION	63
2. WHY APPEAL TO FUNCTION?	66
3. THE SELECTION PROBLEM.....	71

4. CAN WE SOLVE THE SELECTION PROBLEM BY REFINING OUR NOTION OF FUNCTION?	75
4.1 Contextually stable functions	77
4.2 Designed functions	79
4.3 Proper functions	81
4.4 System functions	85
5. CONSEQUENCES FOR THE FUNCTIONAL APPROACH.....	86
 Chapter 4	
Inferentialist Conceptual Engineering	89
1. INTRODUCTION	89
2. AN OBJECTION TO CONCEPTUAL ENGINEERING	91
2.1 Extensions	93
2.2 Intensions	94
3. WHAT MAKES CONCEPTUAL ENGINEERING WORTHWHILE?	97
3.1 Semantic inferentialism.....	100
3.2 Inferentialism about conceptual engineering.....	104
4. OBJECTION	107
5. CONCLUSION.....	110
 Chapter 5	
Against Amelioration as Revelation	112
1. INTRODUCTION	112
2. WHAT A TERM OUGHT TO MEAN	114
3. IS AMELIORATION ALWAYS REVELATION?	117
3.2 An unacceptable constraint on semantics.....	124
4. EVALUATING SPEECH.....	126
6.CONCLUSION	131
 References	133

Acknowledgements

This dissertation is the culmination of three years of writing, and many more years of thinking about conceptual engineering, revisionary analysis and conceptual choice. But most of all, it is the product of participating in a philosophical community and enjoying discussions with intelligent and caring friends. I am deeply grateful for all the thoughtful discussions and social occasions I have had with colleagues and friends over the course of working on this project. I single out a number of individuals below, but I will not be able to name them all. I would therefore like to express my gratitude to all those who have engaged in discussion about these topics with me. The most valuable experiences I have had working on the project have been together with other people, and I do not dare to imagine what the dissertation would look like without the exchanges I have had with you all. Thank you.

My main supervisor, Herman Cappelen, has helped shape my thinking about most of the issues covered in this dissertation. We have many points of disagreement about them, but the disagreements have aided me. He has provided detailed and fruitful feedback on papers, always in a friendly and encouraging fashion, facilitated contact with the wider philosophical community, and alleviated my worries about what seemed like a daunting task. Thank you, Herman. My other supervisor, Amie Thomasson, has provided sharp and valuable comments on my work, and is a significant intellectual inspiration for the kinds of views I am trying to develop. Thank you, Amie.

I am grateful to my fellow conceptual engineering enthusiasts, in particular Steffen Koch, Mark Pinder, Joey Pollock, Manuel Gustavo Isaac, Delia Belleri and Mirela Fuš. Thank you for many fruitful discussions and sharp comments. Additional thanks to Joey for helping me out with proof-reading in the final stretch. Special thanks to Guido Löhr for extensive and valuable discussion, not least in the process of writing “Inferentialist Conceptual Engineering” together. Many thanks

to David Plunkett for giving me detailed comments and valuable advice in his capacity as midway evaluator.

I am lucky to have had so many kind, funny, smart and caring people with whom I have shared the fate of being a PhD student in Oslo. Yael, Feroz, Hallvard, Mirela, Aksel, Tomas, Jonnie, Maria, Yelena, Hilde, Hans-Robin, Joona, Kim, Joost, Jeroen, Sara, Lars, Magnus and more: Thank you all! Special thanks to Christopher Masterman for being the office-mate dreams are made of, and for detailed comments on multiple parts of this work, to Max Johannes Kippersund for extensive discussion, and to Francesca Secco for many helpful comments. To all my other friends – thank you so much for your support and interest in the project. Special thanks to Zacharias Andreadakis for unwavering support.

Many people at my department have aided this project in various ways. I thank Pia Søndergaard, Øystein Linnebo, Solveig Aasen, Rachel Sterken, Christine Amadou, Beate Elvebakk, Karen Haugland, Caroline Strutz Skei, Bente Larsen, the members of ConceptLab and all the people that make the department tick.

To my near and dear, in Norwegian. Mamma, pappa og Magnus: Tusen takk for all støtten dere har gitt meg, og hjelpen jeg har fått av dere. Astrid: Tusen takk for at du er den du er. Det hadde ikke vært lett å gjennomføre dette om ikke du var en så fantastisk skjønn og grei jente. Sist, men ikke minst, tusen takk til deg Silje, for å være en vidunderlig kjæreste. Du har vært forståelsesfull, omsorgsfull og interessert, og dine skarpe kommentarer har hjulpet meg betydelig.

Finally, I would like to acknowledge the fact that being paid a salary to think and write about the issues that interest me the most is a deeply privileged position to be in. I am sincerely grateful to the Department of philosophy, classics, history of art and ideas and the University of Oslo—and ultimately to Norwegian taxpayers—for the opportunity to carry out this project.

Summary

English summary

There is a long-standing tradition in philosophy of *analyzing* our concepts, e.g. of knowledge, free will and causation. But rather than trying to capture the content of our concepts—or indeed, the nature of the object-level phenomena we use our concepts to think and talk about—we may engage in the normative project of asking which concepts we ought to be using. Perhaps there is some concept of causation that is more useful for the purpose of explaining events and states of affairs than the concept of causation that was handed over to us by our ancestors? We may thus take a constructive and revisionary attitude to our conceptual repertoire. To evaluate, construct and revise concepts is to engage in so-called conceptual engineering.

This dissertation presents four, independently readable essays on the theoretical foundations of conceptual engineering, as well as a general introduction to the topic, why we should engage in it and its historical roots in philosophy. In the first essay, I argue that the existence of meaning determinants beyond our control does not inhibit us from making worthwhile changes to what words mean, or what we use words to communicate. In the second, I show that so-called functional approaches to conceptual engineering face a problem pertaining to harmful and methodologically insignificant concept functions. I argue that the problem forces advocates of function to make a weaker claim than what they might want, but that this does not spell defeat for the approach. In the third essay, my co-author and I argue that we should endorse an inferentialist view of what is being engineered. According to this view, the things being engineered are identified by their role in practical and theoretical reasoning. In the fourth and final essay, I argue that we should reject metasemantic theories that see (would-be) cases of semantic improvement as cases of revealing what we already meant. With a

pragmatist outlook, the essays support an optimistic and favorable attitude to conceptual engineering.

Norsk sammendrag

I filosofien er det en lang tradisjon for å *analysere* begrepene våre, for eksempel begrepene vi har om kunnskap, fri vilje og kausalitet. I stedet for å prøve å fange innholdet til begrepene vi finner oss med—eller, for den saks skyld, egenarten til de fenomenene vi bruker begrepene våre til å tenke og snakke om—kan vi engasjere oss i det normative anliggende å finne ut hva slags begreper vi burde ha i bruk. Kanskje finnes det et begrep om kausalitet som er mer nyttig i forklaringssammenheng enn det begrepet om kausalitet vi har arvet fra våre forfedre? Vi kan på denne måten ta en konstruktiv og revisjonistisk tilnærming til begrepsapparatet vårt. Å evaluere, konstruere og revidere begreper er å engasjere seg i såkalt *conceptual engineering*.

Denne avhandlingen består av fire ulike artikler om det teoretiske grunnlaget for *conceptual engineering*, i tillegg til en introduksjon til temaet, med begrunnelse for hvorfor vi skal engasjere oss i praksisen og en beskrivelse av temaets filosofihistoriske røttene. I den første artikkelen argumenterer jeg for at det er mulig å gjøre verdifulle endringer til et språkuttrykks mening, selv om vi skulle anta at mening bestemmes av faktorer utenfor vår kontroll. I den andre artikkelen viser jeg at *funksjonsorienterte* tilnærminger til begrepsrevisjon støter på et problem som har å gjøre med at det finnes skadelige og insignifikante begrepsfunksjoner. Jeg argumenter for at problemet tvinger den funksjonsorienterte teoretikeren til en svakere påstand enn hva hun kanskje kunne ønske, men ikke forkaste hele synet sitt. I den tredje artikkelen argumenterer jeg og min medforfatter for at vi bør omfavne et inferensialistisk syn på hva som blir konstruert og revidert når vi driver med begrepskonstruksjon og begrepsrevisjon. Dette synet identifiserer tingene som blir konstruert og revidert med hvilken rolle de har i praktiske og teoretiske slutninger. I den siste artikkelen argumenterer jeg for at vi bør forkaste

metasemantiske teorier som ser på (det som ville vært) tilfeller av semantisk forbedring som tilfeller av å avsløre hva vi allerede mente med språkuttrykket. Med et pragmatistisk grunnsyn støtter artiklene et optimistisk og positivt syn på begrepskonstruksjon og begrepsrevisjon.

Chapter 1

General Introduction

1. Crafting tools for thought and talk

For whereas he speaks of *the* conceptual frame, *the* system of formal and material rules of inferences, we recognize that there are an indefinite number of possible conceptual structures (languages) or systems of formal and material rules, each one of which can be regarded as a candidate for adoption by the animal which recognizes rules, and no one of which has an intuitable hallmark of royalty. They must compete in the market place of practice for employment by language users, and be content to be adopted haltingly and schematically. In short, we have come out with C. I. Lewis at a ‘pragmatic conception of the *a priori*’. (Sellars 1953, 337)

Wilfrid Sellars’ imaginary opponent represents a tradition in philosophy that is not only real, but one that has been dominant for some time. The tradition may be defined as one of treating our language as given. The philosopher is tasked with identifying what it takes for, say, a mental state to count as knowledge, an action to count as intentional, or a relationship between two events to count as causation. She is to develop descriptively accurate theories about concepts, bits of language or the phenomena that our terms as-a-matter-of-fact refer to. She is not tasked with finding out what kind of concept of knowledge we *ought* to be using, which meaning candidate for ‘intentional action’ would best serve our practical interests or which phenomenon in the vicinity of what we call ‘causation’ most deserves our attention.¹ The latter endeavors are characterized by a normative and revisionary

¹ I alternate between talk of concepts, meaning and object-level phenomena to show that the distinction between descriptive and revisionary projects is tangential to distinctions between what the objects of inquiry are (concepts, language or things in the world), c.f. (Eklund 2017, 192–94)

approach to the tools we use to speak and think, to be contrasted with the descriptive and, frankly, uncritical approach sanctioned by the still-dominant tradition.

On the view of language Sellars advocates, we should not limit the objects or tools of our inquiry to our current language. Contrary to the tradition of treating our language as given, this suggests that philosophical practice is badly impoverished if it fails to encompass normative and revisionary theory about our tools for thought and talk—either in addition or as an alternative to exercises in traditional conceptual analysis (and other descriptive projects). Using a recently popularized phrase, Sellars’ view of language demands that we engage in *conceptual engineering*.

In this introduction, I first describe what conceptual engineering is and give reasons for being interested in it. I then give a brief, broad-strokes history of conceptual engineering in philosophy. Finally, I outline the content of the essays that make up the subsequent chapters and briefly describe the method I have employed.

1.1 What is conceptual engineering?

Conceptual engineering is the practice of evaluating and designing concepts, and implementing new concepts or revisions of existing concepts.² In academic philosophy, recent examples include Sally Haslanger’s effort to develop concepts of gender that can serve as “effective tools in the fight against injustice” (Haslanger 2000, 36); James Woodward’s (2003) effort to design concepts of causation that

² In giving this definition, I do not use ‘concept’ in some narrow, technical sense restricted to, say, mental particulars. Nor do I presume a particular account of what concepts are. Concepts, in the present, loose sense, are the tools we use to speak and think. They are the sort of things we use sub-sentential linguistic items to express. For example, if we say ‘She knows that Berlin is the capital of Germany.’ we use the word ‘know’ to express a concept, i.e. a concept of knowledge.

may improve our explanations of events and states of affairs; and Kevin Scharp's (2013) argument for replacing our (allegedly) inconsistent concept of truth so as to avoid the alethic paradoxes. None of these projects aim at a faithful analysis of existing concepts. Rather, they aim to improve our conceptual toolbox.

Conceptual engineering in philosophy is of special interest, because it constitutes a method that is in tension with prevailing metaphilosophical beliefs and methodological norms. Nonetheless, it is important to recognize that conceptual engineering does not exclusively occur in academic philosophy. If a sociologist wants to investigate how much social mobility there is in a particular society, she needs an interpretation of the data available to her in order to test competing hypotheses. This requires that she operationalizes concepts of, say, social class. This part of her work would fall squarely in the category of conceptual engineering: It is to design a concept or set of concepts that allow her to do something, i.e. reach new, potentially significant truths. Scientists frequently have to operationalize concepts for the purposes of data interpretation and theoretical inference, so there is plenty of conceptual engineering in science. In politics and law, practitioners are frequently obliged to determine not just which cases fall under a particular term or concept of interest (e.g. 'person' in the case of a fetus), but with determining whether the target concept ought to be such that it applies to the case in question (in view of its role in the law, ought our concept of a person be such that it applies to a fetus?). Conceptual engineering also occurs in ordinary discourse. Peter Ludlow describes many examples of what he calls "meaning litigation", such as his personal memory of litigating the meaning of 'doll' with his father (Ludlow 2014, 9–10). As a child, Ludlow and his father had a conflict over whether a toy figure he had received counted as a doll. The figure had the resemblance of a native American and was named 'Chief Cherokee'. The disagreement was not one of identifying the actual extension of the term 'doll', but of determining what the extension of 'doll' ought to be. Understandably, the young

Ludlow did not want his action figure to be colored with the feminine valence of ‘doll’. The meaning of ‘doll’ has proven to be more significant than this, however:

For the most part how we define ‘doll’ doesn’t have important consequences. But there are exceptions. For example, until recently doll imports were taxed a higher rate than other toys. It thus became necessary to sharpen up the definition of ‘doll’ (should Chief Cherokee be taxed at the higher rate?). Accordingly, the Harmonized Tariff schedule defined dolls as being distinguished from toys by “representing only human beings and parts and accessories thereof.” (Ludlow 2014, 9)

The examples I have given are all meaningfully grouped together as exercises in conceptual engineering. What they have in common is a normative, revisionary and constructive approach to the tools we use to speak and think.

The present work aims to contribute to a better understanding of conceptual engineering. In the component essays, I address questions about whether it is feasible to change meaning, what the normative constraints on conceptual engineering are, what kind of thing it makes practical and theoretical sense to engineer, and whether we, in talking about conceptual engineering, are misclassifying changes in belief as changes in meaning.

1.2 Why should we engage in it?

I deal with general and fairly abstract issues about conceptual engineering in the essays. To motivate interest in the theoretical issues, we should want to know why we ought to engage in the practice. Naturally, our motivation for engaging in conceptual engineering may be specific to the individual project one is engaged in. For example, Scharp’s (2013) proposal to replace our concept of truth is motivated by the belief that our concept of truth encodes inference rules that inevitably lead

us to contradiction.³ We should replace our concept of truth if using it commits us to contradictions. Beyond such specific motivations, however, there is also something general to say about why we ought to engage in conceptual engineering.

We may vindicate engagement in conceptual engineering by considering the fact that our languages are historically contingent systems, shaped by forces that do not unmistakably select for the features our languages ought to have. At each point in time, a natural language has a particular set of features that other time slices of the language, or other languages, do not necessarily share.⁴ Current American English has a particular set of syntactical rules, a set of phonemes, rules of pronunciation etc. These are, by and large, contingent features of current American English—evidently so, because there are a vast number of languages with entirely different features. The *semantics* of American English is also a contingent matter: The words, phrases and sentences of American English could have meant something different from what they in fact do. For the purposes of thinking about conceptual engineering, we are particularly interested in what we may call the *lexicon* of a language. A lexicon, in the present sense, is a set of sub-sentential expressions paired with a set of meanings, not an actual book cataloguing those expression-meaning pairs. So understood, our lexicon is aptly described as a conceptual repertoire: It is a set of concepts (meanings) paired with a set of linguistic items. As most other language components, the exact content of a lexicon is a contingent matter.

When we realize that our lexicon could have been different, we need to consider the ways in which it could be different—what possible lexicons are there?

³ Note that Scharp speaks of “constitutive principles”, but “inference rules” captures his idea well enough for present purposes (see Scharp 2013, 36–43). Elsewhere, Scharp (2020) has claimed that most concepts studied in philosophy are inconsistent at some level of analysis, and argued that philosophy is the study of inconsistent concepts.

⁴ This is not to preclude that there may exist universal features of language, e.g. necessary syntactical features of the sort generativists study. If there is a universal grammar, it only accounts for a limited number of the features of a language.

(c.f. Dever 2020). We also need to consider why we find ourselves with this particular lexicon as opposed to any of the other possible lexicons we could have been using. Arguably, our lexicon is the product of both biological evolution and cultural development. Using Sellars' phrase, expression-meaning pairs "must compete in the market place of practice for employment by language users". Thus, our lexicon is determined by processes that are to some extent selective. The content of our lexicon is not perfectly random. Only by serving some function does an expression-meaning pair proliferate and enjoy sustained use over time.⁵ Through this lens, we have *prima facie* reason to trust the tools we use to speak and think. Selection mechanisms ensure that our lexicon is not hopelessly inadequate for coping in the world and dealing with each other. However, selection mechanisms do not necessarily solve for the best possible lexicon. Concepts of race, e.g., may proliferate because they serve the interests of privileged groups, but that does not make them morally benign.

There is a growing literature on conceptual genealogy, where theorists try to understand why we have the concepts we do and which purposes our concepts serve (see Queloz 2021). This kind of work may aid projects in conceptual engineering, by identifying purposes or functions that we should pay heed to when we purport to revise the meaning of a term. Thus, the factors that help explain why we have the lexicon we do may play a normatively constraining role for how to improve it. It is also clear, however, that whatever the identity of the forces that cause us to have the particular lexicon we do, they do not in addition guarantee that our lexicon is in some optimal, unimprovable state. Any lexicon that falls short of the (hypothetical) optimal could be improved. This may happen by expanding the lexicon with a new expression-meaning pair, or by revising the meaning of an existing expression. In either case, there is conceptual engineering to do.

⁵ The *proper function of a language device* is the core notion with which Ruth Millikan builds her semantic theory (Millikan 1984). Needless to say, recognizing the historical evolution of language is not to endorse all the details of Millikan's theory.

Since the content of our lexicon is a contingent matter, shaped by forces that do not unmistakably select for the meanings our terms ought to have, we need to attend to our lexicon with a critical and constructive attitude. This just is to engage in conceptual engineering. Clearly, however, we would have little reason for doing so if concepts could not aptly be evaluated as good or bad. Thus, another part of our motivation for engaging in conceptual engineering consists in an appreciation for idea that there are such things as conceptual virtues and conceptual defects. We have already mentioned a few reasonable candidates: A concept that encodes inconsistent inferences rules is arguably defective: Using it will commit us to contradictions, and we do not want to be committed to contradictions. Exactness, meanwhile, is a reasonable candidate for a conceptual virtue—at least in some range of speech contexts. Scientists need to determine whether a concept applies on the basis of data they are able to attain. This may require that the concepts they use are made exact, in which case exactness is a conceptual virtue in the context of scientific inquiry. Carnap, who was an influential proponent of more exact speech and thought, also heralded a conceptual virtue he called “fruitfulness”: A concept is fruitful if it allows us to articulate true generalizations or laws (Carnap 1962, chap. 1).

Carnapian, science-oriented conceptual virtues may be juxtaposed with moral evaluations of concepts. For example, a concept of marriage that only applies to couples consisting of one man and one woman is arguably harmful, because organizing our practice around it would constitute injustice towards non-heterosexuals. We want our conceptual repertoire to be morally benign, so this concept of marriage would arguably stand in need of revision. We may also want our conceptual repertoire to have critical potential, with concepts to capture social realities that are being ideologically disguised as natural—I take this to be the core objective guiding Haslanger’s project of engineering new concepts of gender and race (Haslanger 2012; 2020).

These remarks are merely intended to give reason to think that concepts—and the conceptual repertoires of which they are members—are aptly evaluated as good or bad along various dimensions. I do not intend to provide a tidy, exhaustive taxonomy of conceptual virtues and conceptual defects. One reason for this is that I do not have a firm grasp of the full range of plausible candidates for conceptual virtues and conceptual defects. The literature on conceptual engineering and conceptual ethics is underdeveloped on this point, and I have not made a systematic effort to push it forward. More interestingly, there are reasons to be skeptical about the idea that there is a tidy and exhaustive taxonomy forthcoming. Different purposes, contexts, functions, tasks etc. may place different sorts of requirements on which concepts we ought to be using. For instance, exactness is plausibly a conceptual virtue in science. In politics and diplomacy, by contrast, there may be good reasons to favor vague and imprecise manners of speech, e.g. in order to reach preliminary agreements. If contingent factors such as contexts, purposes or functions help determine what makes a concept good, the prospects for a neat taxonomy become dim. There are an indefinite number of different sorts of contexts and purposes in which and for which we use concepts, so we may not know the range of conceptual virtues and defects in advance of finding ourselves in a particular context with a particular purpose or task. I elaborate on this issue in the third chapter, where I discuss the so-called functional approach to conceptual engineering.

Thus, we do not have a neat taxonomy of conceptual virtues and defects, but we do have reason to engage in conceptual engineering. Our languages and the conceptual repertoires they encode are not spotless, unimprovable systems, marked by—again quoting Sellars—an “intuitable hallmark of royalty”. And we cannot naïvely trust the processes behind language evolution. Rather, we must take responsibility for assessing and crafting our tools for thought and talk.

In the following section, I sketch a history of interest and engagement in conceptual engineering in philosophy, and offer a rough explanation for why it is making a resurgence as a methodological framework and topic of discussion.

2. The resurgence of conceptual engineering

In philosophy, there has lately been a surge of interest and engagement in conceptual engineering. Of 150 entries in the Conceptual Engineering category on philpapers.org, roughly half are works published in 2020 and 2021, and only one entry is from before 2000.⁶ This would suggest that conceptual engineering is a novelty in the history of philosophy—something that has been brought up in arguments over the value and viability of the practice. Max Deutsch writes that “If the advice to conceptually engineer amounts to no more than, ‘good job, philosophers, keep doing what you’re doing!’, then fine. But the rhetoric surrounding recent discussions of the practice suggests that this is not the view.” (Deutsch 2020, 3949). In my understanding, Deutsch poses a false dichotomy, and the purpose of the present section is to show why and how it is false. I show that conceptual engineering is not simply a novelty. Nor is it simply an old practice with a new, fancy label. I thus argue for a picture in agreement with what Matti Eklund has suggested:

Some friends of the idea of philosophy as conceptual engineering can tend to present the idea as something importantly new. Care is needed here. First, it can reasonably be argued that conceptual improvement has in fact been a traditional concern of philosophy, and philosophers have de facto been concerned with technical concepts differing from their ordinary counterparts, even while they have not been explicit about this methodology and have not explicitly distinguished between what they have been concerned with and its ordinary counterpart. Second, some philosophers have been quite explicit

⁶ <https://philpapers.org/browse/conceptual-engineering> (visited May 19, 2021).

about adopting a methodology like that currently being promoted under the name of conceptual engineering. (Eklund 2017, 194)

Relative to the recent history of English-speaking philosophy—roughly, from the 1960’s onwards—interest and engagement in conceptual engineering may be something of a novelty, but that does not make it a novelty relative to all time-frames or reference classes. We may detect commitment to a conceptual engineering-framework in pieces of philosophy from long before the turn of the millennium, sometimes made explicit in the form of claims. We can see this when, e.g., a philosopher treats concepts as subject to normative assessment, purports to revise or replace a concept, or makes claims in favor of a conceptually normative or revisionary methodology. I show that several prominent philosophers in our past have thought of concepts as the sort of things we may assess normatively and revise if need be. If this holds true, a normative and revisionary approach to concepts is not a 21st century innovation. On the other hand, the current rise of normative and revisionary thought about concepts is a real shift of interest and methodological commitment from what has dominated philosophy the last few decades, where conceptual/philosophical analysis and strands of naturalism have prevailed. In the following subsections, I describe a significant part of the historical background for conceptual engineering and explain its resurgence in a rough, preliminary fashion.

2.1 From Kant to Carnap

As early as Kant’s *Critique of Pure Reason* ([1781–1787] 1998) we find the idea that concepts are the sort of things that stand in need of justification. In the transcendental deduction, Kant does not argue that we possess the twelve “categories” out of metaphysical or physical necessity. Rather, the argument is that we need to make use of these particular concepts in order to form objective judgments about the world we encounter in experience. Kant makes the tacit but fair assumption that we should want to form objective judgments based on experience. If he can prove that some concepts are preconditions for our exercise

of objective judgment, then those concepts have as sound a foundation as any empirical (“non-pure”) concept could have. In this way, Kant’s deduction of the categories is a pragmatic and normative argument: We ought to accept the categories because they allow us to do something, viz. engage in objective, empirical judgment. Hegel follows Kant in thinking about concepts as being amenable to rational assessment and subject to justification. In Hegel’s thought, as opposed to Kant’s, concepts are essentially the products of social interactions in particular historical contexts; even the concepts most fundamental to our capacity for empirical judgment. A common theme is that the concepts we possess are our responsibility. They are not Platonic forms which we can merely grasp, as opposed to tools we can assess and, at least in Hegel’s case, amend.

Nietzsche goes further than Kant and Hegel. He advocates what is, for the time, a radical view in favor of conceptual engineering:

What dawns on philosophers last of all: They must no longer accept concepts as a gift, nor merely purify and polish them, but first make and create them, present them and make them convincing. Hitherto one has generally trusted one’s concepts as if they were a wonderful dowry from some sort of wonderland: but they are, after all, the inheritance from our most remote, most foolish as well as most intelligent ancestors. [...] What is needed above all is an absolute skepticism toward all inherited concepts... (Nietzsche [1901] 1968, section 409)

It is also fair to interpret the classical American pragmatists as engaged in revisionary projects. We may view Peirce, James and Dewey as seeking to revise concepts e.g. of truth and experience, that in their view had become detached from our practical needs. C. I. Lewis explicitly endorsed a program of conceptual engineering. As intimated in the quote from Sellars with which I opened this chapter, Lewis argued for a pragmatic conception of the *a priori*. According to the view set out in *Mind and the World-Order* (C. I. Lewis 1929), our endorsement of statements that function as *a priori* in our engagement with the world—and

consequently our selection of concepts as they are made explicit by such statements—is to be determined on pragmatic grounds:

There may be alternative conceptual systems, giving rise to alternative descriptions of experience, which are equally objective and equally valid, if there be not some purely logical defect in these categorical conceptions. When this is so, choice will be determined, consciously or unconsciously, on pragmatic grounds. (C. I. Lewis 1929, 271)

In the continental tradition,⁷ works are frequently characterized by radical, revisionary ambitions: “The project of analytic philosophy contrasts with that of ‘continental’ thinkers, who are concerned more with creating new languages for describing the human condition than offering (analytic) explications of the entrenched languages.” (Gutting 1998, 10). The Heidegger of *Being and Time* ([1927] 1962) is not concerned with translating his peculiar idiom into established language precisely because he rejects the conceptual framework associated with the latter. He aims to rebuild a framework with which we can understand, most importantly, what it is for someone or something to *be*. Insofar as *Being and Time* is paradigmatic in this regard, it is unsurprising that works in the continental tradition often make for difficult reading.

Conceptual engineering reaches a highpoint in early analytic philosophy. While conceptually revisionary thought would later take a hiatus from analytic philosophy, it was clearly present in its inception. The early analytic philosophers were concerned with developing the tools of modern logic and semantic theory, and they could not access those tools via faithful analysis of existing concepts or pieces of language. Frege makes a point of devising a new notation in the *Begriffsschrift* because German did not stand up to the task of developing what would become modern logic:

⁷ Here I have in mind a varied group of philosophers including e.g. Heidegger and Deleuze.

I believe I can make the relation of my “conceptual notation” to ordinary language clearest if I compare it to the relation of the microscope to the eye. The latter, because of the range of its applicability and because of the ease with which it can adapt itself to the most varied circumstances, has great superiority over the microscope. [...] But as soon as scientific purposes place strong requirements upon sharpness of resolution, the eye proves to be inadequate. (Frege [1879] 1972, 104–5)

And shortly thereafter:

These deviations from the traditional find their justification in the fact that logic up to now has always confined itself too closely to language and grammar. In particular, I believe that the replacements of the concepts of subject and predicate by argument and function will prove itself in the long run. (Frege [1879] 1972, 107)

Frege created the *Begriffsschrift* notation and the concepts expressed by it to serve scientific purposes. Specifically, they were to serve as tools for disciplined inference and rigorous proof. The project was not to analyze the concepts expressed by ordinary language words such as ‘and’, ‘or’, ‘if’, ‘not’, ‘some’, ‘all’, etc. but to construct concepts that could play something like the role of the former, except better—at least in the context of scientific inquiry.

Among early analytic philosophers, no one practiced conceptual engineering more clearly nor with more methodological self-consciousness than Frege’s student Carnap. In fact, the term ‘conceptual engineering’ was coined by Richard Creath to describe Carnap’s view of philosophy (Creath 1990, 7). Carnap engineered concepts of intension and extension as replacements to our pre-theoretic concepts of meaning and reference, giving them exact contents for sentences as well as sub-sentential expressions (Carnap 1956). He engineered concepts of analyticity, probability and inductive confirmation (Carnap [1934] 1937; 1962). Even the much-maligned *Aufbau* ([1928] 1967) may be viewed as an exercise in conceptual engineering. It puts forward a conceptual framework that promises to substantiate

objective knowledge based on perception: Not simply as an analysis of our actual language, nor as a straightforward revisionary proposal, but as a conceptual framework we *could* use, and would have a reason for using if we wanted to track the evidential relations from perception to language closely (c.f. Carnap 1963a, 16–20; Friedman 1987).

In addition to engaging in conceptual engineering, Carnap also theorized about it. This part of his work consists in his outline and defense of “the method of explication” (Carnap 1956; 1962; 1963b). To apply Carnap’s method of explication is to purport to replace—for purposes of better inquiry—a concept of ordinary discourse (the *explicandum*) with a more exact concept (the *explicatum*). The desiderata Carnap placed on a good explication are in line with the scientific purposes the *explicatum* is to serve. According to Carnap, the *explicatum* should be 1) exact, in the sense of not being vague or otherwise indeterminate, 2) preferably higher in the order of quantitative > comparative > categorical (what Brun 2016 calls “precision”), 3) fruitful, in the sense of permitting generalizations, 4) similar to the *explicandum*, since there is no genuine replacement relation if not, and 5) simple.⁸ With these criteria, Carnap defined a particular way to conceptually engineer. Carnap’s outline and defense of explication remains a leading work on how to engineer concepts for purposes of better inquiry.

The early history of conceptual engineering in philosophy is, with the exception of Carnap, not a tale of dedicated practitioners and theorists who self-consciously apply or theorize about a conceptually revisionary method. Nevertheless, the prior examples show that normative and revisionary thought about concepts was present in philosophy before 1960, and we did not need to look to peripheral thinkers to make this case. Still, we should assume that the strands of a conceptual engineering methodology we have identified were intermixed with

⁸ There are only four desiderata in Carnap’s original exposition, as he does not formally distinguish precision, in the sense specified, from exactness (Carnap 1962). Here I follow Brun (2016) in treating precision as an independent desideratum.

thoughts and tools from what we now can identify as non-revisionary analysis and naturalist metaphilosophy. Only later, with the crystallization of non-revisionary analysis—the sort of analysis that we evaluate with thought experiments—does conceptual engineering stand out as a real methodological alternative. That is, it is the philosophy that would follow after Carnap’s heyday that provides a foil against which we can identify conceptual engineering as a method, and against which we are in a position to recognize the philosophical value of conceptual engineering. Before we get that far, let us consider why philosophers generally ceased to think of concepts as tools we can evaluate and amend. Why did Carnap’s view of philosophy not prevail?

2.2 Quine, ordinary language philosophy and conceptual engineering’s hiatus

Quine was one of Carnap’s closest friends over several decades, but also his greatest critic. Quine criticizes Carnap’s views in multiple articles; most famously in “Two Dogmas of Empiricism” (1951), which became one of the most influential philosophy publications of the 20th century. In Two Dogmas, Quine argues that the analytic/synthetic distinction is unintelligible. If Quine’s arguments are successful, the consequences are grave for Carnap’s framework, where the distinction between analytic and synthetic statements play a pivotal role.

The truth of a synthetic statement is supposed to be determined by the way the world is. To justify a synthetic statement one has to show that the world is as the statement has it, paradigmatically by appeal to observation. For example, the truth value of

- 1) The sun is shining.

may be settled by appeal to what we can see through the window. We cannot justify an analytic statement the same way, since the truth of an analytic statement is supposed to be determined by the meaning of the statement, irrespective of the way the world is. To illustrate, the truth of

2) Necessarily, A knows that p only if A believes that p.

is not properly supported by looking at knowers of p to check whether all of them are believers of p, or anything quite like this. What we can do is appeal to how we use the words “know” and “believe” etc. Our linguistic conventions make sure that, strictly speaking, one cannot know what one does not believe. We may therefore appeal to our linguistic conventions to support the truth of 2). However, in line with his enthusiasm for conceptual engineering, Carnap is not concerned with what our actual conventions are. Rather, he is concerned with what they *could* and what they *should* be. Accordingly, he holds that analytic statements are to be evaluated according to the consequences of adopting a convention on which the statement is true. Therefore, on Carnap’s view, the truth of an analytic statement is not something up for discussion except as an indirect way of discussing the pragmatic virtues of adopting the convention(s) that support its truth. Finding out which analytic statements we ought to endorse is to find out which concepts we ought to use. In this way, analytic statements are the locus of conceptually normative evaluation and argument in Carnap’s system.⁹

It is well-known that Quine’s rejection of the analytic/synthetic distinction affects a crucial piece of Carnap’s framework, but less has been said about its specific consequences for the topic of our interest, i.e. Carnap’s view that conceptual engineering is a crucial part of inquiry, to be guided by pragmatic considerations like exactness and fruitfulness. If the present account is accurate, that view of Carnap’s is collateral damage to Quine’s objections. This point deserves nuance. Quine’s influential view is not that pragmatic considerations play no role in inquiry. Indeed, Quine was a self-avowed pragmatist and declares that his argument in Two Dogmas has the effect of “a shift toward pragmatism” (Quine

⁹ Meanwhile, pragmatic considerations do not intrude into our evaluation of synthetic statements, except insofar as we are concerned with which content the statement is (should be) an expression of. In this way, Carnap can isolate the contribution of pragmatic virtues to the goodness of discourse.

1951, 20). Nor does he deny that there are conventions of language use, and that such conventions help determine which statements are true and which are false. Rather, Quine's view is that the conventional element of language use is more elusive than Carnap would have it: We cannot single out a set of statements that we privilege as analytic. As Quine writes in another paper where he develops his objection to Carnap:

I do not see how a line is to be drawn between hypotheses which confer truth by convention and hypotheses which do not, short of reckoning *all* hypotheses to the former category [...] Evidently our troubles are waxing. We had been trying to make sense of the role of convention in a priori knowledge. Now the very distinction between the a priori and empirical begins to waver and dissolve, at least as a distinction between sentences. (It could of course, still hold as a distinction between factors in one's adoption of a sentence, but both factors might be operative everywhere.) (Quine 1963, 397)

For our purposes, the crucial consequence of Quine's objection to Carnap is the following. If we cannot distinguish analytic statements from synthetic statements, we cannot single out the locus of conceptually normative evaluation and argument so important to Carnap's program of conceptual engineering. In Quine's competing, holistic system, it is harder to see exactly what kind of role pragmatic considerations have, what legitimate pragmatic considerations there are, and when they hold priority over traditional considerations pertaining to truth-evaluation. To illustrate, endorsing a statement may ensure greater simplicity in our system of beliefs, but when are we justified in doing so if the statement has no support in observation? What if our observations run counter to a statement—can pragmatic considerations in favor of the statement overturn those observations?

This is not the place to evaluate Quine's arguments against the analytic/synthetic distinction or examine how Carnap could respond to them. What is worth noting is that Quine's influence in philosophy made it less popular and hence less viable to assume an analytic/synthetic distinction in the

philosophical environment that developed. Searle writes that “Most philosophers today accept some version or other of Quine’s rejection of the analytic-synthetic distinction.” (1996, 7; quoted in O’Grady 1999) The fact that Quine convinced the philosophical community that they ought not assume an analytic/synthetic distinction had the methodological consequence of detracting from the appeal of engaging in conceptual engineering. It does so because on Quinean premises it is less clear when we are justified in invoking pragmatic considerations about conceptual choice, and less clear how those considerations rank against traditional, truth-pertaining considerations such as a statement’s having support from observation. In addition, Quine’s influential argument is all but addressed to Carnap, so it would be unsurprising if further views of Carnap suffered collateral damage. Finally, Quine endorsed a naturalist program of explaining how we form and adjust our system of beliefs, as opposed to a program of prescribing how we ought to form beliefs. This part of Quine’s philosophy has been less influential than his objections to the analytic/synthetic distinction, but it adds to a shift away from Carnap’s normative approach to thought and talk.

Another, equally significant source of resistance to Carnap’s metaphilosophy was the rise of the movement that became known as ordinary language philosophy. Prominent philosophers like Ryle, P.F. Strawson and Austin saw it as the philosopher’s task not to construct artificial languages or engineer additions to our language, but to better understand existing language in its ordinary use. They advocated that we pay close attention to the way terms and phrases are actually used, as opposed to the ways philosophers in our past had treated language; stretching, refining and creating terms as they saw fit. One core idea behind the movement, influentially argued by Wittgenstein in *Philosophical Investigations* (1953),

is that philosophers entangle themselves in unsolvable puzzles (e.g. realism v. idealism), only in virtue of stepping beyond the limits of ordinary language.¹⁰

Again for the sake of nuance, it is worth noting that the ordinary language philosophers were not blind to the need for or value of normative and revisionary thought about language. If Austin is representative, the point is rather that we should not deviate from ordinary language with a light heart:

First, words are our tools, and, as a minimum, we should use clean tools: we should know what we mean and what we do not, and we must forearm ourselves against the traps that language sets us. [...] Thirdly, and more hopefully, our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our armchairs of an afternoon—the most favoured alternative method. (Austin 1956, 7–8)

Despite not naïvely trusting the concepts we have, to borrow Nietzsche's phrase, inherited from our ancestors, Austin thought that we would grasp most of the concepts we need to grasp if only we took sufficient care to understand language in its ordinary use. Allowing for exceptions, the advice is to stick to ordinary language. This advice is directly antithetical to Carnap's view of ordinary language and his program of conceptual engineering, so there is no need elaborate how these metaphilosophies clash. What is noteworthy for our purposes is that ordinary language philosophy became a dominant force in the world of English-speaking

¹⁰ Interestingly, Carnap has a similar argument against the value of traditional, metaphysical debates such as whether there are numbers. Using a distinction between questions internal and external to a language form, Carnap argues that the question "Are there numbers?" is devoid of cognitive content when understood as an external question, and should be replaced by something along the lines of "Should we admit talk of numbers in our language?" (Carnap 1950).

philosophy. As it did, the environment became less accommodating to projects in or about conceptual engineering.

2.3 Thought experiment-based philosophy

If there is one characteristic to define the method favored in English-speaking philosophy from the 1960's onwards, it is the use of thought experiments as a means for assessing theories.¹¹ Let us take a traditional view of the practice, understanding the theories thereby tested as *conceptual analyses*. The core claim of a conceptual analysis is, paradigmatically, a bi-conditional statement. The right-hand side of the statement is intended to provide individually necessary and jointly sufficient conditions for the application of the concept on the left-hand side. Where the target concept is expressed with a one-place predicate 'C', this may take the form of

3) x is C if and only if x is F, x is G and x is H

where 'F', 'G' and 'H' denote properties. Critics of 3) may then venture into thought experiments to probe circumstances where the proposed application conditions and the concept come apart. If they find a case where a particular has properties F, G and H, but C does not apply, it shows that the conditions were jointly insufficient. If they find circumstances where C applies to a particular that is not F, or not G or not H, the condition that fails to obtain is proven unnecessary. In either case, 3) would face a counterexample. The analysis would therefore fail: It would need to be amended if not discarded (perhaps in favor of an altogether new analysis). The circumstances probed may be real or hypothetical: Since the analysis is supposed to provide conditions that obtain whenever the concept applies (and/or always then), it does not matter whether the case is actual or merely possible.

¹¹ What Machery (2017) calls "the method of cases".

Edmund Gettier set the example for this methodology with “Is Justified True Belief Knowledge?” (1963), where he used a thought experiment to convince the philosophical community to reject the long-standing belief that knowledge could be analyzed as true, justified belief. Gettier identifies a set of scenarios where an epistemic agent has a true, justified belief in a proposition, but yet fails to know what she believes. Other thought experiments are not hard to find. Putnam’s Twin Earth, Kripke’s Gödel case, Burge’s arthritis patient, Goldman’s barn façade county, Singer’s drowning child, Searle’s Chinese room, Jackson’s Mary, countless variations of trolley cases in ethics, countless hallucination and illusion cases in the philosophy of perception, countless cases of preemption and over-determination in the philosophy of causation. The philosophical literature of the last six decades is full of thought experiments, and several of the mentioned thought experiments are from exceptionally influential philosophy publications in this period. Many of these thought experiments are widely considered to be devastating to the (hitherto celebrated) theories they undermine.

Thought experiment-based philosophy is non-revisionary. When philosophers make a verdict about a case, they go on to make straightforward inferences about the truth of the theory or analysis that the thought experiment was designed to test. For example, we judge that the subject in a Gettier-case does not know that *p*, despite having a justified, true belief that *p*. From this it follows that the analysis of our concept of knowledge—as justified, true belief—is false. On the prevailing methodology, we are not tasked with further evaluation of whether our concept of knowledge *ought* to have the content it has, and therefore, whether our case verdict is in line with the concept of knowledge we ought to be using.

The method remains non-revisionary if we think of the theory being tested not as an analysis of our *concept* of knowledge, but as an object-level theory about knowledge. After engaging in thought experiment to test the theory so construed, we are not tasked with considering whether *knowledge* is the epistemic state we should be theorizing about—as opposed to an epistemic state much like knowledge,

except not exactly what happens to be the semantic value of the English word ‘knowledge’. Whether the theory being tested is a conceptual analysis or a theory about some worldly phenomenon, the truths we learn through thought experiments (supposing we do learn truths via thought experiments) are constrained by the concept we currently possess. Either it is a truth about that concept, or it is a truth articulated with that concept.

In the late 1990’s, metaphilosophical discussion reignited, when philosophers started to question the epistemic value of judgments elicited in thought experiments. After all, our judgments in thought experiments are more or less immediate responses to case descriptions and they are not corroborated by external checks. Quite simply, they are intuitions. But if it turned out that our intuitions were not reliable, then the practice of thought experiment-based philosophy would be in jeopardy. On what grounds could we rest assured that our intuitions are reliable sources of information about our concepts of knowledge, the moral right etc.? Or, using a different construal of the study object, on what grounds could we presume that our intuitions were reliable sources of information about the *nature* of worldly phenomena, such as knowledge or morality?

Doubts about the epistemic value of intuitions arose from research in cognitive psychology, most famously that conducted by Kahneman and Tversky (see Kahneman, Tversky, and Slovic 1982). They uncovered a host of cognitive biases that affect our intuitive judgments, and there was no good reason to think that philosophers were exempt. Doubts about intuition grew more acute when so-called experimental philosophers exposed social and cultural variations in judgments toward philosophical cases (e.g. Weinberg, Nichols, and Stich 2001). The findings suggest that judgments toward philosophical cases correlate with ethnicity, gender and other traits that should not bear any systematic connection (if any) to the truth of those judgments. Moreover, there is no independent court of appeals to dismiss judgments that are inconsistent with what is accepted as common ground in philosophy. In a recent book, Machery (2017) systematizes the

objections that one may make on the basis of these findings. He poses a trilemma of problems for philosophers who argue on the basis of their own intuitions about cases. Either their method is unreliable (not sufficiently truth-conducive), or parochial (truth-conducive, but simply indicative of *our* concepts), or dogmatic (dismissing judgments inconsistent with our own for no good reason).

In addition to discussions about the nature and epistemic value of intuitions, a number of other debates gained traction in the late 1990's and 2000's. As intimated, we find arguments over whether we should understand the target theories as conceptual analyses or object-level theories (c.f. Williamson 2007). What, by the way, is conceptual analysis? What is the relation between philosophy and the sciences? What can we learn from probing the judgments non-philosophers make about philosophical cases? At the moment of writing, discussions about these and similar issues are still live; some in general metaphilosophy and some in field-specific meta-theory (in metaphysics, ethics and epistemology).

2.4 The resurgence of conceptual engineering

As the metaphilosophical debate developed and disseminated, more and more philosophers found reason to question the nature, viability and value of the method they employed. In this environment, it is no surprise that a methodology *not* centered on thought experiments should gain traction. The methodology of conceptual engineering asks us to articulate the content of the concepts we ought to use in speech and thought. Reaching this objective does not hinge on using thought experiments (or, for that matter, real cases) to elicit judgments articulated with the concepts we already possess. The fact that the subject of a Gettier scenario does not know what she truly and justifiably believes is not an argument for a claim about what kind of concept of knowledge we ought to have. Entirely different kinds of reasons are required to support this kind of claim. In this way, the methodology of conceptual engineering is not vulnerable to the problems identified with the use of intuition. At best, a thought experiment serves to elucidate the content of a

concept, but the program of conceptual engineering asks us not to end our inquiry here:

Let us suppose, however optimistically, that the problem of the analysis of the everyday meaning of ‘know’ had both been shown to exist and subsequently solved, so that agreed necessary and sufficient conditions for the ascription of knowledge were now on the table. That would be a considerable technical achievement, and no doubt a long round of hearty applause would be in order, but I hope that philosophers would not regard it as a terminus, as many writers make one feel they would. I should like it to be seen as a prolegomenon to a further inquiry: why has a concept demarcated by those conditions enjoyed such widespread use? (Craig 1990, 2)

The program of conceptual engineering asks us to go further still. Not only should we ask the descriptive question Edward Craig asks here, of why we have the particular sort of concept of knowledge we do have. We need to ask what kind of concept of knowledge we ought to be using. Given that our concept of knowledge serves a particular purpose or set of purposes, it is conceivable that there is a concept that better serves those purposes than the concept we in fact find ourselves with.

Interest in conceptual engineering develops over the 2000’s and the 2010’s. Articles with titles containing ‘conceptual engineering’, ‘conceptual ethics’ and ‘metalinguistic negotiation’ show up; several revisionary works are published or attract attention, e.g. Haslanger’s (esp. 2000) work on gender and race; and renewed interest in Carnap’s philosophy intensifies—something that had already begun in the 90’s, but now, in the 2010’s, with increased attention to Carnap’s method of explication. It is beyond both my knowledge and the scope of this historical brief to detail all the ways in which normative and revisionary thought returned to the center of attention, or to offer individual explanations for why particular philosophers chose to engage in projects in or about conceptual engineering at this point in time. The aim of this section has been to narrate a

history of conceptual engineering in philosophy in accurate but broad strokes. The fact that thought experiment-based philosophy faced a methodological reckoning around the turn of the millennium offers us a preliminary explanation for why the time was ripe for conceptual engineering to make a resurgence. Now, to say that conceptual engineering has made a resurgence is not to deny that one may be able to identify strands of conceptually revisionary argument in philosophy between 1960 and 2000. However, these strands of revisionary methodology are overshadowed by the prevalence and centrality of thought experiments in this period. At the time of writing, the presence of revisionary thought about concepts is more pronounced than it has been in a long time, and only since around 2010 has there been a systematic debate about conceptual engineering: What it is, how it should be done, whether it is viable and what its value is.

The advice to conceptually engineer does not in any useful sense amount to saying “good job philosophers, keep doing what you’re doing!”, nor to propose that we engage in a practice without any precedence, as per Deutsch’s criticism. The advice to conceptually engineer is a recommendation that we ask questions of the form “How ought we conceive of X?” and “What concept of X best serves our purposes for having a concept of X?”, rather than the traditional question frames “What is X?” or “What is it to be X?”. The advice to conceptually engineer asks us to embrace theoretical virtues that have not prevailed in the philosophy of the last few decades. Not least, the advice to conceptually engineer asks us to look beyond the limits of what we can learn by engaging in thought experiment. While conceptual engineering is not a 21st century innovation, engaging in it is to depart from the method(s) that have dominated philosophy for the last six decades.

3. The essays

This work consists of the present introduction and four essays (chs. 2–5). The essays are independently readable articles on the theoretical foundations of conceptual engineering.

The essay that makes up chapter 2, “Conceptual Engineering and the Implementation Problem”, is a minimally edited version of a paper that has appeared in *Inquiry* (Jorem 2021). It discusses an issue that Herman Cappelen (2018) takes to be a significant obstacle to conceptual engineering. The implementation problem, as I call it, is the problem of implementing the concept or meaning one has designed for an expression, and so, of having an impact on actual thought and talk. It confronts us when we assume that meaning is determined by factors beyond our control (as many philosophers of language believe), and also assume that conceptual engineering is the project of trying to change meaning. Taken together, these two assumptions suggest that conceptual engineers are not able to make the changes they aim for. I argue that the implementation problem dissolves if we expand our view of how one may implement the product of an exercise in conceptual engineering. I describe four implementation options: *Standing Meaning*, *Meaning Modulation*, *Speaker-Meaning* and *Different Language*. I argue that the latter three options constitute feasible and worthwhile ways by which conceptual engineers could achieve what they aim for. Even the first option, of changing standing meaning, is less problematic than what some authors have claimed, because they have not accounted properly for the fact that languages come in different sizes, as measured by the size of the speaker population. I thus argue that the implementation problem is not a problem worth worrying about.

In chapter 3, “The Good, The Bad and the Insignificant – Assessing Concept Functions for Conceptual Engineering”, I discuss an approach to conceptual engineering that is centered on the idea that the *functions* of concepts play an important methodological role for projects in conceptual engineering. The so-called functional approach is attractive, because it offers a promising explanation of what makes a concept good, and a plausible account of the normative limits of revision. I raise a problem for the functional approach that I call “the selection problem”. The selection problem confronts us when we realize that there are

functions that we should not (or need not) engineer in the service of. Some functions are harmful, and some are insignificant for purposes of determining the limits of revision. Therefore, we have to make a selection among the functions a concept serves if we want to determine what revisions of the concept should be like. However, we only have question-begging ways of specifying the range of functions conceptual engineers ought to engineer in the service of. I review existing ways of operationalizing the notion of the function of a concept. None of the notions on offer discount the harmful or methodologically insignificant function candidates that pose a problem for the functional approach. I elucidate the consequences this has for the functional approach. I argue that the selection problem, despite being a real problem, should not lead us to reject the functional approach.

The essay in chapter 4, “Inferentialist Conceptual Engineering”, is co-authored by Guido Löhr, Eindhoven University of Technology. An inferentialist view of concepts identifies concepts with their role in reasoning. Our inferentialism concerns conceptual engineering specifically. We give an argument for conceiving of what is being engineered in conceptual engineering projects as the things that inferentialists identify concepts with. We argue for this view by way of an objection to the rationality of engaging in conceptual engineering. The objection is based on a representationalist understanding of what is being engineered. On the representationalist view, conceptual engineering is the practice of changing extensions via changes in intension. If what is being engineered is simply a device for accessing extensions (via intensions), conceptual engineering has a bad rationale. However, the conclusion that conceptual engineering has a bad rationale may be resisted if we reject the premise that what is being engineered is a representational device. Thus, we assume that conceptual engineering is worthwhile, and argue that this supports being an inferentialist about what is being engineered. Only if what is being engineered is something that has practical and theoretical consequences of application do we have reason to care about what it takes for it to apply. In addition to accounting for the assumption that the practice

is worthwhile, inferentialism about conceptual engineering offers a plausible interpretation of what is at stake in concrete examples.

The essay in chapter 5 is titled “Against Amelioration as Revelation”. I criticize a group of externalist views—advocated in various forms by Haslanger (2006; 2020), Ball (2020) and Sider (2011)—that I call “Amelioration as Revelation”. According to Amelioration as Revelation, the metasemantics of our languages have made our terms mean what they ought to mean. It follows that conceptual engineers can rarely if ever improve the meaning of sub-sentential expressions. According to the view, many of the cases that we might think of as conceptual engineering or meaning revision are actually cases of *revealing* what a word means. For example, if meaning is determined by “future reasonable agreement”, as Ball claims, then the word ‘planet’ did not change meaning when the International Astronomical Union voted to define ‘planet’ so as to give the word the meaning it currently has, but rather revealed the meaning it has had all along. I argue that Amelioration as Revelation commits us to unacceptable verdicts about semantic change and that it fudges the difference between normative and descriptive assessments of meaning. In doing so, it deprives us of an important way to evaluate speech. Once we recognize that we may assess meaning normatively, there is no need to bake norms of concept selection into our metasemantics.

As a collection, the four essays support a favorable and optimistic attitude towards conceptual engineering. Each essay addresses, directly or indirectly, objections one might have to the practice:

Chapter 2 defends conceptual engineering against the objection that engagement in the practice is futile (because it is not in our power to change meaning).

Chapter 3 contributes to the development of a functional approach to conceptual engineering, albeit critically. In itself, it is important to get the functional approach right. This is all the more important since the approach offers what is, in my understanding, the best response to the objection that revising a concept achieves

little more than “changing the subject” (c.f. Strawson 1963; Cappelen 2018, chap. 9).

Chapter 4 addresses an objection to the effect that conceptual engineering has a bad rationale. We show how that objection may be resisted by endorsing an inferentialist view of what is being engineered. In illustrating the inferentialist view, it also shows why and how it matters what it takes for a concept to apply—thus strengthening the positive case for engaging in conceptual engineering.

In Chapter 5, I criticize views that threaten to undermine what is unique about conceptual engineering. According to the views criticized, projects in (would-be) conceptual engineering involve changes in belief, not changes in meaning. If this were true, there would hardly be anything special about the methodology of conceptual engineering to set it apart from traditional methodologies centered on the aim of producing true beliefs (compare with the aim of identifying useful and benign concepts that we in turn can use to articulate truths).

Another recurring theme in the essays is a pragmatist outlook:

Chapter 2 considers the practical value of particular implementation options. Since conceptual engineers can make *worthwhile* and *feasible* accomplishments without changing the standing meaning of a natural language expression, there is no need to hold conceptual engineering hostage to the latter implementation option.

Chapter 3 offers reasons in favor of the view that what we *use* a concept for puts a normative constraint on how we may revise it. Although I raise a problem for the functional approach, the problem only shows that we need to reconsider what the functional approach should consist in, not that the view should be discarded altogether.

The inferentialist view we defend in Chapter 4 makes use of a view of meaning that has been championed by prominent pragmatist philosophers Wilfrid Sellars and Robert Brandom. According to our argument, conceptual engineering has a sound

rationale only if what is being engineered is something that has practical and theoretical consequences of application. What is being engineered cannot simply be a device for accessing things in the world in representation; it needs to be something with significance for action or further thought and speech.

In Chapter 5, I argue that normative assessments of meaning have independent merit. We do not need to deem an utterance to be false in order to have an objection to it: Considerations pertaining to, say, the moral quality or usefulness of a word-meaning pair (or the use of one) stand on independent grounds. Moreover, such considerations do not need to be baked into our hypotheses about how meaning is determined, as if we were forced to do all our speech evaluation with descriptive assessments of meaning and assessments of truth based on the latter.

4. Method

The topic of this dissertation, conceptual engineering, is a methodological framework in philosophy. I conceive of what I am doing in addressing this topic as itself, at least in part, an exercise in conceptual engineering. This is to say that several of the arguments I present in the essays hinge on (or develop) normative considerations about which concepts we ought to be using. Thus, the essays do not (only) aim at finding non-normative truths about or articulated with our existing concepts. Below, I give a few of examples where I think this methodological strand is particularly clear, but first a few words about the method I employ at a rudimentary level of description. Note that my purpose in saying something about the method I have employed is to share how I, in broad outline, conceive of it, not to give a detailed methodological account. The essays are supposed to stand on their own feet.

As most works in philosophy, the present work is not empirical in any useful sense of the word ‘empirical’. I do not use any data to test my hypotheses. Rather, the arguments are driven by *assumptions* of various kinds. The range of assumptions

I make is constrained by my intended audience. For the project as a whole, the intended audience consists of philosophers who have some kind of interest in conceptual engineering. However, there are fine-grained shifts in my intended audience between the essays, and sometimes between sections. For example, in chapter 4, I address theorists who are enthusiastic about conceptual engineering, but think of meaning and content in representationalist terms. Taking as an assumption that conceptual engineering is worthwhile, I argue that one should not think of what is being engineered in austere, representationalist terms. This assumption would obviously be inadmissible if my audience in this essay consisted of representationalist philosophers who are skeptical of the value and viability of conceptual engineering, such as Deutsch (2020), but they are not.

My arguments consist for the most part of pieces of informal reasoning. Several of them may be faithfully paraphrased into formal, deductive inferences. For example, in section 3 of chapter 2, I describe a case of a language used by two speakers, Tim and Martha, who have made a decision to deviate from ordinary English. After describing the case, I make use of *modus ponens*: “If Tim and Martha would succeed in the changing the language they use to communicate with each other, the implementation problem does not affect very small languages. Since it seems that they would succeed, there seem to be languages for which it is feasible to change standing meaning.” Other arguments are inferences to the best explanation, i.e. *abductive* arguments. For example, in section 2 of chapter 3, I argue that the functional approach to conceptual engineering gains *abductive* support from its construal of the limits of revision. It does so because it can explain the idea of there being normative limits of revision as well as our verdicts particular cases of crossing those limits.

As intimated, some of the arguments only make sense in a methodological framework that is conceptually normative or revisionary. For example, the essay in the third chapter investigates what kind of concept of function we ought to be using to identify normative constraints on conceptual engineering. The aim is not to

capture the content of our concept of a concept function, nor to give a true description of the “nature” of concept functions. Rather, it is to figure out what kind of concept of function we are best served with for the purpose of developing the functional approach to conceptual engineering. Similarly, chapter 4 treats our concept of what is being engineered in conceptual engineering exercises as itself a target of conceptual engineering. We ought to endorse an inferentialist understanding of what is being engineered not because it is *the* true theory about what is being engineered, but rather because the things inferentialists think that concepts are would be a reasonable thing to engineer. In chapter 5, meanwhile, I consider what the point of semantic and metasemantic theorizing is. Indirectly, these considerations say something about what kind of concept of meaning we have need for.

To summarize, I conceive of the method I employ as assumption-based, informal reasoning that aims at a more useful understanding of the topics and phenomena discussed.

Chapter 2

Conceptual Engineering and the Implementation Problem

1. Introduction

Conceptual engineers attempt to revise or replace our defective representational devices. They purport to revise meanings and replace concepts. In order to do so, they have to be able to affect that which makes it the case that an expression means what it means, or that which makes it the case that we think by means of the concepts that we in fact do. To use a term of art: They have to hook up with the *metasemantic functions* governing our thought and talk. If conceptual engineers fail to do so, they will not succeed in effecting semantic or conceptual change. To their apparent peril, it is widely believed that the functions governing our thought and talk include *externalistic* functions. Externalistic functions take as input such matters as what experts believe, what the actual world is like, the naturalness of the candidate meaning and patterns of use that have been initiated by baptisms long past (Burge 1979; Putnam 1975; D. Lewis 1984; Sider 2011; Kripke 1980). These matters are largely beyond our control. To make matters worse, we do not know the details concerning how the metasemantic functions operate, individually or in concert.¹ It would seem that the conceptual engineer is like a blindfolded oarsman, foolishly challenging ocean tides.

Apparently then, metasemantic functions obstruct our efforts at implementing the product of an exercise in conceptual engineering. Let us call this

¹ Herman Cappelen (2018) makes the stronger claim that metasemantic functions are inscrutable. See Schroeter and Schroeter (2020) for criticism.

the implementation problem. Several authors express worries about it.² In this paper, I focus on linguistic meaning. I argue that the implementation problem only has bite if we take an unjustifiably restrictive view of what it is to implement a meaning. On this restrictive view, the only way to implement a meaning is to change the standing meaning of an expression in an ordinary language like English. Against this, I direct attention to the fact that there are multiple ways a conceptual engineer could implement the meaning she prescribes. I examine the following, possibly non-exhaustive list of alternatives:

Standing Meaning. She could cause semantic change so that the standing meaning of a natural language expression changes to comply with her prescription.

Meaning Modulation. She could get people to modulate the meaning of an expression in context so that it complies with her prescription.

Speaker-Meaning. She could get people to use an expression with speaker-meanings that comply with her prescription.³

Different Language. She could get people to speak a constructed language in which a particular expression has the meaning she prescribes.

² The implementation problem was first noted with reference to externalist metasemantics by Burgess and Plunkett (2013a), and Plunkett and Sundell (2013). Cappelen (2018) claims that it constitutes a significant obstacle and discusses it at length. Deutsch (2020) argues that we cannot stipulate the semantics of terms. Koch (2021a) holds that externalism poses a real hurdle, but argues that we nonetheless have ‘collective long-range control’ over meaning. Riggs (2019) disagrees with these contributions. He argues that if some metasemantic thesis is true such that we do not have control over what our words mean, then conceptual engineers did not want to change what our words mean.

³ Pinder (2021) argues that we should view conceptual engineers as engaged in the project of changing speaker-meaning. I agree with Pinder that it is feasible and worthwhile to change speaker-meaning, but see no reason to restrict conceptual engineering to this one implementation option.

Think of the above as options a conceptual engineer could pursue, if she were to have an intentional plan for implementation. I clarify them later. We are interested in two properties of these implementation options. First: Is the option feasible? Are conceptual engineers in a position to do what the implementation option requires of them? The considerations given in this introduction suggest that e.g. *Standing Meaning* may be unfeasible. Second: Given that the option is feasible, is it worth pursuing? Does it get conceptual engineers what they could reasonably desire? Many conceptual engineers want to promote more exact, precise and fruitful thought and speech. Many want to promote social justice. If an implementation option rendered such goals unattainable, the feasibility of that implementation option would be of little comfort.

To establish that conceptual engineers face an implementation problem worth worrying about, it is insufficient to show that one of the above options are unfeasible. In particular, it is not sufficient to show that it is unfeasible to change the standing meaning of a natural language expression. Rather, one would have to argue *for each* of the above options, that they are unfeasible or that they are not worth pursuing. In the absence of good reason to dismiss each option, we are justified in holding that the implementation problem is insignificant; that it is not truly a problem. One might object to this line of thought if one held that conceptual engineers' *sole* objective is to change the standing meaning of natural language expressions. However, I argue against this claim in section 3.

The plan is as follows. In section 2, I briefly describe what conceptual engineering is. I present an example of a project in conceptual engineering, identify the prescribed meaning and distinguish ways in which it may be implemented. In section 3, I discuss *Standing Meaning*. I explain why it is largely unfeasible to change standing meaning, but also why we should ask for other implementation options. In sections 4—6 I discuss, in order, *Meaning Modulation*, *Speaker-Meaning*, and *Different Language*. I argue that they constitute feasible and worthwhile options for implementing a meaning.

2. Design and Implementation

The main rationale for engaging in conceptual engineering is the contention that there are devices we could use to speak and think that are superior to the devices we currently use to speak and think. The devices in question may be linguistic meanings, such as the meanings of ‘marriage’ and ‘cause’, or concepts, such as a concept of marriage or a concept of causation.⁴ To get some purchase on the contention that some meanings and concepts are better than others are, consider first an example of a concept designed to be of little use. Consider the concept of *an electron or cow*, as opposed to the concept of *an electron* (c.f. Sider 2011, 1–2). It is not by chance that physicists use the concept of an electron rather than the concept of an electron or cow. Zoologists are equally prudent to avoid the latter concept. The union of the set of cows and the set of electrons does not permit any interesting generalizations for either theorist. For a different example, consider a concept of marriage that applies exclusively to couples consisting of one man and one woman. One could organize social life around this concept, viz. by assigning a range of legal, moral, practical and symbolic significances to falling in its extension. However, organizing social life around such a concept of marriage would arguably breed or constitute discrimination of non-heterosexuals. It is better to organize social life around a concept of marriage that also applies to same-sex couples. Such a concept would not engender harm by discrimination. We could thus argue, on moral grounds, that the inclusive concept of marriage is a more benign concept of marriage. In general, we can evaluate concepts as better or worse along various

⁴ In this paper, I individuate concepts and meanings at a fine granularity (and words at a coarser granularity). This convention has implications for matters of conceptual variation, such as whether your concept of knowledge is the same as my concept of knowledge, or whether ‘knowledge’ in your idiolect has the same meaning as ‘knowledge’ in my idiolect. I want to avoid presuming that there is no cross-individual, cross-cultural or cross-societal conceptual variation where there in fact is such variation (c.f. Machery 2017). That would require avoiding locutions such as ‘the concept of F’ as if there were one and only one concept of F. For ease of reading, I sometimes indulge in presuming definiteness, as in ‘the concept of an electron’.

dimensions.⁵ Conceptual engineers attempt to design concepts and meanings that will serve us well in some capacity.

Conceptual engineers also want to implement the meanings they have designed. The implementation stage of conceptual engineering is the topic of this paper. For the purposes of evaluating claims about implementation, it may be helpful to have an actual example of conceptual engineering in mind. In that errand, let us consider James Woodward’s (2003) work on causation. Woodward presents his theory of causation as a revisionary account—as an account of what we ought to mean by ‘cause’ and ‘causal explanation’:

[M]y project has a significant *revisionary* or *normative* component: it makes *recommendations* about what one ought to mean by various causal and explanatory claims, rather than just attempting to describe how we use those claims. [...] we introduce concepts (including concepts of cause and explanation) and characterize them in certain ways at least in part because we want to *do* things with them: make certain distinctions, describe certain situations (which usually requires being able to tell whether the concept applies, on the basis of evidence that we have some possibility of getting), calculate with them, use them in proofs or arguments, and so on. Concepts can be well or badly designed for such purposes and we can evaluate them accordingly. (Woodward 2003, 7–8)

Woodward wrote *Making Things Happen* before the terms ‘conceptual engineering’ and ‘conceptual ethics’ were popularized, but his methodological commitments are squarely within such a framework. To fulfill his commitments, he constructs multiple concepts of causation and corresponding meanings for ‘cause’. If his

⁵ See Burgess and Plunkett (2013a; 2013b) for an introduction to conceptual ethics, Ludlow (Ludlow 2014, 41–51) for criteria guiding modulation of meaning, Cappelen (Cappelen 2018, 34) for a taxonomy of conceptual defects, Carnap (1962, chap. 1; 1963b, 933–40) for a classical exposition of science-friendly conceptual virtues, and Brun (2016) and Dutilh Novaes and Reck (2017) for further discussion of Carnapian virtues.

project is successful, we can use these meanings when we want to make more exact, precise and fruitful causal claims than what we can make with the ordinary meaning of the word ‘cause’. Woodward engineers four different concepts: of a total cause, of a direct cause, of a contributing cause and of an actual cause. Here, I simplify his account and focus only on his concept of a total cause. Here is how he presents this concept:

(TC) X is a total cause of Y if and only if there is a possible intervention on X that will change Y or the probability distribution of Y. (Woodward 2003, 51)

Notice that this statement is not explicitly metalinguistic or metaconceptual. It makes no mention of a concept, a meaning or a linguistic expression. Still, it is clear from how Woodward has framed his project that (TC) plays the role of presenting a concept or meaning—a concept or meaning he argues we should use.⁶ This raises the question: Precisely which meaning is being presented? For lack of plausible alternatives, we may identify the meaning as one for ‘total cause’ such that (TC) would hold true if ‘total cause’ in fact expressed it (never mind whether it actually does so). Let us call this meaning ‘Woodward’s prescribed meaning’.⁷

By assumption, Woodward’s prescribed meaning is non-identical to the meaning of the English word ‘cause’. Yet it does not require a major departure from actual semantic fact to imagine ‘cause’ having Woodward’s prescribed

⁶ Notably, (TC) has the same structure as other core theoretical statements in recognized works in conceptual engineering, e.g. Haslanger (2000). I do not assume a particular account of this kind of speech act. One option is to interpret it as part of a metalinguistic negotiation (Plunkett and Sundell 2013; Plunkett 2015). The important distinguishing feature is that the arguments in favor of accepting (TC), as opposed to a superficially identical statement—e.g. a conceptual analysis or a claim about ‘the nature of’ causation—have to do with conceptual ethics.

⁷ Strictly speaking, multiple possible meanings for ‘total cause’ could sustain the truth of (TC), given a fine-grained individuation of meanings (cf. fn. 4). I do not assume that ‘Woodward’s prescribed meaning’ picks out a single, unique meaning at the finest level of granularity, but rather a range of meanings so individuated.

meaning. In other words, ‘cause’ could have had a meaning such that the following sentence would be true:

(‘TC’) X is a cause of Y if and only if there is a possible intervention on X that will change Y or the probability distribution of Y.

If Woodward somehow succeeded in changing the standing meaning of the word ‘cause’ in English, in compliance with his prescription, that would make it the case that the sentence (‘TC’) become true in English. Woodward would then have succeeded in implementing his prescribed meaning in a particular way: He would have implemented it *for* ‘cause’, *at* the level of standing meaning *in* English.

It is crucial for my argument that there are other ways of implementing a meaning, beyond changing the standing meaning of a particular word in a particular language, such as changing the standing meaning of the word ‘cause’ in English. However, once we grasp the identity of the meaning in question, it is readily conceivable how it could be implemented in other ways. The very same meaning we identified above could

- (1) be the meaning of ‘cause’ on an occasion of use,
 - (2) be a component of a content communicated by a speaker, by means of the word ‘cause’,
- or
- (3) be the meaning of the word ‘cause’ in a different language than ordinary English.

Here, (1)—(3) correspond to the implementation options I sketched in the introduction, i.e. *Meaning Modulation*, *Speaker-Meaning* and *Different Language*, respectively. I argue in later sections that some or all of these implementation options are not only conceivable, but also feasible and worth pursuing.

Note finally that it is not necessary to target ‘cause’ as the word for which to implement Woodward’s prescribed meaning. Woodward himself uses a different lexical item in his actual theoretical statement (‘total cause’). My reason for focusing on ‘cause’ as the vehicle for Woodward’s prescribed meaning has to do with the idea that it is metasemantically more problematic to implement a meaning by way of a pre-existing lexical item (*lexical improvement*) than by way of a novel expression (*lexical expansion*). A pre-existing lexical item has a history of use and a novel expression does not. If we should not worry about the implementation problem with respect to cases of lexical improvement, there is *inter alia* no reason to worry about the implementation problem with respect to cases of lexical expansion. It is no less worth noting that the lexically expansive variety of conceptual engineering may be both feasible and worthwhile on independent terms.⁸

I now turn to examine our implementation options. In the next section, I discuss *Standing Meaning*.

3. Standing Meaning

On an innocuous description, conceptual engineers purport to revise what words mean. This description suggests that conceptual engineers try to change the standing meaning of natural language expressions. For example, Woodward would be in the business of trying to change the meaning of the word ‘cause’ in English. His efforts would be futile: It is unfeasible to change the meaning of ‘cause’ in English. The unfeasibility of changing the standing meaning of expressions in languages like English has impelled Cappelen (2018), Deutsch (2020) and, albeit less so, Koch (2021a) to hold that conceptual engineers face an implementation problem worth worrying about.

⁸ The distinction maps on to what Chalmers has called *de novo* engineering and conceptual re-engineering (Chalmers 2020). For further discussion of lexical expansion vs. lexical shift/improvement, see Armstrong (2016) and Cappelen (2020).

My main argument in this paper is that we have no reason to worry about the implementation problem once we expand our view to other implementation options. I discuss feasible options in later sections. In this section, I examine *Standing Meaning* in closer detail. First, I elaborate on why we should think that it is largely unfeasible to change standing meaning. Second, I show that the unfeasibility of changing standing meaning is a language-specific feature. We may not be able to change what words mean in English, but we need to consider other languages. In particular, we need to consider idiolects and other small languages. Neither Cappelen (2018), Koch (2021a) nor Deutsch (2020) assess particular languages for the feasibility of effecting semantic change in them. However, just as we should take into more serious consideration other meaning-phenomena than standing meaning (and corresponding implementation options), we should consider the feasibility of changing standing meaning across languages. Finally, I argue that we should not restrict conceptual engineering to the project of trying to change standing meaning.

To appreciate why it is largely unfeasible to change standing meaning, consider a word whose semantics has earned much theoretical attention: ‘water’. It is a semantic fact that ‘water’ denotes stuff composed of H_2O -molecules. This would hold true even if many, perhaps most, English-speakers lacked knowledge of the chemical composition of water. But why does ‘water’ denote H_2O -composed stuff? Which facts ground this fact? The following facts have all been adduced to explain why ‘water’ denotes H_2O -composed stuff:

H_2O -molecules compose the stuff that comes out of our taps and fills our lakes.

Chemists believe that water is H_2O .

Our use of ‘water’ stands in a complex causal relation to stuff composed of H_2O -molecules.

Stuff composed of H_2O -molecules is a more natural referent for ‘water’ than are alternative candidates.

Facts about (complex) patterns in the use of ‘water’, including how it was introduced.

I draw these *explanantia* from celebrated proposals in the externalist tradition, including Putnam (1975), Lewis (1984), Kripke (1980), Burge (1979) and Devitt (1981). We may ignore the differences between the views. The interesting feature is common to them: The listed metasemantic determinants are largely beyond our control.

There must also be metasemantic determinants within our control. H_2O -composed stuff is not, as such, a more natural meaning candidate than NaCl -composed stuff, so naturalness cannot by itself select the former for ‘water’ and the latter for ‘table salt’. Similarly, the fact that H_2O -composed stuff comes out of our taps cannot singlehandedly make the word ‘water’ denote the stuff so identified. The fact that ‘water’ denotes H_2O -composed stuff must have *something* to do with how English-speakers use and have used that word. Indeed, I mentioned facts about complex patterns in the use of ‘water’ in the previous paragraph, but claimed that they were largely beyond our control. Surely, however, the right thing to say is that such patterns are beyond our *individual* control, not beyond our *collective* control. As a collective, we could start using a word differently, and this may provide leeway to effect semantic change. Steffen Koch (2021a) has argued that we have ‘collective long-range control’ over meaning on roughly these grounds. He likens the control we have over meaning to the control we have over the climate, and concludes with ‘modest optimism’ about the prospects conceptual engineers face. The idea is that a conceptual engineer could succeed in effecting semantic change if she could cause a significant portion of the linguistic population to start using a target expression differently, viz. in accordance with her prescription. It is clearly possible to change the way people use words, hence the modest optimism.

However, our modicum of optimism should vaporize when we take a moment to consider just how many speakers one would have to influence in order to cause a change in the standing meaning of ‘water’ or ‘cause’ in this way. While it is not metaphysically impossible to change the meaning of ‘water’ or ‘cause’, it is beyond what we may count as practically feasible.⁹ The same would hold true for most expressions of most ordinary languages. This result should not come as a surprise. The very idea of *standing* meaning—of the meaning a word has in a language; its lexical meaning—is about something stable across speakers, times and places. It cannot be an easily manipulable commodity. Thus, the problem is not so much that there are externalistic determinants of meaning. The standing meaning of a word could be determined exclusively by internalistic factors—such as what the speakers of a language intend or mean by a word—and yet it would be unfeasible to change because it were practically impossible to change sufficiently many intentional states. Arguably, it is within our power to change some of the relevant intentional states. One could e.g. give convincing arguments for what we should mean by a word. But in languages like English there are just too many individuals one would have to affect. Whether we endorse some form of externalism or some form of internalism, then, the standing meaning of expressions in languages like English will be beyond our control.

The reference to English is non-accidental. When assessing the (un)feasibility of changing the standing meaning of natural language expressions, we need to keep in mind that expressions only have meaning in virtue of belonging to a particular language. There is no such thing as the semantic value of a sound or inscription unless we conceive of the sound or inscription as a particular expression in a

⁹ Neither Cappelen (2018) nor Deutsch (2020) assume that it is metaphysically impossible to change meaning, so my assuming that is metaphysically possible to change meaning does not beg the question against their views. Clearly, there are metasemantic views (in the vast space of logically possible views) on which any given semantic fact depends exclusively upon facts that are utterly beyond human control. I do not address such views.

particular language. This much is evident when we consider the sound we make when we pronounce ‘see’. The sound means different things in English ([\approx view] or [sea]), Spanish (‘sí’ [yes]) and Norwegian (‘si’ [say]), and it means nothing in e.g. Finnish or Arabic. This point can get lost on us when we are inquiring into what kind of metasemantic determinants natural languages are subject to in general, but the identity of the language matters in the present context. It matters because we cannot infer that it is unfeasible to change the standing meaning of natural language expressions from the fact that it is unfeasible to change the standing meaning of English expressions. In particular, we should assess smaller languages.

Let us therefore consider an opposite extreme of English with respect to size: The language used by two people, Tim and Martha, to communicate with each other. Tim and Martha have frequent discussions about all sorts of philosophical and non-philosophical issues. With other people, they speak plain English, but when they are discussing with each other they often find reason to use expressions that are not part of ordinary English, or to use English expressions differently. An empirical linguist could therefore investigate the language they use to communicate with each other as a language in its own right. Now, suppose Tim and Martha dislike the intension of ‘water’ in English (according to the semantics we are taught about that word). Tim says, ‘Let’s start speaking as if “water” denotes anything that has the sensory and practical properties of the stuff that comes out of our taps, but could vary in chemical composition.’ Martha replies ‘Brilliant, let’s go.’

First, it is clear that they could, between the two of them, go on to use words in accordance with their agreement, thereby doing everything in their power to institute the wanted meaning in the (two-speaker) language they use to communicate with each other. Second, we need a good reason to believe that, once they have, other metasemantic determinants will counteract. According to one strand of externalism, comparatively natural meaning candidates exert a kind magnetism, so that terms come to denote them rather than alternative, unnatural meaning candidates (D. Lewis 1984; Sider 2011). But surely, H₂O’s alleged

reference magnetism would not override the conventions Tim and Martha abide by. On Lewisian metasemantics at least, reference magnetism only kicks in after the conventions of language use have determined a range of initially admissible interpretations (Williams 2007). The conventions Tim and Martha comply with do not permit us to assign H_2O -composed stuff as the exclusive denotation of ‘water’. After all, they use ‘water’ to talk about logically possible stuff that is not composed by H_2O . What about the meaning-grounding fact, identified by Putnam, that what we call ‘water’ in the actual world is composed by H_2O ? It is implausible that this would obstruct their efforts. Putnam’s Twin Earth thought experiment purports to show that psychologically identical individuals could mean different things by ‘water’ if their environments were different (Putnam 1975). The converse does not follow: Sameness of environment does not ensure that ‘water’ means the *same* thing for psychologically *distinct* individuals. Putnam never argued that our H_2O -laden environment would make ‘water’ refer to H_2O in any idiolect, no matter the beliefs, intentions and use-patterns of the speakers. Indeed, that would be an implausible claim. I project that Putnam’s argument would not convince many readers if the speakers transported to Twin Earth were Tim and Martha, as I have described them here. Tim and Martha seem to succeed in speaking a language in which ‘water’ denotes stuff not necessarily composed by H_2O .

It might help clarify this discussion to note that it does not follow that it is unfeasible to change meaning property *M* from the hypothesis that there are metasemantic determinants of *M* beyond our control, such as naturalness or past usage. As long as there also are metasemantic determinants of *M* we can control, then, provided that it is sufficient to change these to change *M*, it is feasible to change *M*. Thus, Deutsch is mistaken when he writes “[E]xternalism *does* imply that semantic shifts require corresponding shifts in external factors: stipulation, by itself, never suffices.” (Deutsch 2020, 3952–53). It may be that a speaker’s intention or act of stipulation does not suffice for a change in meaning in a given case. On the other hand, semantic change does not necessarily *require* change in external

factors, even on strongly externalist views. Externalism is the denial of internalism, not the claim that external factors exhaustively determine meaning. I am unfamiliar with any externalist account committed to the excessively strong claim that intentional states play no role in determining meaning, and see no need to address such a view.

If Tim and Martha would succeed in the changing the language they use to communicate with each other, the implementation problem does not affect very small languages. Since it seems that they would succeed, there seem to be languages for which it is feasible to change standing meaning. On the other hand, it still seems unfeasible to change what words mean in English. We therefore find that it is unfeasible to change the standing meaning of expressions in some but not all languages. The feasibility of changing standing meaning depends importantly, though not solely, upon the number of individuals that speak the language one is trying to effect semantic change in. In fact, if use-patterns and speaker intentions play any role in determining standing meaning, then, *notwithstanding the operation of other metasemantic mechanisms*, it will be more feasible to change the standing meaning of expressions the smaller the language is.¹⁰

We could also use ordinary, coarsely individuated languages to illustrate the language-relative feasibility of changing standing meaning. Consider Lule Sámi, a language spoken by 1,000—2,000 people in northern Sweden and Norway. Clearly, if use-patterns or speaker intentions play any role in determining meaning, then it is more feasible to change what a Lule Sámi expression means than it is to change what an English expression means. If a conceptual engineer changes how 1,000 speakers use a Lule Sámi expression, she would be making reasonable progress towards semantic change. If a conceptual engineer changes how 1,000 speakers use an English expression, she would barely be getting started. It is

¹⁰ A minor qualification: If, e.g., past or expert usage play important metasemantic roles, then the target languages may need to be alike in other regards, but all things considered, size still matters.

therefore more feasible to change the semantics of Lule Sámi than to change the semantics of English, much like it is more feasible to climb Mont Blanc than it is to climb Mount Everest.

This leads to a strange consequence for those committed to the claim that conceptual engineers *only* succeed in implementing the product of their work if they succeed in changing the standing meaning of an ordinary language expression. They are forced to hold that conceptual engineering is more feasible, hence easier, in small languages than in big languages. Though perhaps not absurd, this is a strange consequence for the view, and it needs to be squared with the significance one accords to succeeding in conceptual engineering.¹¹ Instead of looking for reason to maintain this view, I propose we widen our view of the implementation options conceptual engineers could pursue.

This reasoning may be further substantiated. Notably, failure to effect semantic change in an *ordinary* language does not entail failure to effect semantic change in any language spoken among the same population. To illustrate what I have in mind, suppose Woodward manages to get a set of people S to use the word ‘cause’ in accordance with his prescription. For practically realistic membership assignments to S, he would fail change the standing meaning of ‘cause’ in English. But there may be other languages he could succeed in effecting semantic change in, depending upon the exact members of S. Failing to effect semantic change in English, he could still have caused semantic change in American English¹², the language spoken by American scientists, the language spoken by experimental

¹¹ On Cappelen’s (2018, 137–47) ascription of worldly significance e.g., changing what ‘cause’ means is to change in what it is to be a cause. Arguably, however, it is not easier to change what it is for things to be what they are by working in Lule Sámi than in English.

¹² The main differences between British and American English are in spelling and phonetics, but semantic differences are not unheard of. For example, in its standard usage in the respective languages, ‘corn’ refers solely to maize in American English, but may be used for any cereal grain in British English.

scientists at the University of Pittsburgh, or... Restrictions on language individuation limit how long this list can go on, but those restrictions are not entirely clear, if, indeed, there are any hard restrictions. It is equally well worth considering counterfactual circumstances where English has fewer speakers. In some of these, the changes in behavior among S-members would constitute semantic change in English, despite not constituting semantic change in English in the actual world. While it may be that our counterfactual Woodward would succeed in something that our actual Woodward would not—viz. changing the meaning of ‘cause’ in English—it seems clear that the actual Woodward has also had a significant impact on speech and thought, never mind his failure to change the meaning of ‘cause’ in English. I suggest we attend to implementation options that can articulate the kind of success he has, irrespective of his causing semantic change in any given language.

Changes in standing meaning are metaphysically downstream from the changes in speaker behavior and speaker intention that conceptual engineers can in the first instance hope to make. Indeed, they are quite far downstream, in the sense that they are harder to effect than other changes in speech and thought. I propose we rather identify success conditions that can articulate the changes conceptual engineers have a chance at accomplishing. The implementation options described in proceeding sections do just that. As long as these other options leave conceptual engineering worthwhile doing, there is poor reason to restrict conceptual engineering to the project of trying to effect semantic change in ordinary languages. In this light, the standing meaning of expressions in ordinary languages is a rather arbitrary locus of change. Moreover, it is uncharitable to view conceptual engineers as engaged in largely futile projects, when we could view them as engaged in fruitful projects. Therefore, even though it may be feasible to change the standing meaning of expressions in some languages—small languages being the most probable—there seems to be poor reason to hold on to *Standing Meaning* as the only implementation option.

4. Meaning Modulation

We have found that it is largely unfeasible to change the standing meaning of natural language expressions—at least in most ordinary, coarsely individuated languages. For the sake of argument, we can now assume that there are no natural languages for which it is feasible to change standing meaning. It would not follow that speakers are unable to change the semantic values of expressions *in context*. And if that were the case, conceptual engineers could convince them to change them in particular ways. In this section, I discuss Peter Ludlow’s view of meaning modulation, and Herman Cappelen’s objection to the thesis of meaning control it relies upon (Ludlow 2014; Cappelen 2018).¹³

According to Ludlow (2014), analytic philosophers have too often assumed that natural languages have static lexicons with fully determinate meanings. He presents an extensive battery of examples to show, first, that meanings are more dynamic than commonly thought, and second, that meanings are generally underdetermined. The examples range from ‘book’, ‘hacktivist’, ‘journalist’, ‘doll’ and ‘planet’, to ‘rape’ and ‘person’. Conspicuously, Ludlow provides evidence that we change the meanings of such expressions on a conversation-by-conversation basis, often intentionally and often by argument. He enlists a couple of technical terms to describe this: When speakers engage in narrowing, broadening, sharpening or loosening the meaning of an expression, reflectively or not, then, if the hearer accepts, they have *modulated* that meaning. The result of modulating one or more meanings is that the interlocutors are speaking a *microlanguage*, in which the affected expressions have the meanings so modulated.

Ludlow’s view of language in *Living Words* (2014) is one of constant flux. However, his focus on demonstrating the ubiquity of meaning modulation comes

¹³ To represent Ludlow’s view fairly, note that he avoids talk of the semantic values of expressions, in favor of talk of meanings and modulations thereof. He intends to remain as neutral as possible towards ways of thinking about the semantics-pragmatics interface (Ludlow 2014, 96).

at a cost: We lose sight of the continuity of meaning and language use, about which Ludlow writes close to nothing.¹⁴ But presumably, there must be something like the standing meaning of an expression, no matter how underdetermined and subject-to-long-term-change it may be. If there were not, we would have a hard time explaining successful communication in impoverished contexts, e.g. with remote strangers (c.f. Cappelen 2018, 164–65). In addition, it is reasonable to think that the very possibility of modulating meanings requires the existence of some stable meaning that interlocutors then go on to mold in accordance with their needs, interests and values. Thus, while it might be the case that ordinary languages depend upon patterns of local use, there is surely dependency the other way around too. Acknowledging this is important for the reason that it gives externalistic determinants of meaning a stronger and perhaps more plausible role than what they have on Ludlow’s view. We see this in the following.

Ludlow specifically addresses the compatibility of his view of meaning with externalism:

The doctrine of *Meaning Control* says that we (and our conversational partners) in principle have control over what our words mean. The meanings are not fixed by convention, nor by our conversational setting alone. If our conversational partners are willing to go with us, we can modulate meanings as we see fit. [...] *Meaning Control* does not exclude the possibility of externalism about content [...] The idea is that it is within our control to defer to others on elements of the meaning of our words (for example, a doctor on the meaning of ‘arthritis’ and a botanist on the referents of ‘beech’ and ‘elm’) and it is also within our control to be receptive to discoveries about the

¹⁴ Here is an exception: ‘When we engage with others in conversation, all of the participants come to the table with a robust linguistic background already in place. We could say that the words the participants are deploying have residual value from previous entrainments.’ (Ludlow 2014, 39–40). ‘Entrainment’ refers to the outcome of coordinating meaning with conversational partners.

underlying physical structure of the things we refer to (for example, the discovery that ‘water’ refers to H₂O and not XYZ). (Ludlow 2014, 83)

Some externalistic features of meaning may be accommodated on the strength of this appeal to deference. However, certain externalists may be displeased with the idea that speakers do not use externalistic meanings until they *choose* to defer to experts or be receptive to discoveries. Cappelen gives voice to this concern:

Externalism as I have understood it—and as it has been understood in the philosophical tradition that springs out of the work of Kripke, Burge, and Putnam—is not the view that conversational partners at any one point in time can just decide that externalistic constraints on semantics don’t apply. (Cappelen 2018, 166–67)

Ludlow would have had a plausible reply if he had allowed something like the standing meaning of an expression. He could then reply that speakers cannot simply change the standing meaning of an expression by opting out of externalism. Standing meaning could remain constrained by e.g. broad patterns of use long pre-dating the conversations in question, and hence held constant. If the standing meaning of an expression is that meaning speakers use by default, this gives externalistic determinants a strong and plausible a role to play. It would still leave room for speakers to use deviant meanings locally, when they make active choices to modulate meanings (or choose to speak in ways that modulate meanings by consequence).

Pace Cappelen, I am unsure of the exact shape of the externalist theses that spring out of the work of Kripke, Burge and Putnam. Are facts about meaning in context supposed to be grounded the exact same way as facts about standing meaning, or are the former manipulable in a way that the latter are not? How much context-sensitivity is there? Cappelen’s objection to Ludlow seems to rest more on an anti-contextualist view than on externalism *per se*. Here, anti-contextualism is the thesis that the semantic values of uses of language are fixed

with no *pragmatic intrusion*: The only role for context is to fill in the gaps left open by a restricted set of context-sensitive expressions—e.g. indexicals like ‘I’ and ‘now’—not to change what terms like ‘marriage’ or ‘cause’ denote (c.f. King and Stanley 2005; Recanati 2010). One may, or may not, embrace a form of externalism that embodies anti-contextualism. If an anti-contextualist version of externalism holds true, it is unfeasible to change the semantic values of expressions in context through agreement between conversational partners. In that case, conceptual engineers could still influence people to adopt *speaker-meanings* that align with their prescription, as described in the next section. Meanwhile, there is clearly logical space for a view in the externalist family that would leave contextual meaning within the control of conversational partners. If such a Ludlow-friendly externalism holds true, speakers could modulate meanings in context. And then, if a conceptual engineer could get people to modulate meanings in accordance with their work, I submit that she would accomplish what she could reasonably desire to accomplish. If Woodward could get n people to modulate ‘cause’ in accordance with his prescription on x occasions, I see no reason for him to feel that there is something left to be accomplished; some way that he could further implement his prescribed meaning for those n people on those x occasions of use. His efforts would be worthwhile.

5. Speaker-Meaning

Now suppose that an anti-contextualist version of externalism holds true. On this externalism, there is no leeway to change the meaning of expressions in context by conversational agreement. Things being so, conceptual engineers could still aim to affect what speakers communicate or what they attempt to communicate. In other

words, conceptual engineers could attempt to affect the speaker-meanings that we could assign to the expressions people use in context.¹⁵

It takes no more than traditional Gricean pragmatics to show that what we communicate, implicate or intend to communicate, may be non-identical to what is said by an utterance, in a pure, semantic sense of ‘said’ (Grice 1989). Therefore, speaker-meanings cannot be fixed by the very same facts and functions that determine the standing meaning of the expressions used. Moreover, the abundance of examples in the pragmatics literature suggests that there are no clear limits to what we can communicate by means of an expression. Famously, we can use ‘ham sandwich’ to refer to a person (as in ‘The ham sandwich left without paying.’ where a waiter is referring to a person who ordered a ham sandwich). On these grounds, it should be feasible to use ‘cause’ to communicate one of the meanings Woodward has designed for that word.

One might worry that it is one thing to intend to communicate such a content, quite another to succeed in communicating it. Conceptual engineering would not be worthwhile if its practitioners could only initiate futile efforts at communication. Thus, one may object that I have merely substituted one problem for another. But the exchange is favorable: We have traded an intractable metaphysical problem for a feasible practical challenge. Granted, it may be difficult to communicate what one means when one is using an existing expression with a deviant meaning (or what one means by a novel expression, for that matter). However, by no means is it impossible. We can always say what we mean. For example, I can say that I am using the word ‘cause’ with the meaning Woodward has designed for ‘total cause’; I can say that I am using ‘cause’ in such a way that x is a cause of y if and only if...

¹⁵ Pinder (2021) argues that we should view conceptual engineers as engaged in changing speaker-meaning, chiefly because it presents a more plausible interpretation of the practice than what we get if we view conceptual engineers as trying (and failing) to bring about semantic change. Pinder’s arguments are different but compatible with the present defense of the option to change speaker-meaning.

I can say that I am using ‘marriage’ such that same-sex couples may be married. All the materials for communicating clearly what one means by an expression should already be in place if one is attempting to communicate what a conceptual engineer has argued that we should mean by that expression. If those materials are not in place, the design-stage of that piece of conceptual engineering has not been successful, but that is a separate issue.

Changing what speakers communicate by their use of an expression does not involve changing any semantic fact as such. Since this implementation option does not guarantee any semantic change, locally or globally, one may question whether it gets conceptual engineers what they want. Does not the success of a project in conceptual engineering require that one has changed a semantic fact at some level or another? Deutsch writes that

a conceptual engineer might convince some speakers to speaker-refer to something other than the semantic reference of ‘free action’ when using the term. But this seems like a rather trivial and easy thing to do. Surely it is not the sort of thing the exciting terminology of ‘conceptual engineering’ was designed to describe. (Deutsch 2020, 3941)

Deutsch leaves his claim about triviality unexplained. The claim cannot, however, provide grounds to conclude that it is not worthwhile to change speaker-meaning. The only sense in which it is trivial to change speaker-meaning is that there is no implementation problem inhibiting it as there is for the project of changing standing meaning. Moreover, I see no reason why ‘the exciting terminology of “conceptual engineering”’ should entail that implementing a meaning cannot be trivial, in *this* sense of ‘trivial’. To be sure, it would not follow that changes in speaker-meaning are inconsequential. It is question-begging to assume that conceptual engineers have to face an implementation problem to be engaged in a worthwhile project.

More importantly, there are independent reasons to value speaker-meaning as a locus of change. To see this, consider the converse situation of changing the semantic facts without changing what speakers communicate. The downstream effects of speech on belief and action depend upon what is communicated, not on what is semantically encoded by an utterance (in case what is semantically encoded is not what is communicated). This means that philosophers who want to promote social justice by engaging in conceptual engineering, such as Sally Haslanger (2012), would be very unhappy with this predicament. She could then succeed in effecting semantic change without doing one bit for the cause. By changing communication facts without changing semantic facts, on the other hand, she would be able to make relevant changes: changes that could constitute the promotion of social justice. The same goes for projects aiming for better inquiry, scientific or non-scientific. If Woodward could get scientists to communicate by ‘cause’ what he argues we should mean by that word, without changing any semantic fact about it, it would still result in scientists reaching conclusions that are logically licensed by the inferential role he has designed. Moreover, insofar as Woodward has succeeded in designing an exact and fruitful meaning for ‘cause’, beliefs instilled by communicating that exact and fruitful meaning would themselves be exact and fruitful. These are worthwhile changes. It is therefore worthwhile to change speaker-meaning.

The worth and feasibility of changing speaker-meaning would not count for much if you could only cause a severely limited amount of speakers adopt the meaning you prescribe. However, there are only familiar, practical limitations on how many people you can convince, and how often they will end up meaning by an expression what you argue they should mean by it. Therefore, this implementation option leaves conceptual engineers in a no worse predicament than those who engage in the familiar project of trying to change what people believe.

Objection: According to Cappelen, “The content of what’s called ‘speaker’s meaning’ is just as externalistically determined as linguistic meaning: we have no more control over that content than we have over what we say when we utter sentences in a public language.” (Cappelen 2018, 76).

Reply: This is an unattractive view. The existence of non-identity between an expression’s linguistic meaning and what a speaker means by it entails that the former cannot be fixed by the very same determinants and functions that fix the latter. But then, if it is true that speaker-meaning is just as externalistically determined as linguistic meaning, we need to tell two separate externalist stories, one for speaker-meaning and one for linguistic meaning, and the former story has to leave speaker-meaning beyond our control. The only remotely plausible way of constructing that story—that I can think of, anyway—is to recognize that what a speaker means is a content she entertains in thought, and then to see externalism about speaker-meaning as a special case of externalism about thought content.

This might seem like a reasonable approach. The problem, however, is how an externalism about thought content can deliver as strong results as is needed in this context. Suppose I want to go along with Tim and Martha and speak about all and only the stuff that share the sensory and practical properties possessed by what we call ‘water’ here on Earth, including stuff that is not composed by H_2O . I seem to be able to entertain thoughts about this stuff. I can simply think about it in the above terms. If that is the case, then these thoughts are available to me when I want to communicate something by means of a word, including the word ‘water’. As always, I just have to make my communicative intentions known. Surely, the externalist about thought content will urge that we cannot typify the thoughts in question as thoughts about water, since water-thoughts are essentially about H_2O -composed stuff. However, there is no need to typify them as water-thoughts. No matter how we typify them, they are still available to provide a content for what I want to communicate by means of an arbitrary expression (including ‘water’). For

these reasons, I am not optimistic on behalf of the claim that speaker-meaning is beyond our control in the relevant way.

6. Speak a Different Language

We have so far considered how conceptual engineers could implement their prescribed meanings in (or through the use of) existing languages. However, they could also implement their prescribed meanings by way of constructing new languages. In section 3, we found that the standing meaning of natural language expressions are largely beyond our control, such as ‘cause’ or ‘water’ in English. For the sake of argument, we can now assume that not even tiny languages constitute an exception. Perhaps because, contrary to my own belief, naturalness or past usage play a much stronger role than I have given them credit for. On this basis, one may worry that any language is such that the meanings of its expressions are beyond our control.

However, it is just not true that all languages are subject to externalistic determinants. Artificial languages comprise a significant exception. Consider programming languages. The rules that constitute a programming language alone determine the semantic values that syntactically valid strings attain. Those values need not conform to what experts believe (whatever that means), what would be the most ‘natural’ value, and they are not affected by any causal interaction between language users and the environment. Similarly, when we use the language of classical logic (or the language of an alternative logic), the rules governing that language alone determine the meaning of the symbols used. Programmers and logicians have defined the rules for such languages. Their acts of defining rules were in their control, insofar as anything is ever in our control. Hence, the metasemantic determinants for such languages are in our control. By consequence, so are the meanings of the expressions in such languages. It is in our power to construct languages. When we do, the meanings of the expressions in the target language seem to be in our firm control.

Therefore, even if externalist metasemantics inhibit conceptual engineers from changing the languages we already have in use, and even if externalist metasemantics inhibit them from getting speakers to communicate deviant meanings through their use of an existing language, conceptual engineers could still construct new languages. Indeed, if we individuate languages semantically—i.e. that language $L \neq$ language L^* if L and L^* encode different mappings from lexical items to semantic values—it is hard *not* to see conceptual engineers as proposing that we speak different languages than the languages that are currently in use.

It would be practically unfeasible to construct languages from scratch if this requires us to provide rules that would make it the case that an expression in the constructed language has an identical meaning to, say, the one we saw Woodward design for ‘cause’. That meaning was designed with the help of English expressions, and it would take an inconceivable amount of work to recreate the meanings of the latter expressions in a language built up from scratch (if it is at all possible). This may breed suspicion that it is not possible for conceptual engineers to construct the meanings they are interested in by way of constructing new languages.

However, conceptual engineers do not have to construct languages from scratch. They can instead propose that we speak a language L^* , that has the same morphology, phonetics, syntax and semantics as an existing natural language L , except for the imposition of a semantic rule and concomitant adjustments. In that way, it is practically feasible to construct L^* . Imagine Woodward propose that we speak English* rather than English, where English* inherits meaning for all its expressions from English, except for the meanings that ‘cause’ and its cognates have, which are instead defined by the rules he puts forward in his work, and cognate supplements. This is certainly something Woodward *could* have proposed. If people follow up on his proposal and start speaking English*, well, then he will have succeeded in getting people to use the meaning he prescribes. So long as the usual effects of speech obtain in the event of speaking English* (rather than

English), it would be a worthwhile accomplishment. I see no reason why those effects would not obtain. Getting people to speak a different language therefore seems like a viable option to grant conceptual engineers what they could reasonably desire. In the following, I respond to three potential objections to this implementation option.

First, it will do no good to object that we do not grasp what it is to speak English*. Our grasp of what it is to speak English* is at least as good as our grasp of what it is to speak English. Speaking English*, after all, just is speaking English except for the substitution of a meaning we have found it hard to pin down in precise terms, for a meaning that has been given by precise, bi-conditional statements.

Second, one may worry that speakers of English* would not be able to communicate with speakers of English. However, it is not in general true that we are incapable of communicating across languages. Consider Spanish and Portuguese, or Swedish and Norwegian. Admittedly, our case is different: It involves semantic variation, rather than variation in phonetics or spelling. Therefore, in order to communicate properly, the speaker has to let the hearer know of her semantic deviance. If a speaker of English* hides the fact that she is speaking a language in which ‘cause’ has a different meaning from the homonym in English, or fails to take other steps to ensure that the intended content is retrieved, she will not meet the demands for good communication. I offer no safeguards against bad communication. Clearly, however, there *are* straightforward ways of communicating well, such as saying what you mean. Conceptual engineering is, if anything, an excellent tool for getting clear on exactly what we mean, and hence to avoid verbal disputes and other miscommunications. In addition, conceptual engineers are concerned to minimize or eliminate properties of meaning that make us liable to communicate badly, such as vagueness or imprecision. Since the present implementation option requires being precise about

the semantic features of the language one is speaking, it does not undercut this virtue of conceptual engineering.

Third, one may worry that I am relying on an individuation criterion for languages that is too fine-grained. Would speakers of English* *really* be speaking a different language from English? Well, this obviously depends on how we individuate languages. On a coarse-grained individuation, use of English* will just be use of English.¹⁶ But we can still single out English* if need be. Compare with Lewis on the languages spoken in a population *P*:

Not one but many—perhaps infinitely many—languages are used by *P*. [...]

That is so, but it is no problem. Why not say that any rich fragment of a language used by *P* is itself a used language? Indeed, we will need to say such things when *P* is linguistically inhomogeneous. (D. Lewis [1975] 1983, 186)

If it is in our interest to be precise, we can identify different fine-grained languages spoken among speakers of English. There is the language of politicians, the language of theoretical physics, the language of jihadists, the language of ‘counter-jihadists’, the language of drug dealers etc. In practice, we rarely need to descend to a very fine-grained level of individuation, but that is a poor reason to prohibit fine-grained language individuation in our theorizing. How we individuate languages is and ought to be driven by our (passing) interests. Sometimes, the small differences matter. This is not unique to languages. We employ different individuation criteria for many, if not most, other objects, abstract and concrete. For example, different passing interests guide whether we count book *a* and book *b* as the same book: Same copy? Same string of marks? Same informational content? Either we individuate languages finely, and allow the possibility described

¹⁶ C.f. Davidson: ‘There is much to be said for restricting the word ‘language’ to systems of signs that are or have been in actual use: uninterpreted formal systems are not languages through lack of meaning, while interpreted formal systems are best seen as extensions or fragments of the natural languages from which they borrow life.’ (Davidson [1984] 2001, 71 emphasis added).

in this section, or we individuate them coarsely. If we do individuate them coarsely, then, on pain of putting the cart before the horse, we have to adjust our metasemantic theses for the languages so individuated. Our choice of an individuation criterion for languages cannot carry the metaphysical weight that this worry asks of it.

It is important not to confuse the preceding discussion with questions about the primacy of idiolects or momentary rules of interpretation on the one hand, and large, ordinary languages on the other (e.g. as discussed by Davidson 1986; Dummett 1986). It is a matter of contention whether English is to be explained as a complex amalgamation of idiolects, or, conversely, whether there must be a larger social practice in place before there can be any linguistic meaning. The preceding discussion does not avail itself of the independent existence of idiolects: On this proposal, the conceptual engineer is using a large, ordinary language to construct what we might call an idiolect or a constructed language. The large, ordinary language must therefore be in place to develop the constructed language. In the paradigm case, the conceptual engineer is producing a sentence—such as Woodward’s (TC’)—that is false or indeterminate in the large, social language, but true in the constructed language. The conceptual engineer, on this construal, is proposing that we speak a language in which the sentence is true.

Given that we can choose which language to speak, this is a feasible way of implementing a meaning. Conceptual engineers could succeed in implementing a meaning by convincing people to speak the language they prescribe.

7. Conclusion

It may be largely unfeasible to change standing meaning in the natural languages we ordinarily recognize (e.g. English and Spanish). However, we have found reason to reject the claim that conceptual engineers must do so or else accomplish nothing. In the last three sections, we have considered other ways in which conceptual

engineers could succeed in implementing the products of their work; other options for getting people to use the meanings they prescribe. The options are not intended to be mutually exclusive, and it is not my belief that conceptual engineers have to commit to the one or the other. Rather, the options I have identified are tailored to address a variety of theoretical views on language and thought, with the aim of showing that conceptual engineers should not worry about the implementation problem. Without additional, controversial assumptions, none of the implementation options are rendered unfeasible by metasemantic externalism. The feasibility of the modulation option could be defeated if we conjoin externalism with an anti-contextualist view, but one could be externalist without being anti-contextualist. As for the option to change speaker-meanings, we found no good reason to deny its feasibility. The only route would be to develop an argument based on externalism about thought content, but we found reason to doubt the cogency of that approach. And then, even if conceptual engineers could not change what speakers mean by their use of an expression in an existing language, they could still construct new languages and convince people to speak them. Rejecting the latter option requires adopting a dismissive view about the possibility of fine-grained languages or of constructing languages, but we have not seen any good reasons for taking such views.

I make no claim to have reviewed every consideration that bears upon the feasibility and worth of the implementation options I have discussed. Rather, I hope to have shown that a whole lot more needs to be argued, including arguments for what seem to be unattractive views, if one wants to maintain that conceptual engineers are engaged in a futile project. Pending those arguments, we can conclude that conceptual engineers face no more of an implementation problem than thinkers who engage in the familiar project of trying to change what people believe.

Chapter 3

The Good, the Bad and the Insignificant – Assessing Concept Functions for Conceptual Engineering

1. Introduction

To engage in conceptual engineering is to design and implement concepts that will, in the good case, serve us well in some capacity. Commonly, the aim is not just to add new concepts to our conceptual repertoire, but to replace or revise the concepts we already possess. A conceptual engineer may e.g. seek to revise our concepts of gender with the goal of promoting social justice (Haslanger 2000), propose that we replace our actual concept of truth with concepts of truth that avoid the alethic paradoxes (Scharp 2013), recommend concepts of causation that improve our explanations of events and states of affairs (Woodward 2003), or propose a concept of knowledge that serves our need for having a concept of knowledge (Craig 1990; Hannon 2019). These projects take a normative approach to our conceptual repertoire, to be contrasted with the descriptive approach characteristic of traditional conceptual analysis. While a conceptual analysis should accurately capture a concept in use, exercises in conceptual engineering are not so constrained. Now, lifting the requirement to capture a concept in use does not mean that anything goes. Exercises in conceptual engineering are simply subject to different success conditions. The relevant success conditions should befit the aim of improving our conceptual repertoire. As a rough approximation, an exercise in conceptual engineering is successful only if

(*Goodness*) the engineered concept is good by standards that apply to the concept in question; e.g. it is useful or benign.

(*Limits*) it does not involve revising a concept beyond limits, *viz.*, so that we merely “change the subject” or undermine legitimate purposes for having the concept that is being revised.¹

These conditions are not fleshed out, and they are not intended to be. In this paper, I examine a view of how we should develop them. According to what I call *the functional approach* to conceptual engineering, we should understand success in conceptual engineering in terms of the function(s) that a concept performs. More precisely, the functional approach understands the limits of revision in terms of the function(s) currently performed by the concept to be revised, and what makes some concept good in terms of which function(s) the engineered concept is to perform.

The approach has many advocates. Recently Weinberg (2006), Brigandt (2010), Plunkett and Sundell (2013), Plunkett (2015), Prinzing (2018), Nado (2021a; 2021b), Simion and Kelp (2020), and Thomasson (2020a) have all advocated a functional approach to conceptual engineering. Some of these speak of the ‘role’, ‘purpose’ or ‘aim’ a concept serves, but we may interpret all of them as latching on to roughly the same idea; one that is described well with the word ‘function’. Looking further back in time, we may identify a functional approach to conceptual engineering in Strawson’s (1963) objection to Carnap—that using his method of explication to solve philosophical problems changes the subject rather than solving the problem—and in the latter’s response (Carnap 1963b). As Nado (2021b) observes, both Strawson’s objection and Carnap’s response are drenched in functional terms:

¹ *Goodness* and *Limits* are not supposed to be jointly sufficient for success. Arguably, unless the engineered concept is implemented somehow—e.g. becomes the semantic content of a word in a language, or used in some other way—the project is not yet fully successful. I defend an optimistic, pluralist view about implementation in (Jorem 2021). Meanwhile, I hold the working assumption that *Goodness* and *Limits* are necessary conditions on success in conceptual engineering. It is possible to define ‘success in conceptual engineering’ in alternative ways, e.g. by equating it with *successful implementation*, but that would not capture the interest with which this paper is written.

The kinds of concept we employ are not independent from the kinds of purpose for which we employ them; even though some concepts can fulfil more than one kind of purpose. (Strawson 1963, 506)

The explicatum is intended to take the place of the explicandum, and that means, of course, that it is to be used for the same purpose as the explicandum. (Carnap 1963b, 936)

In addition to philosophers thinking about conceptual engineering, several philosophers *engaged in* conceptual engineering have appealed to function or purpose, including Craig (1990) and Hannon (2019) on knowledge, Haslanger (2000) on gender and race, and Woodward (2003) on causation.

However, it is not immediately clear what it is for concepts to perform functions. Proponents of the functional approach need to clarify their ‘function’-talk. In doing so, they face several plausible demands. First, there need to exist functions in the sense being proposed. Second, the functions so construed need to help us understand the success conditions on conceptual engineering (i.e. *Goodness* and *Limits*). Third, the candidate functions should be functions we *ought to* preserve when we revise concepts. Reviewing extant proposals, I argue that there is probably no sense of ‘function’ that can meet these demands, so as to give us a fairly complete, non-question-begging account of conceptual goodness and the limits of revision. Whereas Cappelen (2018) argues on similar grounds that the functional approach fails, I aim to show that this conclusion is too hasty. Firstly, there are clear, non-empty notions of a concept function. If nothing else, we may understand a concept function simply as whatever we happen to use the concept for on some occasion.² However, not any given function performed by a concept

² The idea that we *use* concepts may provoke questions about how I conceive of concepts; perhaps driven by a worry that there is no good understanding of what concepts are, such that they are the sort of things we can use or engineer (Cappelen 2018, 141, 200; Isaac 2020; Koch 2020). I assume, minimally, that concepts are the sort of things words can mean. Perhaps the set of concepts is broader than the set of linguistic

should constrain revisions of the concept. Not all functions are good and methodologically significant. Some are harmful and some are insignificant. Even so, there is comparatively weak claim the proponent of function can make, which still has methodological bite: For any given concept there are functions this concept performs that help determine what a successful revision must be like. What we cannot tell in advance, I argue, but must determine case-by-case, are exactly which functions we need to do conceptual engineering in the service of.

In section 2, I show that the functional approach offers an appealing construal of *Goodness* and *Limits*. In section 3, I raise a problem for the functional approach that I call *the selection problem*. According to the selection problem, there are functions that we should not engineer in the service of. Therefore, we have to make a selection among functions, but it is not clear how we can specify the range of functions conceptual engineers need to pay heed to. In section 4, I review whether extent ways of understanding function can deal with the selection problem without falling prey to other problems. In section 5, I argue that the advocate of function should accept that the selection problem cannot be dealt with in a general manner. I elucidate the consequences this has for the functional approach.

2. Why appeal to function?

We use concepts to carve up the world, to explain and predict events, to evaluate actions and states of affairs, to prescribe thought and action, etc. The guiding thought behind the functional approach, as I understand it, is that what we *use* a given concept for matters for how we should assess revisions of that concept. On the side of *Goodness*, what a concept is to be used for matters for what kind of standards we should use to evaluate that concept. On the side of *Limits*, what we

meanings, but it will do no damage to my main line of argument if ‘concept’ is understood as the meaning of a substantive linguistic expression, and ‘using a concept’ as using some expression with a given meaning (or given speaker-meaning).

currently use a concept for matters for how we should circumscribe the limits of revision of that concept. In the following, I motivate and clarify the functional approach with respect to both of these claims; first *Goodness*, then *Limits*.

On a functional understanding of *Goodness*, what it is for a concept to be good is for it to perform its functions well. The functions a concept is to perform thereby determine the standards of evaluation we should apply when assessing the concept. As Thomasson writes,

no detailed evaluation can be made without an assessment of the functions that are to be served. Once a purpose (or multiple purposes) is/are identified, we can go on to use that in engaging in conceptual engineering—determining what sorts of rules or constraints that would best (or better) enable it to fulfill its function(s), going forward. (Thomasson 2020a, 449)

We can motivate this view by a comparative example. It is sometimes claimed that concepts ought to “carve nature at its joints” (e.g. Sider 2011). *Joint-carving* is a plausible candidate for a conceptual virtue or norm, i.e. as a standard of evaluation.³ Take a concept that plays an important role in natural science, such as the concept of an electron. Presumably, there is a joint in nature that our concept of an electron carves out. This is, I assume, part of what makes it a good concept. However, it is dubious to evaluate all concepts according to whether they carve out a joint in nature. For instance, our concept of a table is a useful concept, despite failing badly to carve a joint in nature (tables can be made out of any number of materials, have any number of different shapes and sizes etc.). The functional approach can explain why its failure to carve a joint in nature is not a flaw: Our concept of a table does not perform any functions that would justify us in evaluating

³ I set aside the question of whether we should understand joint-carving in a deflationary or metaphysically heavyweight way (Thomasson 2020a). On a heavyweight reading, facts about the joints of nature are deep facts about the metaphysical structure of the world. On a deflationary reading, talk of “the joints of nature” is something like a shorthand for talk about what it turns out to be useful to identify in the process of forming scientific explanations and predictions.

it according to this standard. By contrast, many scientific concepts perform explanatory and predictive functions that do justify us in evaluating them according to whether they carve out a joint in nature.

There are a number of other conceptual virtues that are important in science, but less important in non-scientific discourse. Consider being a quantitative concept vs. being a categorical concept, or being exact vs. being vague. Exact, quantitative concepts drive scientific progress. In everyday discourse, by contrast, we get along fine with categorical and vague concepts, such as our concepts of being bald or of being wet; or even our concept of democracy. We have no acute need to make these concepts more exact, or to replace them with quantitative alternatives. However, it seems arbitrary to set apart conceptual virtues that apply to scientific discourse from conceptual virtues that apply to non-scientific discourse unless there is *something* about scientific discourse that warrants our differential treatment. Absent better explanations, we are justified in thinking of the relevant *something* as what we use scientific concepts for, i.e. generalization, explanation, prediction, etc. In turn, it makes sense to classify what we *use* scientific concepts *for* as functions performed by scientific concepts. Thereby, we find reason to endorse the following.

(*Functional Goodness*) The functions a concept C is to perform determine the standards of goodness for C.

What, then, of *Limits*? The need to address limits on conceptual revision came squarely in view with Strawson's objection to Carnap: Unless we preserve something from the concept to be replaced or revised, we run the risk of changing the subject rather than improving what we had. This bears upon a different success condition from *Goodness*, because a concept can be good on independent merits and yet fail to be acceptable as a revised state of, or replacement to, a given concept in use. To illustrate, our concept of belief is independently useful, but we would achieve nothing if we revised our concept of *desire* so as to attain the same content as our concept of belief. The functional approach to conceptual engineering can

explain this by appealing to the fact that our concept of desire performs a different function from our concept of belief.

As with *Goodness*, the functional approach gains abductive support from its construal of *Limits*: It offers a good explanation of why we reject revisionary proposals that, intuitively, involve too much change and accept—or hesitate to reject—proposals that do not seem to involve too much change. To see this, compare two revisionary proposals. First, consider the silly proposal to revise our concept of knowledge so as to apply to all states of belief; not just those that are true and justified. Next, consider the proposal to revise our concept of knowledge so as to apply to all and only those states of belief that are true and justified (c.f. Weatherson 2003). Let ‘KNOWLEDGE^{JTB}’ name the concept that would result from the latter revision. Philosophers since Gettier (1963) have been painfully aware that our actual concept of knowledge is non-identical to KNOWLEDGE^{JTB}, but the existence of Gettier cases merely shows the non-identity of the two concepts, not that KNOWLEDGE^{JTB} would be too drastic of a departure from our current concept of knowledge. For that matter, Gettier cases are few and far between. Extending our concept of knowledge to *all* states of belief is problematic in a way or to a degree that the KNOWLEDGE^{JTB} revision is not. The functional approach to conceptual engineering offers a promising explanation of this. If we suppose—as is plausible—that we use our current concept of knowledge to extend epistemic praise, we can explain our diverging judgments towards these two proposals. If we were to revise our concept of knowledge so as to apply to all states of belief, we would undermine the function of epistemic praise. If, on the contrary, we revised our concept of knowledge so as to apply to all and only states of belief that are true and justified, we could still use the concept for this purpose.

Zooming out, it is plausible to think that the theoretical and practical value concepts have for us reside in the use we have for them. Unless there is use for a given concept—to carry out some set of tasks or perform *some kind* of function—there would be no reason to target that concept, as opposed to any other concept,

as the object of revision. But if the value of a concept resides in the functions it performs for us, then it is plausible that the capacity to perform those functions is what we should preserve when we engage in revision. We are therefore justified in drawing the limits of revision accordingly:

(*Functional Limits*) A revision of concept C is acceptable only if it preserves the capacity to perform functions already performed by C.

As will become clear in the next section, there are problems with *Functional Limits*, in particular having to do with how to understand ‘functions’ and exactly which ones we ought to preserve. For now, let us observe that the claim is well-motivated.

Notably, *Functional Limits* makes no mention of “subjects” or “topics” (c.f. Cappelen 2018, chaps. 9–10); nor does it require that a concept preserves numerical identity through a revision process (c.f. Prinzing 2018). For reasons spelled out by Nado (2021a; MS), Koch (2021b) and others, I think these ways of drawing the limits of revision are dead ends. Unless we specify an individuation convention for subjects or concepts in a way that piggybacks on facts or properties we care about for independent reasons, there is no good reason why conceptual engineers should not sometimes change the subject, or sometimes change which (numerically individuated) concept we use. I have therefore specified *Functional Limits* without requiring sameness of topic or concept.

We seem to have good reason to accept *Functional Goodness* and *Functional Limits*. As these two commitments comprise the functional approach to conceptual engineering, we have *prima facie* reason to endorse the functional approach.⁴ However, our motivation has rested on an unsophisticated understanding of what

⁴ Note that one could endorse a weaker version of the functional approach than one comprised by *Functional Goodness* and *Functional Limits*. One could e.g. think of function-preservation as a point in favor of a revision, to be weighted against other good-making features, rather than as a necessary condition. Pending arguments for taking a weaker approach, I find it more interesting to explore the prospects of the comparatively strong version encapsulated by *Functional Goodness* and *Functional Limits*.

it is for a concept to perform a function. In the next section, we attend to a problem that arises on this unsophisticated understanding.

3. The selection problem

On the understanding of function we used to motivate the functional approach, the function of a concept is what it is used for. In general, we use concepts for all sorts of tasks: To categorize, explain, predict, prescribe, evaluate, command etc. We may identify more specific tasks when we look at concrete examples. For example, Craig hypothesizes that we use our concept of knowledge to “flag approved sources of information” (Craig 1990, 11). Woodward suggests that we use our concept of causation to identify “relationships that are potentially exploitable for purposes of manipulation and control” (Woodward 2003, 25). These seem like plausible ascriptions of function. Moreover, it seems reasonable to evaluate revisionary proposals according to how well the engineered concept (of knowledge or causation) can perform these functions.

Importantly, however, for all we have assumed about concept functions so far, we have no reason to think that concepts perform exactly one function each. We use our concept of knowledge for a number of different tasks. Some of these tasks seem less well fit to determine what our concept of knowledge should be like. We use our concept of knowledge to discredit interlocutors (e.g. “And what do you know about that?”), to position ourselves (“I don’t *believe* that there is anthropogenic climate change, I *know* there is.”), to acknowledge mistakes (“I should have known better.”) and so on (c.f. Cappelen 2018, 182 on uses of ‘women’). Now, it may sound awkward to think that our concept of knowledge *has the function* of, e.g., discrediting interlocutors just because we sometimes use the concept for this purpose. I address this concern in the next section, where we explore more demanding notions of function. For now, we need only observe that it sounds dubious to ask that revisions of our concept of knowledge should aid us in discrediting interlocutors.

Recall the two claims that comprise the functional approach as I have construed it here:

(*Functional Goodness*) The functions a concept C is to perform determines the standards of goodness for C.

(*Functional Limits*) A revision of concept C is acceptable only if it preserves the capacity to perform functions already performed by C.

These schematic statements make indefinite reference to the functions performed by a concept. They do not tell us exactly *which* functions of C that are supposed to determine the goodness of C or its limits of revision. And then, since any given concept performs a number of different functions, the functional approach lapses into obscurity if it does not clarify whether *all* the functions of a concept are supposed to count for *Goodness* and *Limits*, only some, or only one. In the event that only one function or only some functions count, we should want to know which one(s). As noted, one could supply a different, more demanding notion of function to replace the unsophisticated notion of function we have used so far. Then we could discount would-be functions such as *the use of our concept of knowledge to discredit interlocutors* as not being functions in the operative sense of ‘function’. But for now, I ask for patience with my equation of functions with uses as I spell out the present problem.

In the absence of a more sophisticated notion of function, we can still ask whether there is really a problem here. Is it really so bad to hold that *all* the functions of a concept are to determine its goodness and limits of revision? The answer is “yes”. As I show in the following, there are both *harmful* functions and functions that are otherwise *insignificant* (for the purposes of *Goodness* and *Limits*). The fact that our concepts perform harmful and insignificant functions means that we have to make a selection among the functions a given concept performs. Not all of them count toward *Goodness* and *Limits*. I call this *the selection problem*.

Some functions are methodologically insignificant. Take our concept of planethood as an example. Not too long ago, when we still considered Pluto a planet, our concept of planethood applied to nine objects in our galaxy. We used the concept to think and speak about a property exhibited by nine objects in our galaxy. Therefore, on our unsophisticated understanding of function, our concept of planethood had the function of categorizing nine objects in our galaxy. In the wake of the International Astronomical Union’s decision to redefine ‘planet’ in 2006, our concept changed.⁵ According to the new definition, “A planet is a celestial body that [...] has cleared the neighborhood around its orbit.”⁶ On this criterion, Pluto no longer counts as a planet. While the revision preserved several functions performed by our pre-2006 concept of planethood (such as the function to categorize bodies of mass orbiting a star), it failed to preserve the function to categorize nine objects in our galaxy. And yet the revision is acceptable. It follows that we would be wrong to require the preservation of the function considered. By consequence, we cannot require that revisions preserve all functions performed by the concept to be revised.

A similar example is due to the change from a pre-Linnaean concept of fish to the Linnaean concept of fish. The former concept functioned to classify whales as fish, but with Linnaeus’ taxonomy we transitioned to a concept of fish that does not apply to whales and that, by consequence, does not function to categorize

⁵ Is this a genuine example of conceptual change, as opposed to a change in belief?

Haslanger (2012, 398) sometimes describes her project as revealing what we meant all along by terms like ‘woman’ and ‘black’. Externalist views, especially temporal externalist views, may lead to such verdicts. See (Ball 2020) for a defense applied to conceptual engineering. Following Ball, one might think that our concept of planethood disapplied to Pluto all along. However, even if we accept this, we may still evaluate the example as a hypothetical case of conceptual change. Our interest is with what it takes for a revision to be acceptable, and this interest is equally well served by considering hypothetical cases.

⁶ C.f. Resolution B5 – Definition of a Planet in the Solar system: https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf (accessed October 2020).

whales. Yet we would consider this an acceptable revision. It turns out that some of the tasks we use concepts for are less important, methodologically speaking, than others. To classify whales as of the same kind of animal as cod and trout is not an important task. Neither is the task of denoting a property exhibited by nine objects in our galaxy. I project that we can find many more examples of insignificant functions. In virtually any case of conceptual change we can probably specify something the old concept was used for that the new concept is no longer used for.

While we should not be obliged to preserve insignificant functions, we actively want to avoid preserving harmful functions. Consider a concept of rape that does not apply to intra-marital sexual transgressions. Let us call it ‘the extra-marital concept of rape’. Unfortunately, we do not have to go far back in time to find evidence that the extra-marital concept of rape was our actual concept of rape. Edward Schiappa quotes Edwin Silberstang on US legal practice from only a few decades ago: “A husband cannot be guilty of an actual rape, or an assault with intent to rape his wife even if he has, or attempts to have, sexual intercourse with her forcibly and against her will.” (Silberstang 1972, 775; quoted in Schiappa 2003, 54). In the legal context, use of an extra-marital concept of rape has helped protect wrong-doers from liability to punishment. Indeed, the extra-marital rape *performed the function* of protecting wrong-doers from liability to punishment.⁷ The extra-marital concept of rape thereby performed a harmful function. What I take to be our current concept of rape is marriage neutral. It does not perform the function of protecting wrong-doers from liability to punishment. Despite not performing this function, it remains an acceptable replacement to the extra-marital concept of

⁷ If the reader finds it dubious that the extra-marital concept of rape performed this function, she may instead consider instead the function to *categorize a kind of violence that occurs exclusively outside of marriage*. Given the empirical fact that we lacked (in law and in general thought) other concepts to identify intra-marital rape, the concept’s performing this latter function has had the harmful *effect* of masking an extensive kind of wrongdoing, and of protecting wrong-doers.

rape. Indeed, the very point of broadening our concept of rape would be to have it cease performing the function considered.

The extra-marital concept of rape is not a unique case. Concepts of race may function to consolidate social injustice in virtue of underwriting dubious explanations of socio-economic differences. A ‘both sides’ concept of objective journalism may function to distort our ability to identify and disseminate facts. The former is a morally harmful function and the latter is an epistemically harmful function. We do not want concepts to perform these functions. We should not require their preservation or judge concepts to be good according to how well they perform them.

Conceptual revisions cannot be normatively constrained by insignificant functions or harmful functions. We need a way to filter them out, preferably in a principled and non-question-begging manner. If we achieve this, we will have solved the selection problem.

4. Can we solve the selection problem by refining our notion of function?

It is fair to object to some of the function ascriptions I made in order to raise the selection problem. I have flouted the principle that “All functions are uses; but not all uses are functions.” (Prinz 2018, 868) To be sure, many of the function ascriptions I made turn on what I called an unsophisticated notion of function, and would not count as functions on various, more demanding accounts. In this section I review strategies for solving the selection problem that are based on existing ways of operationalizing what it is for a concept to perform a function.

Some advocates of the functional approach hint at a specific understanding of function, but do not make it fully explicit. E.g. Haslanger (2000, 35) speaks of the “central functions” of a term, but does not spell out what it takes to be a central function. Brigandt (2010) speaks of “the epistemic goals” of scientific concepts, and

can be interpreted as meaning the epistemic function *intended* to be served by those who deploy or have developed the concepts in question—but whether he would accept this interpretation is not clear to me. Thomasson does propose explicit accounts of function—Millikan’s (1984) account of proper function and Cummins’ (1975) account of system function—but does not commit to using either to develop the functional approach (Thomasson 2020a, 444–46). The idea to use the notion of a system function is echoed by Haslanger in a recent paper: “Functions are attributed to parts of a system when they contribute to the ongoing workings of the system.” (Haslanger 2020, 250, fn.). Simon and Kelp (2020), meanwhile, propose that the conceptual engineer is to construct a concept with a designed function—some task the concept is designed to serve—with the goal that the designed function becomes that which explains why the concept sustains and proliferates. If this happens, the designed function becomes an etiological function of the concept, which is, in essence, the same as what Millikan calls a proper function. Nado, on the other hand, argues that “neutrality on the nature of function is here not only permissible, but appropriate.” (Nado 2021a). Riggs (2021) advocates a similar position. Arguing that the notion of a concept’s function is being asked to play too many different explanatory roles, Riggs proposes that we do not need a sophisticated notion. He concludes,

Talk of the function, point, or aim of a concept isn’t a way of referring to a useful piece of the theoretical machinery, but instead is a way of directing our attention to what *matters* in a given inquiry, which will change depending on what is at issue. (Riggs 2021)

I agree with Nado and Riggs that we do not need the more demanding notions of function. One reason for this is that the more demanding notions do not, as I show in this section, help us deal with the selection problem.⁸

⁸ Thus, the present argument is different from the argument Riggs (2021) gives. Riggs considers three categories of explanatory roles that the notion of a concept’s function is

4.1 Contextually stable functions

As a first stab at restricting the operative sense of ‘function’, one could narrow it down to *contextually stable functions* (c.f. Cappelen 2018, 182). That is, we could restrict which functions we count as methodologically significant to those functions that are performed across all its contexts of use. By doing so, we avoid several problematic function candidates. For instance, although we do use our concept of knowledge to discredit interlocutors and to acknowledge mistakes, it is not the case that any time we use our concept of knowledge it performs these functions. By narrowing the operative notion of function down to contextually stable functions, then, the functional approach avoids misattributing significance to what are in fact insignificant functions.

Following Cappelen (2018), however, it is not obvious that there are any good candidates for contextually stable functions. Are there specifiable tasks that we always carry when we use a given concept? If there are not, the wanted functions do not exist, and this strategy fails. If, on the other hand, there are contextually stable functions, the strategy does not thereby succeed. Cappelen considers what he calls a trivial version of the functional approach: “According to this view the only *universal*, i.e., stable, function of a concept ‘C’ is to denote Cs. The function of the concept ‘tiger’ is to denote tigers. The function of the concept ‘salad’ is to denote salads, the function of the concept ‘woman’ is to denote women, and so on.” (Cappelen 2018, 182). Although these denotational functions might count as

supposed to play, including the normative role of fleshing out what I call *Goodness* and *Limits*. He argues that no notion of function can play all the explanatory roles, and that it follows that we ought to deflate the notion of function to something contextually variable, such as what we care about in a particular context. The present argument concerns only the normative role of ‘function’-talk about conceptual engineering. I show that sophisticated notions of function do not help us solve the selection problem, and this lends some support, as I explain later, to the conclusion Riggs draws, i.e. that we should deflate ‘function’ talk.

contextually stable functions,⁹ they will not help us understand *Goodness* and *Limits*. Requiring the preservation of a denotational function would only beg the question of what it takes for a revised concept ‘C*’ to denote Cs. If we ask for the meaning of the word ‘aspersion’ and get told that it means *aspersion*, we learn nothing new about the meaning of the word. In a similar vein, it is not helpful to get told that revisions of our concept of knowledge should still denote *knowledge*. On one interpretation of that demand, the revised concept should be qualitatively identical to our current concept of knowledge. But this is unacceptable, as it would preclude revision across the board. If the request is rather that the revised concept still denotes some phenomenon or state worthy of the label ‘knowledge’ we are not much wiser, since if we knew the range of phenomena worthy of the label ‘knowledge’, we would already know the limits of revision of our concept of knowledge. Denotational functions do not help us understand *Limits*.

The kinds of functions *practitioners* in conceptual engineering appeal to are not perfectly stable. Consider the functions Craig and Woodward have appealed to. Craig engineers a concept of knowledge to serve the purpose of *flagging approved sources of information* (Craig 1990, 11). This is not a contextually stable function. When we speak about knowledge in general terms—e.g. “Knowledge has great value.”—we are not using our concept of knowledge to flag an approved source of information. Woodward, meanwhile, engineers a concept of causation to serve the purpose of *identifying relationships that are potentially exploitable for purposes of manipulation and control* (Woodward 2003, 25). Again, this is not a perfectly stable function. When we use our concept of causation to query the hypothesis that impact from an asteroid caused the mass extinction of dinosaurs, we are arguably not trying to identify a relationship that we could potentially exploit (c.f. Woodward 2003, 11). Any given concept can be used in an indefinite range of contexts. Chances are,

⁹ Note, however, that there are good reasons to resist the idea that every concept or piece of language serves a *descriptive* or *representational* function. See e.g. Brandom (1994), Price (2011) and Thomasson (2020b).

there are some contexts in which a given function candidate is not performed. If there is some function that is performed across all contexts of use—as witness highly general functions such as the function of *parsing objects into categories*—chances are it will not set appropriate limits of revision or standards of goodness.

I believe both Craig and Woodward identify important functions served by our concepts of knowledge and causation. That is to say, if a revision of either concept resulted in a concept that could not perform the respective function, this would give us reason against endorsing that revision. However, the importance of the respective functions does not entail that every kind of use of our concepts of knowledge and causation, silly and serious, are in their service.¹⁰ Recall that we are asking not only for the existence contextually stable functions, but for them to be such that we get a better grasp of *Goodness* and *Limits*, and for the set of contextually stable functions to be such that we ought to preserve them when we engage in revision. We better look to other notions of function to solve the selection problem.

4.2 Designed functions

Most concepts are occasionally employed for questionable purposes. As noted, our concept of knowledge could be employed for discrediting an interlocutor. However, a fork could be used as a murder weapon, but we would not say that the function of a fork is to kill people. Forks are created as tools for eating. Thus, a fork could function as a murder weapon on a particular occasion of use, but that does not mean that *the function of the fork* is to kill people. As Wright observes, “The function of a telephone is effecting rapid, convenient communication, but there are many other things telephones do: take up space on my desk, disturb me at night,

¹⁰ The respective functions could be important in an additional sense: They could help explain the proliferation of our concepts of knowledge and causation. Playing this role is again compatible with the function not being served on every occasion of use. What Millikan calls “the stabilizing and standardizing direct proper function of a language device” is not a contextually invariant function, but still explains proliferation and sustained use of the language device (Millikan 1984, 31–32).

absorb and reflect light, and so forth.” (Wright 1973, 141). In these examples, *the function of X* may be understood as something like the purpose for which X was created. Following this line, we could amend the operative notion of function to *the purpose(s) for which a given concept were created*. Let us call these functions ‘designed functions’. Prinzling flirts with a notion of designed function when he writes that “the function of a concept is what it was designed for” and illustrates with the functions that can openers and hammers were designed for (Prinzling 2018, 869).¹¹

Intuitively, narrowing down the operative notion of function to designed functions would preclude the function of protecting wrong-doers from liability to punishment from counting towards *Goodness* and *Limits*, since our previous concept of rape was arguably not constructed for this purpose. It would intuitively preclude the function to denote a property exhibited by nine objects in our galaxy, since we would probably have had our former concept of planethood even if Pluto did not exist. These are desirable consequences. However, there are major problems facing this strategy. There are reasons to dispute, first, the existence or prevalence of designed functions of concepts, second, our epistemic access to them, and third, their adequacy for defining norms on conceptual engineering.

The notion of a designed function is an intentional notion. It denotes the intentions that some individual(s) had when they designed a device. However, the devices in question are *concepts* and it is dubious that many, concepts have been designed intentionally, let alone all concepts. Consider our concepts of knowledge and causation: Who designed these concepts and what were their intentions? I find it more plausible to think that these concepts arose out of interactions between individuals, and between individuals and the environment, guided by practical

¹¹ Prinzling qualifies: “‘Design’, of course, need not mean ‘conscious design’. Darwin’s great lesson, after all, was that the presence of a design does not entail the existence of a designer.” (Prinzling 2018, 869). This suggests a notion of function that applies indiscriminately to products of evolution and conscious design. Millikan’s notion of proper function is one such; I discuss the strategy of using it in the next subsection.

needs for survival, coordination, welfare, etc, but without the metaconceptual awareness of someone who *designs* a concept. Some concepts might have designed functions, but we want our account of success in conceptual engineering to cover more than a narrow range.

Even if a concept of interest were designed intentionally, we would not thereby be in a position to divine exactly which purpose the concept was designed to serve. For instance, our concepts of knowledge and causation have a long history, and even assuming that they were intentionally designed, we do not know who developed them, much less what their intentions were. This would leave us unable to assess revisions of our concepts of knowledge and causation, since we would not have access to the proclaimed determinants of their *Limits*. Finally, concepts may be designed with malicious intent, to perform what are in fact harmful functions. If a concept is designed to serve a harmful function, revisions of that concept should not be functionally continuous. The notion of a designed function thus fails several of our desiderata.¹²

4.3 Proper functions

In answering an objection from Cappelen (2018) about relying on an intentional notion, Thomasson (2020a, 444) appeals to Millikan's (1984) notion of a proper function. The notion of a proper function does not require that there be a conscious purpose behind the candidate function. Thus, we can use it to make sense of our ascriptions of function to biological items, e.g. hearts, kidneys and instinctive behaviors. Millikan herself uses the notion of a proper function to develop a general theory about language and content; it is only natural to consider whether the

¹² As noted, Simion and Kelp (2020) use the notion of a designed function to propose that conceptual engineers should design functions for new concepts to perform. They do not, however, commit the mistake of thinking that this will help explain the limits of revision. On their proposal, conceptual engineers should be more concerned with innovating concepts than with repairing defects in existing concepts.

functional approach to conceptual engineering can be articulated with her notion of a proper function. Although the ensuing argument focuses on Millikan's account, it also applies to other etiological accounts of function, such as Wright's (1973).¹³

On Millikan's account, proper functions crop up in virtue of copying effects in social and natural environments. A copying effect ("reproduction") occurs when there is a causal relation between two items, such that one item attains one or more features of the other. This could be the effect that occurs when the genes of a parent are copied in the genes of its offspring, the mass production of a commercial item, or some individual's copying another individual's linguistic behavior. Such copying effects give rise to what Millikan calls "reproductively established families": Sets of items that share features in virtue of a copying effect. Proper functions are defined on this basis. Roughly, if a copied feature contributes to the survival or proliferation of members of the reproductively established family, then what the copied feature does for members of the family will count as a proper function for those members.¹⁴ On Millikan's view, then, we may view our (token) concepts as members of reproductively established families. Over the course of history, certain (type) concepts have enjoyed proliferation and sustained use, e.g. our concepts of knowledge and causation. Other concepts have been discarded, e.g. the concept of phlogiston. Whether a concept "survives" or not depends on what it does for us, or what it lets us do. The proper function of a concept C, then, is whatever we (and our ancestors) were able to do with C that explains how it enjoys sustained use.

Using the notion of a proper function as our operative notion lets us discount several abhorrent function candidates from misleading our efforts in function-based conceptual engineering. Consider the pre-2006 concept of planethood. The

¹³ There are differences between Millikan's and Wright's accounts of function that are significant for other purposes, c.f. (Millikan 1993 ch. 1).

¹⁴ Millikan's account is more complex than what I can do justice to here. See (Millikan 1984, 18–49).

fact that it applied to Pluto meant that we could use this concept to think and speak about a property exhibited by nine objects in our galaxy. However, our being able to do this cannot explain why we had that particular concept of planethood: The pre-2006 concept of planethood was developed before Pluto had been discovered and, presumably, the discovery of Pluto did not add any evolutionary advantage to the concept.¹⁵ Therefore, the function to denote a property exhibited by nine objects in our galaxy cannot be a proper function of our pre-2006 concept of planethood. The present way of operationalizing *Functional Limits* thereby avoids the undesirable consequence that the 2006 revision of our concept of planethood was unacceptable. I assume that we can write off many problematic function ascriptions in a similar fashion. There is a general reason for thinking this: Selection effects tend to favor concepts that perform valuable functions. If a particular function is practically worthless, it is less likely that appealing to it can help explain why we have a concept that can performs it. In this way, a function's counting as a proper function will tend to coincide with it being a function worth serving. By consequence, proper functions may tend to be the functions we *ought* to do conceptual engineering in the service of.

There is one minor problem and one major problem with the current strategy. The minor problem has to do with our epistemic access to the proper functions of our concepts. If we take seriously the idea that we need to identify a proper function to ascertain whether an instance of conceptual revision is acceptable, conceptual engineers will have to engage in some hard empirical work. Sure enough, if we identify some valuable task F that a concept C enables us to do, we have the beginnings of an explanation for why we have C, and this is *prima facie* evidence in favor of believing that F is a proper function of C. But the belief is highly defeasible. There are any number of reasons why a concept could proliferate

¹⁵ I assume that the discovery of Pluto was an ordinary empirical discovery and did not constitute conceptual change from the concept of planethood we operated with before the discovery.

or survive. The reason why we have a concept of insanity could be, echoing Foucault (1973), that it may aid us in confining socially undesirable people, rather than to identify a psychological condition that is detrimental to the deviant individual. It is not clear how we should go about to assess the quality of these competing explanations. Moreover, if our route to identifying a proper function is to identify some valuable task F that a concept enables us to do, then we should start to wonder why we have to take the detour of determining whether F is a proper function. If we already know that F is a valuable task enabled by the target concept, that is arguably all we need to know to be justified in evaluating revisions of the concept according to whether or how well they let us carry out F.

This brings us to the major problem with articulating *Functional Goodness* and *Functional Limits* in terms of proper function: The fact that a concept has proliferated and survived thanks to a function F does not entail that F is worth (pre-)serving. Recall the extra-marital concept of rape. It is in the interest of actual and potential offenders not to be liable to punishment for sexual violence against their spouses. This can explain why we had a concept of rape that did not apply to acts within the bonds of marriage. Arguably, the extra-marital rape had the *proper* function of protecting wrong-doers from liability to punishment. But just as little as might makes right, an explanation is no justification. It would be wrong to require that this function be preserved by revisions to the extra-marital concept of rape. Millikan avows that “[t]he task of the theory of proper functions is to define this sense of ‘designed to’ or ‘supposed to’ in *naturalist*, *nonnormative*, and nonmysterious terms.” (Millikan 1984, 17, emphasis added). Millikan’s goal is to develop a theory of language and content that, among other things, can aid *explanations* for why we have the concepts that we in fact find ourselves with. Meanwhile, our goal is to define *normative* limits on conceptual revision and what it is for a concept to be *good*. For this purpose, the fact that a function can at once be harmful *and* explain why we have the target concept is deeply problematic.

4.4 System functions

As noted earlier, Thomasson (2020a) suggests that we might use Cummins' notion of system function to flesh out the functional approach to conceptual engineering, and Haslanger (2020) makes a remark favorable to a view along these lines. Cummins' main motivation for developing an account of function is to wrest the notion free from etiological and teleological presuppositions (Cummins 1975). Thus, on Cummins' view, an ascription of function to an item does not (and should not) entail that the functional item *exists* or is present in a system *because* it performs the target function. Instead, Cummins thinks of function more austere, in terms of the effects an item has on the capacity of a system. His example to illustrate the account is an assembly-line production (Cummins 1975, 760). The production line constitutes a system with the capacity to produce a commodity. To explain the system's capacity to produce a commodity, we may appeal to the capacities of machines or workers along the line. On Cummins' view, the various tasks performed by components of the system count as functions of those components relative to an explanation of the higher-level capacity of the system itself (Cummins 1975, 758–65).

Although system functions are different in kind from proper functions, the proposal to use them as the operative notion in *Functional Goodness* and *Functional Limits* runs into the same kind of problems. First, there is an epistemic problem. What exactly is the system, the components of which are concepts? Is it our conceptual repertoire, our discursive practice or something else? This is not at all obvious. But let us set this question aside, because we still face a daunting task: What is the capacity of the system which the concept is supposed to contribute to, and what kind of capacity of the concept makes a contribution to it? A production line has an easily definable capacity, but this is not true of the system candidates that concepts are components of—at least if we exclude capacities that are methodologically insignificant, such as the capacity to think, speak or interact. This inhibits us from knowing what the function of a concept is, because the function of

a concept is only system function relative to its contribution to the capacity of the system of which it is part.

The bigger problem will come as no surprise: Operationalizing Functional Goodness and Functional Limits with the notion of a system function does not go one step toward discounting harmful and insignificant functions. Whatever we consider to be the system of which concepts are components, it cannot be something perfectly benign or optimized along some non-moral dimension. To illustrate, our conceptual repertoire and our discursive practice are both systems with a capacity for oppression. Insofar as a concept consistently contributes to the system's capacity for oppression, the concept is performing a harmful system function. Indeed, the details of Cummins' account are not important on this score, because the notion of a system function does not even begin to select for the functions we ought to engineer in the service of. As the notion of a proper function, the notion of a system function was designed for explanatory purposes—for use in scientific explanation, specifically—not for normative purposes such as determining what revisions of a concept must be like.

5. Consequences for the functional approach

We do not have a notion of function that solves the selection problem for us. None of the notions surveyed in the previous section latch on to all and only the functions we want our concepts to serve. The technical and semi-technical notions we have reviewed may have independent merit, even as applied to concepts, but they do not add much if anything to our effort to solve the normative issues of present concern: of determining what makes a concept good and of determining the limits of revision.

In saying that the limits of revision are defined by functional continuity, or in saying that how a concept is to be evaluated is determined by its function, it may sound like we are solving normative, methodological problems regarding

conceptual engineering by identifying some non-normative property of the target concept, viz. its function. However, the solution is clearly incomplete. First, we have no reason to think that our concepts perform exactly one function each. Any given concept is used for a plurality of different tasks, and is bound up with our speech, thought and action in innumerable ways. On an unsophisticated way of thinking about concept function, the functions of a concept just are the tasks and purposes for which we use a concept. On my understanding, there is nothing problematically unclear about this way of understanding function, but the tasks and purposes for which we use concepts are many and varied. Some are good, some are bad and some are insignificant. This gives rise to the selection problem: The unsophisticated notion of function applies to many functions we do not need, and often better not, engineer in the service of. But since advocates of the functional approach have suggested more demanding notions of function, it would be foolish not to consider what these notions can do for us in our attempt to spell out *Goodness* and *Limits*. However, a review of the notions on offer makes it clear that whatever notion of function we opt for, we need to make further evaluation of the target functions. It does not suffice to certify that they are functions in this or that technical sense of the word ‘function’.

For the specific task of defining normative constraints on conceptual engineering—as opposed to, say, the task of explaining why we have the concepts we currently do—we do not have any real reason to favor the technical notions of function over an unsophisticated notion. The investigation, then, offers some support to the deflated view of function advocated by Riggs (2021) and Nado (2021a). The view may be summed up by the suggestion to “think of a concept’s functions as being merely extrinsic, relational properties like ‘is used by x for y’.” (Nado 2021a, 1522) The notion of what we use something for is unsophisticated yet clear. The fact that it applies to harmful and methodologically insignificant function candidates is less problematic than what it might seem, because no matter how technical we make the operative notion of function, we still have to assess

whether the target functions really are worth performing and preserving. Indeed, if we judge by the functions practitioners in conceptual engineering have actually appealed to, there is nothing to suggest that they are using a more fancy notion. Woodward suggests that we use our concept of causation to identify “relationships that are potentially exploitable for purposes of manipulation and control.” (Woodward 2003, 25), and designs concepts of causation that lets us do this. Something similar may be said for Craig, who assumes that we possess our concept of knowledge in order to “flag approved sources of information” (Craig 1990, 11).

According to the argument presented, there is no general, non-question-begging way to specify the range of functions we should do conceptual engineering in the service of. However, this does not spell defeat for the functional approach. As argued in section 2, there are good reasons to understand *Goodness* and *Limits* as relative to what we use a concept for. In addition, there clearly exist concept functions in several senses of the word ‘function’. Many of them are worth performing, just not all. The advocate of function therefore has warrant to claim that there exist concept functions, and that some of them play a normatively constraining role for exercises in conceptual engineering. These are comparatively weak claims, but they still have methodological significance. Once we have identified some benign and worthwhile function of a concept, we may evaluate revisions of the concept according to whether and how well the revised concept performs the function in question. In view of the selection problem and the absence of a neat solution to it, conceptual engineers just have to assess functions on a case-by-case basis. Conceptual engineering was never supposed to be easy.

Chapter 4

Inferentialist Conceptual Engineering

With Guido Löhr, Eindhoven University of Technology

1. Introduction

To engage in conceptual engineering is to try to make changes to our conceptual repertoire. It occurs within philosophy, e.g. when a philosopher proposes that we use a concept of her design; in politics and law, e.g. when key terms are defined in policy documents and court rulings; in science, e.g. when a researcher operationalizes a concept for the purpose of an experiment; and in social movements, such as the body positive movement's effort to broaden our concept of a beautiful body. To illustrate with examples from philosophy, Scharp (2013) argues that the alethic paradoxes show that our concept of truth is inconsistent, and therefore, that we should replace it with concepts of truth that are consistent. Woodward (2003) admits that his account of causation may not capture exactly what we mean by the word 'cause', but argues that his account provides us with useful concepts of causation, and thus good things to mean by the word 'cause'.

One of the central tasks facing theorists of conceptual engineering is to better understand what is being engineered. Theorists of conceptual engineering come to the field armed with different views on language, mind and practice, and correspondingly different views about what it would be fitting to engineer. Are conceptual engineers proposing new concepts (whatever concepts are, c.f. Isaac 2020), new linguistic meanings (Cappelen 2018), new classification procedures (Nado 2020), new conceptions or beliefs about a subject matter (Sawyer 2020; Machery 2017), or several of the former? One's answer to this question has implications for one's further views about conceptual engineering. If, e.g., one believes that conceptual engineers purport to revise the meaning of natural

language expressions, and one endorses externalist views about the foundations of meaning so construed, one is liable to think that conceptual engineers are engaged in a futile project (Cappelen 2018; Deutsch 2020).

To complicate matters, there are a number of different theories of concepts in philosophy and psychology, a number of different theories of meaning in philosophy and linguistics, and different ways the terms ‘concept’ and ‘meaning’ are used. If, for example, you believe that concepts or meanings are so-called *prototypes*, you might think that conceptual engineers are supposed to revise or construct new prototypes, but then the practice is not being carried out appropriately since few if any conceptual engineers proceed by way of designing a prototype; i.e., by specifying *typical* and *cue-valid* features of a category (see Rosch and Mervis 1975).

The plurality of views on concepts and meaning also comes with a danger of miscommunication: A sentence such as “Conceptual engineering is the practice of constructing and revising concepts.” is susceptible to different interpretations, and this may engender verbal disputes over, say, questions about revision vs. replacement. This all shows a need to establish common ground in the field, or at least mutual understanding (c.f. Cappelen 2018, chap. 12; Isaac 2020; Koch 2020; Nado 2020). As Cappelen demurs, “Many of those who write about conceptual engineering are unclear on the exact nature of the entities being engineered.” (Cappelen 2018, 141).

Some have adopted the term ‘representational device’ (e.g. Cappelen 2018; Simion 2018; Cappelen and Plunkett 2020), presumably to talk about what is being engineered in a more neutral way. However, as Burgess and Plunkett (2013a; 2020) observe, this is not a perfect solution, since there is more to thought and talk than representation (Price 2011; Thomasson 2020b). Indeed, there is an important minority tradition in the philosophy of language and mind according to which we should understand concepts primarily in terms of their role in reasoning rather than their role in representation. For inferentialists such as Sellars (1953; 1997) and

Brandom (1994; 2000), it is more accurate to think of concepts as inferential devices than as representational devices.

An inferentialist approach to conceptual engineering has not been thoroughly explored in the literature. In our view, it should be explored since there are good reasons to view what is being engineered as inferential devices. In this paper, we argue that being an inferentialist about what is being engineered secures a rationale for engaging in the practice that is lost if we conceive of what is being engineered in austere, representationalist terms. On the assumption that conceptual engineering is worthwhile, this provides support for inferentialism about conceptual engineering. Our objective is not to establish that being or having an inferential role is all there is to meaning or conceptual content. Rather, our agenda is to recommend a shift of focus from the representational features of content to the inferential features of content for the purposes of doing and thinking about conceptual engineering. We start by exploring and developing an objection to conceptual engineering. We argue that the objection exposes a flaw in a representationalist conception of what is being engineered. In section 3, we argue that inferentialism provides an appealing solution. Finally, in section 4, we raise and respond to an objection to our argument.

2. An objection to conceptual engineering

In a paper reacting to the recent surge of interest in conceptual engineering, Max Deutsch raises a dilemma for conceptual engineers (Deutsch 2020). According to the argument, conceptual engineers are either engaged in the futile project of trying to stipulate the semantics of terms, or they are engaged in the feasible but somehow trivial project of trying to change speaker-meaning. In the final section of the paper, Deutsch puts forward an altogether different objection to conceptual engineering. The objection leveled here is not that conceptual engineering is unfeasible or trivial, but rather that the very idea of a conceptual defect (and *inter alia* of a conceptual virtue) is incoherent. If the objection is sound, it does not matter much

whether conceptual engineers are able or not to implement the products of their work, nor whether their success at it would count as a trivial accomplishment. Deutsch argues as follows.

A good way to speak of, and communicate about, knowledge, free action, and women is to use terms that semantically refer to these things, and the terms that semantically refer to these things include ‘knowledge’, ‘free action’, and ‘woman’. So, the usual rationale for engaging in conceptual engineering is a bad rationale: since our terms are not, in fact, defective, relative to the purpose of using them to speak of their semantic referents, there is no need, and no value, in trying to improve them. (Deutsch 2020, 3955)

We agree that it is not obvious that terms (or what they express) are the sorts of things that can be better or worse. But what is being engineered has to be the sort of thing that is better or worse, should the practice be worthwhile: There would be no point in attempting to revise a concept if there is no useful sense in which the revised state of the concept is an improvement on the unrevised state. Now, Deutsch assumes that the point of having a term is to allow us to speak of its “semantic referent”. The broader picture is something like this: The world is abundant with objects, properties and relations. Words are our tools for representing them linguistically. Perhaps our toolbox has room for additions—Deutsch has no argument against introducing new terminology for uncharted phenomena—but conceptual engineers are misguided when they try to change the reference of terms that are already in use. Since our terms already have a reference, they already do what they are supposed to do, which is to represent some selection of objects, events, properties or relations. If a term already has a reference, it is thereby a non-defective representational device. Conceptual engineering, therefore, has a bad rationale because it is the practice of repairing non-existing defects.

Let us flesh out the objection with the well-defined semantic notions of extension and intension. Herman Cappelen, one of the critical targets for Deutsch’s

objection, has defended a view of conceptual engineering as the practice of trying to change the extensions of linguistic items via changes in their intension, and also argued that we ought to engage in the practice so construed (Cappelen 2018; 2020). We argue that extensions and intensions are not the sorts of things that can be better or worse as such. However, the assumption that what is being engineered is *simply* a representational device—that it is *simply* a tool for accessing extensions via an intension—is well worth questioning. Instead of concluding that conceptual engineering is not worthwhile, we may instead fit our view of what is being engineered to the assumption that conceptual engineering is worthwhile. This, in turn, supports our inferentialism about what is being engineered.

2.1 Extensions

Extensions are nothing more than sets of particular objects, events, relations, or property instances that linguistic expressions denote or refer to. The extensions of ‘cow’, ‘electron’ and ‘marriage’ are, respectively, the set of all cows, the set of all electrons and the set of all marriages. The extension of a term changes when particulars of the relevant kind go in and out of existence. As a preliminary observation, note that the meaning of a term never improves or deteriorates through these kinds of changes: The birth or death of a cow should not alter our evaluation of the meaning of ‘cow’.

Since extensions just are sets of particulars, it is dubious that they are the sorts of things that can be better or worse. Take any such set E . It is, firstly, absurd to think that there could be some other set E^* that is better at being E than E . The set of *all married couples* is not better at being the set of *all married couples consisting of one man and one woman* than the latter is at being itself. More strongly, there is not a useful sense in which a given set of particulars is good or bad just as such. We may explain why (a range of) same-sex couples ought to count as being married if we assume that being married has some kind of significance, be it practical, theoretical, moral, symbolic, legal etc. Just as a set of particulars, however, there is no reason

to prefer the one extension-candidate over the other. If we say “the number of marriages between one man and one woman has increased”, we have not made use of an objectionable concept. If there is anything objectionable about this set, it cannot consist in its being the denotation of an arbitrary linguistic item.

To support this line of thought further, consider what might seem like particularly bad extension candidates. First, consider the null-set. There is nothing wrong with an empty extension in and of itself. Atheists do not have an objection to the concept of god common to most religions—i.e. the concept of an omniscient, omnipotent and omnibenevolent creator. Rather, when they engage in arguments over the existence of this being, they necessarily have to make use of a concept that they believe has an empty extension. The *concept* is in good shape, even if belief in the existence of God is, from the atheist perspective, not. Next, consider a set that fails, according to metaphysical realists, to carve out a joint in nature, such as the set of all cows and electrons (Sider 2011, 3). The set itself just is what it is: cows and electrons. The set of cows does not become a bad set if you add electrons to it. Moreover, denoting this set is perfectly acceptable. If you use ‘cow or electron’ in a sentence, you have not thereby violated a rule of conceptual ethics. It is, of course, fair to object that we do not need a simple term to denote this set, or that the concept of a cow or electron has no utility in (scientific) description, explanation and prediction. But these are not reasons to think that the set itself is somehow bad, never to be denoted by a linguistic expression.

If there is an unexamined way in which extensions are good or bad, it has not been communicated in the literature. In the meantime, we conclude that extensions are not the sorts of things that are good or bad as such.

2.2 Intensions

If extensions are not good or bad as such, conceptual engineering has a bad rationale if it is the practice of attempting to change the extensions of linguistic items for the sake of it. However, it is well-established that having a meaning

consists in more than having an extension. ‘Creature with a heart’ and ‘creature with a kidney’ are co-extensional, but not identical in meaning. Indeed, Cappelen specifies that the changes in extension that are relevant for conceptual engineering are those that are “driven by changes in intension” (Cappelen 2018, 62). Let us therefore consider whether the addition of intensions can, contrary to Deutsch’s objection, secure a rationale for engaging in the practice. For this purpose, let us adopt a standard construal of intensions as functions from possible worlds to extensions.¹

The output of an intension is an extension. However, as we just argued, extensions are not good or bad as such. It follows that intensions are not good or bad in virtue of having a particular output. The remaining possibility is that the function from possible worlds to extensions can itself somehow be good or bad. Now what could this consist in? In the absence of other obvious answers to this question, we have one suggestion: Perhaps the function is defective if it fails to yield a determinate extension for some or all possible worlds. Several philosophers, including Carnap (Carnap 1962, chap. 1; Leitgeb and Carus 2020), have viewed indeterminacy as a conceptual defect, so this is far from an outlandish suggestion.² If indeterminacy is a defect, we have something for conceptual engineers to repair.

Before assessing the suggestion, let us first clear up a technical conundrum. In the logical/mathematical sense of ‘function’ at play in thinking of intensions as functions from possible worlds to extensions, a function is a binary relation that maps each element from one set (each possible world) to exactly one element from another set (an extension, viz. in/at that possible world). Given that intensions are such functions, something is not an intension if it fails to yield an output for some

¹ It is not crucial that the circumstance of evaluation is a possible world. With uncontroversial adjustments, the argument applies to alternative conceptions of a circumstance of evaluation.

² Note that extensions may also be indeterminate as such: The ensuing discussion applies equally to indeterminacy at the level of extensions.

range of inputs. It would then seem impossible for an intension to fail to determine an extension, and the proposal to fix defective intensions would be a contradiction in terms. Now, the concepts of intension and extension were originally devised for artificial languages that are free of indeterminacy (Carnap 1956).³ To the extent that natural language semantics exhibits vagueness and other kinds of indeterminacy, the notions of an extension and intension do not apply as defined. Conversely, insofar as they do apply, they are not defective, and hence not in need of repair. However, we think it would be unfair to dismiss the suggestion on these technical grounds. Instead we just assume that there is a sense in which an intension can fail to yield a determinate extension.

There are still two major problems for the suggestion that conceptual engineering could consist in repairing indeterminacies in our representational devices. The first problem is the severely limited scope conceptual engineering would have on this proposal. Sure enough, some engineering projects aim, at least in part, to repair indeterminacy. Examples include Carnap's (1962) effort to develop an exact, quantificational concept of inductive confirmation, and the IAU's decision to give a more precise definition of 'planet'. But the suggestion does not provide a rationale for other *prima facie* legitimate projects in conceptual engineering. Consider the effort to broaden our concept of marriage so as to include same-sex couples, or Sally Haslanger's project of developing concepts of gender to serve as "effective tools in the fight against injustice" (Haslanger 2000, 36). These projects are not in any useful sense attempts to repair indeterminacies, but we take it that a theory of conceptual engineering should account for them.

The more fundamental problem with the suggestion is that it is implausible to think that indeterminacy is always a defect. Ordinary discourse is full of terms

³ Unfortunately, this has carried over to thought about the semantics of natural languages as well. As Ludlow objects: "The received wisdom seems to be that semantics demands precision and fully determinate meanings. Whatever the merits of precision and fully determinate meanings, semantics has no need for them." (Ludlow 2014, 7)

that do not have perfectly exact meanings. Indeterminate terms and phrases pervade most areas of discourse. It is hardly plausible to think that all of them are in need of repair. Consider ‘bald’, ‘working class’, ‘kitsch’, ‘accessory’, ‘minimalist’, ‘sandwich’ and ‘harmony’. None of these terms encode an intension that yields a determinate extension, but they are not thereby in need of repair. If this is so, indeterminacy is not always a problem. But then we need something to explain the difference between the cases where it is a problem and the cases where it is not. Arguably, what decides whether indeterminacy is a problem is what we use a term for (or, as we are going to propose, what the downstream consequences of its application are).

As intimated, we cannot see any other way in which it makes sense to discriminate between functions from possible worlds to extensions as good or bad. Our review does not exhaust the space of possibilities, but in the meantime we conclude that an intension is not the sort of thing that is good or bad as such.

If what is being engineered is simply a representational device—if it is simply a device for accessing extensions via intensions—conceptual engineering has a bad rationale. It is not worthwhile to engineer intensions or extensions for the sake of it. We could, like Deutsch, use this line of thought to object to conceptual engineering. Fortunately, the objection is contingent on an austere, representationalist conception of what is being engineered. Rather than concluding that conceptual engineering is not worthwhile, we may instead fit our view of what is being engineered to the assumption that conceptual engineering is worthwhile. To find out how we should do this, we think it is helpful to consider why we care about whether a concept applies to a particular.

3. What makes conceptual engineering worthwhile?

It means something to us, in the common sense of the phrase, whether an action is intentional, whether two events are related as cause and effect, whether a sentence

is true, or whether a state is a democracy. We do not have concepts of intentional action, causation, truth or democracy simply to parse the world into those particulars that do and those particulars that do not fall under the concept in question. Classifying a state as a democracy or classifying a sentence as true has consequences, both theoretical and practical. It would be rather uninteresting whether our concept of truth applied to a sentence, if nothing followed from this. Of course, something *does* follow: From the assumption that a sentence ‘p’ is *true*, it follows that p. We make inferences in accordance with this schema all the time. But if we pretend that our concept of truth does not play this inferential role (or anything like it), we would not have much reason to care about what fell in its extension. We could then ask, “So what if the sentence is true?” and we would not have a good answer. If a word functions simply as label for an extension, whether it does so via an intension or not, we do not have good reason to care about what falls under the label. Jonathan Weinberg reasons similarly:

Our interest in philosophical concepts like PERSON and VOLUNTARY is not just to parse the world in such-and-such a way. Rather, we think that persons should be treated differently than non-persons (only they get rights, perhaps), and voluntary actions should be treated differently than involuntary behaviors (only they are morally evaluable, perhaps). (Weinberg 2006, 32)

In the previous section, we noted that having a meaning consists in more than having an extension—it must at least consist in having an extension and an intension. However, as we argue presently, we need to suppose that there is yet more to meaning than having an extension and intension, if meaning is to be an apt object of revision. What seems to be missing is the downstream significance of using a word with a given meaning, i.e. the downstream significance of applying a concept in speech.

Once we assume that concepts have consequences of application, we have reason to care about what it takes for a concept to apply. What it takes for a concept to apply should make the consequences of its application appropriate. To illustrate,

the far-right terrorist Anders Behring Breivik was nearly ruled exempt from liability to punishment on the grounds of a proposed diagnosis of paranoid schizophrenia.⁴ If being classified as a paranoid schizophrenic makes a person exempt from liability to punishment, it is crucial that what it takes to count as paranoid schizophrenic are conditions that make this consequence appropriate. Alternatively, the conditions for counting as liable to punishment could be revised. In either case we see a normative interplay between conditions and consequences of application.

The interplay between conditions and consequences of application is also lucid in the case of our concept of marriage. We have reason to care about who gets to count as married in light of the consequences of being married, especially the rights and duties one acquires as a married couple. It is unjust to preclude same-sex couples from enjoying these rights and duties. If we leave out these consequences, the project of trying to change our concept of marriage begins looking absurd: Proponents of same-sex marriage did not fight for the possibility of same-sex couples being in the extension of an austere representational device. As Dummett writes, “A naïve view of language regards the assertability-conditions for a statement as exhausting its meaning: the result is to make it impossible to see how meaning can ever be criticized, revised or rejected...” (Dummett 1973, 455). The view of language Dummett is objecting to is an instance of what Brandom (1994, 121) calls a “one-sided theory of meaning”; a theory that identifies meaning exclusively with what it takes for a concept to apply, or exclusively with its downstream significance for action and further thought and speech. What is absent in the representationalist view we have reviewed is an appreciation of the downstream significance applying a concept has. However, it seems crucial to take

⁴ <https://www.nytimes.com/2011/11/30/world/europe/norway-killer-of-77-was-insane-during-rampage-prosecution-says.html> (accessed July 18, 2021). The court ordered, and eventually sided with, a second psychiatric evaluation according to which the subject was sane at the time of committing the acts and hence liable to punishment (TOSLO-2011-188627-24, 47-71, accessed July 18, 2021 at <https://lovdata.no/static/file/1282/toslo-2011-188627-24-eng.pdf>).

a concept's consequences of application into account when assessing what its conditions of application ought to be.

Inferentialists like Brandom and Sellars identify concepts with inferential roles, which is to treat concepts as things with both circumstances and consequences of application. To better understand exercises in conceptual engineering, and to appreciate their rationale, we propose that we view what is being engineered as the things that inferentialists believe concepts are.⁵ In line with our diagnosis in this section, this promises to make better sense of conceptual engineering. Not least, it fits the assumption that conceptual engineering is worthwhile. In the remainder of this paper, we first explore inferentialism as a theory of meaning or content, then explain what is to be inferentialist specifically about what is being engineered, and finally we spell out and respond to an objection to our proposal.

3.1 Semantic inferentialism

Any time we use language to make claims, the statements we produce are essentially the sort of things for which reasons may be asked, and which may themselves be given as reasons for further statements, beliefs or actions. To illustrate, consider

- 1) There is a shark in the bay.

This may be given as a reason for any number of beliefs and actions, including

⁵ Thomasson (MS) also proposes that we understand what is being engineered in inferentialist (and “artifactualist”) terms, because it makes better sense of various features of conceptual engineering, including that what is being engineered (or targeted for revision) is a functional artifact with norms of use, and because it promises to rationalize projects in conceptual engineering. In essence, what we do in the present paper is develop the latter reason for being inferentialist about conceptual engineering at length.

2) There is an animal in the bay.

and

3) You should not go for a swim.

We may also give reasons for producing or endorsing 1), such as

4) I saw a fin moving through the water.

Any statement has something for which it is a reason, and *any* statement is something for which reasons may be given. Thus, all statements play the dual role of premises and conclusions in possible arguments. This is a simple, uncontroversial fact, but what to make of this fact is not obvious. Many would view this fact just as the consequence of further facts about the speech act in question, or of the features of arguments and logical connectives, and not itself a deep fact about meaning or content. For inferentialists like Sellars and Brandom, by contrast, having a role in reasoning is what meaning most fundamentally consists of. They identify the content of a statement with *what the statement is the production of a reason for* and *what is a reason for producing or endorsing the statement*. Only propositional contents—the contents we express with declarative sentences—are thereby identified, because only these contents are admissible as premises or conclusions. However, propositional contents may be decomposed. By substituting one sub-sentential expression for another, we may detect the different contributions the expressions make to the inferential role of the sentences in which they occur. For example, if we substitute ‘shark’ for ‘duck’ in 1) we get

5) There is a duck in the bay.

which has a different inferential role than 1), owing to the different contributions that our concepts of a shark and of a duck make to the content of the statements. To be sure, 2) follows from both 1) and 5). But 3) does not follow from 5), and while 4) is a reason for 1), it is not a reason for 5). Thus, while only propositional contents are admissible as premises and conclusions, concepts also have inferential roles, to

be identified by the contribution they make to the inferential role of the propositional contents they are components of.⁶

Now, the inferences we have surveyed are formally non-valid. Many would understand these inferences as holding only in virtue of further, tacit premises; paradigmatically conditionals, e.g.:

- 6) If there is a shark in the bay, then you should not go for a swim.

By contrast, Sellars and Brandom do not see goodness of inference as something that rests on logical form. For them, the notion of a good inference is prior in the order of explanation to the notion of a formally valid inference.⁷ This is not to say that logic is superfluous on their view. Instead, they develop an expressive understanding of logic. On this view, logical vocabulary serves to make explicit inferential proprieties that are otherwise implicit in our discursive practice. This expressive function is important. Once we have a practice of claiming and inferring, logical vocabulary allows us to endorse inferences in the form of claims, which in turn allows us to give and ask for reasons to endorse those inferences. We

⁶ There are at least two ways of understanding the target inferential relations. First, one may be concerned with the inferences we are disposed to make. This has us treat concepts as subject to empirical, psychological inquiry, e.g. as described by Machery (2009; 2017). Second, following Sellars and Brandom, one may be concerned with the inferences we are committed or entitled to make. Our inferentialism about conceptual engineering remains neutral between the former, dispositional construal and the latter, normative construal (however, see Löhr 2021). Attempts to change our conceptual norms and dispositions are closely tied together: A change of norms may engender a change of dispositions and vice-versa. We believe that conceptual engineers in the first instance care (and should care) about the normative relations, and we think it is fair to assume that dispositional changes follow suit to normative changes, but our argument does not hinge on these assumptions.

⁷ This is also why we follow Brandom in using the phrase “goodness of inference” and not “validity of inference”. If there are proprieties of inference without (or prior to the introduction of) logical vocabulary, we need to distinguish good and bad inferences without assimilating the former to formally valid inferences, and the latter to non-valid inferences.

may endorse an inference from p to q by stating “If p , then q ” or endorse an inference from a ’s being F to a ’s being G by stating “If Fa , then Ga ”.

If we combine the idea that concepts are inferential roles with an expressivist view of logical vocabulary, it is reasonable to predict that philosophers would try to capture the content of concepts by means of conditional and bi-conditional statements. And of course, this *is* widespread practice in philosophy. For example,

(*Knowledge*) A knows that p if and only if p , A believes that p , and A ’s belief that p is justified.

If the content expressed by ‘know’ is an inferential role, it makes perfect sense to try to capture it by a bi-conditional claim, as per (*Knowledge*). The claim expresses when we may infer that someone knows something, and what follows from someone’s knowing something.⁸ If we grasp what it takes for a concept to apply to a particular and what follows from a concept’s application, there is nothing more to grasp, according to the inferentialist, to grasp that concept. This way of identifying the content of a concept is at least as suitable for concepts we could have been or ought to be using as it is for concepts in actual use. It is thus a fitting way of identifying or presenting engineered concepts. This was anticipated by Frege in the *Begriffsschrift* (Frege [1879] 1972, 103–7; 1979, 12–13). As Brandom writes, “Employing the explicating logical locutions of which the conditional is the paradigm is to enable what Frege calls ‘the scientific formation of concepts.’ Such concepts will wear their contents on their sleeves; the inferential proprieties in

⁸ This is not to say that it captures everything that follows from someone’s knowing something, or everything from which you may infer that someone knows something. For example, it may follow that we ought to trust what A says about a matter M , if we have found that A knows that p , where p is some proposition that pertains to M . Generally, concepts have consequences of application that exceed what we may capture by the conditions that are necessary for its application. There may be consequences that hold, for example, as a matter of what has been described as typicality effects by prototype theorists, or what follows only in virtue of auxiliary hypotheses/beliefs.

virtue of which they mean what they mean are written down for all to read.” (Brandom 1994, 109) Indeed, this agrees with how concepts are identified by contemporary conceptual engineers. Both Haslanger (2000, 42) and Woodward (2003, 51), e.g., present the concepts they prescribe by means of bi-conditional statements. In producing such statements, they specify an inferential role. As an additional small step in making our case, then, we may note that inferentialism about concepts offers a natural way of understanding core theoretical claims made in prominent exercises in conceptual engineering.

3.2 Inferentialism about conceptual engineering

Our purpose in this paper is not to defend inferentialism about concepts or meaning. Nor, in fact, do we assume inferentialism about concepts or meaning. Our object of interest is *what is being engineered* when we engage in conceptual engineering. Whatever concepts happen to be, we argue that inferentialism is the right approach for understanding the object of conceptual engineering. We propose that what is being engineered are the things that inferentialists believe that concepts are, although we do not commit to, and much less do we argue for, the view that concepts are what inferentialists believe that concepts are. In this way, our view is compatible with any number of ideas about what concepts are and how to individuate them, even thoroughly representationalist theories of concepts.

There are many kinds of content we have reason to associate with speech acts, bits of language and thought. We have argued that inferentialists identify a kind of content to associate with sub-sentential expressions—their inferential roles—that fits the bill for what it makes sense to engineer, but this is not to preclude that one may associate other kinds of contents with sub-sentential expressions, nor to preclude that some of the latter are more deserving of the word “concept”.⁹ This

⁹ Riggs (2019) and Thomasson (MS) reason similarly. Riggs argues that it is our job to engineer a sense for ‘meaning’ that captures what conceptual engineers are trying to

pluralist attitude is warranted if we remind ourselves that the concept of a concept is itself apt for engineering. We have no objection to associating intensions and extensions with linguistic expressions. What we have argued is rather that this kind of content is not one that we have reason to revise. By contrast, the content identified by inferentialists is one we may have good reason to revise.

A conceptual engineer, on our proposal, purports to change a range of inferences we are committed, entitled or disposed to draw. In the paradigm case, the conceptual engineer will try to change when we are committed or entitled to infer that some particular counts as a woman, a marriage, as knowledge etc, where the consequences of counting as a woman, a marriage, as knowledge etc. are generally kept intact. For instance, proponents of same-sex marriage try (tried) to change the conditions a couple has to meet in order to qualify as being married, while not altering the consequences of so qualifying, e.g. the rights and duties of marriage. As we have argued, there is a straightforward motivation for this kind of project that is lacking from other conceptions of what is being engineered: Only in the light of having practical and theoretical consequences does it matter when our concept of marriage applies.

Let us consider a less traditional example, taken from the British version of the Swedish TV show *Real Humans*. Joe, a father of two, has sexual intercourse with the attractive service humanoid robot Anita. For this to be possible, all he has to do is enable the robot's adult content function. Anita is conscious and happens to be attracted to Joe but at the time, Joe has no reason to think that Anita is anything but a non-sentient mildly intelligent machine. Thus, he classifies his action as masturbation with a sex toy rather than as sexual intercourse. Even though Joe is trying to hide the fact he had intercourse with Anita, he does not think he has committed a moral wrong. He tries to hide his action not because he thinks it is

revise and construct. Thomasson observes that how we need to conceive of concepts to better understand exercises in conceptual engineering need not be identical to how we need to conceive of concepts for other purposes.

unethical but because he feels embarrassed. His emotional response is shame, not guilt. When Joe's action is revealed to the family by the company who made Anita, Joe's wife, Laura, considers his action adultery. Joe is surprised by this classification and argues against it. He says that Anita is just a machine and that acting with a machine or tool is not the same as cheating or adultery, which normally requires a sentient person (again at this point nobody knows that Anita is sentient). Eventually, he loses the battle against his wife and is asked to move out for a while.

The example of Joe is both challenging and complex. It is challenging because it confronts us with a novel situation that our conceptual tools were not shaped to resolve. It is complex in that it bears upon multiple inferences that are, could be or should be encoded in the concepts we are equipped with to make sense of the situation. Let us focus on whether Joe should count as having cheated on his wife. First, recall that Joe thinks of his act as some form of masturbation. However, does it follow from this that he has not cheated on his wife? If it does follow, it becomes pertinent to address what the application conditions of our concept of masturbation should be in the first place. Second, bracketing possible inferential connections between our concepts of masturbation and of cheating, it seems clear that classifying Joe's act as cheating has grave consequences. Classifying Joe's act as cheating entitles moral blame. Not least it entitles Laura to leave him. These inferential consequences figure in the background when we consider whether we should classify Joe as having cheated. They do so for good reason: They help decide what it *ought* to take for an act to count as cheating and thus whether Joe's act falls under the concept of cheating we ought to be using. It is possible that our actual concept of cheating applies to Joe's act, but also that his act does not warrant the consequence that Laura is entitled to break up with him. If so, the only appropriate way of resolving the situation is to repair our concept, or engineer new ones.

The example illustrates just how much inferential relations matter. If conceptual engineering is about engineering these relations, it is clearly an important enterprise. The gravity of its consequences makes it matter greatly how

we classify the husband's action. We find reason to inquire into what it *ought* to take for an act to count as cheating since what it *does* take to count as cheating is given by our actual concept of cheating, and that concept is not necessarily the tool we ought to be using to cope with the situation. We are tasked with figuring out what our concept of cheating should be, and, if not cheating, what concept could capture Joe's act and license appropriate inferential consequences.

4. Objection

We have argued that inferentialism about what is being engineered makes better sense of projects in conceptual engineering; not least because it secures a rationale for engaging in such projects, which is lost if we think of what is being engineered in representationalist terms. Against our proposal, one might object that our rationale for engaging in conceptual engineering need not derive from properties that are somehow inherent to what is being engineered. Although we have considered marital rights and duties as something that follows from the very application of our concept of marriage, one could think of these consequences as extraneous to the concept; as something we attribute to marriages by adopting certain further beliefs or commitments (e.g. a commitment to the effect that if A and B are married, then A and B may pay tax T as if they were one individual). Thereby, one might resist the conclusion that what is being engineered needs to have an inferential role.

The objection is naturally paired with the idea that conceptual engineering is normatively constrained by the function a concept serves (see e.g. Haslanger 2000; Brigandt 2010; Nado 2021a; 2020b; Thomasson 2020a). In the case of our concept of marriage, the idea would be that, in our practice, the concept performs the function of marking couples as the subject of certain rights and duties, where the rights and duties in question are not necessarily part of what it *means* to be married. Rather than ascribing consequences pertaining to rights and duties to the content of our concept of marriage, they could be ascribed to something we *do* with our

concept of marriage. To use a distinction from the philosophy of language, they would be something like pragmatic features, as opposed to semantic properties.¹⁰

Jennifer Nado's "Practical Role Account" falls in this category of views. According to the Nado, the conceptual engineer is in the business of "devising a classification procedure that successfully fills an intended practical role." (Nado 2020, 11) A classification procedure is "is a set of steps or rules which, when followed, determines an intension-candidate" (Nado 2020, 9) and a practical role is just the set of purposes for which people employ the target term. A purpose for applying the term 'marriage' to a couple is to mark the couple as the subject of a range of rights and duties; just what we previously described as an inferential consequence of applying our concept of marriage to a couple.

Our response to this objection is partly concessive. If our argument so far is sound, we might as well conclude with a disjunction: either the thing being engineered is an inferential device by virtue of its content, or it is an inferential device by virtue of the significance we have accorded it in our practice. In itself, this would be an interesting result. If conceptual engineering's having a sound rationale entails that the object of engineering plays an inferential role (by virtue of its content or our practice), then conceptual engineers need to pay heed to the inferential roles played by our concepts (no matter what they are grounded in). No matter their source, it will be crucial for conceptual engineers to consider the consequences of a concept's application. Only in the context of having theoretical and practical consequences does it matter what it takes for the concept to apply, and only in that context do we have good grounds for determining how a concept is to be engineered. Once this is accepted, we may further ask, first, what the inferential consequences of a concept's application are, and second, what those consequences hold in virtue of. One's answer to the second question may of course

¹⁰ This echoes Haslanger: "instead we begin by considering more fully the pragmatics of our talk employing the terms in question. What is the point of having these concepts?" (Haslanger 2000, 33).

be that some but not all consequences hold in virtue of beliefs, commitments or practices that go beyond the content of the concept, and some but not all consequences hold in virtue of the content of the concept.

How to circumscribe what is part of the content of a concept—as opposed to what pertains to beliefs, theories, commitments, etc. articulated with the concept—is a long-standing theoretical problem. We do not aim to solve it here. What we would like to note is that how we draw this line is tangential to the claim that we need to think of concepts as inferential devices in order to secure a sound rationale for engaging in conceptual engineering. If we conceive of the object of engineering austere, as something that has an extension and intension, and do not enrich this conception by assuming that what is being engineered is something whose application has theoretical and practical consequences (in virtue of *something*), then our view of conceptual engineering falls prey to the objection that the practice lacks a rationale.

Without attempting to identify exactly what pertains to the content of concepts, there is positive reason to favor the present proposal over accounts that ascribe consequences of application exclusively to the pragmatics of discourse. The present account makes it clear that inferential consequences need not be treated as fixed or given. They, too, are up for normative assessment and engineering in the light of such assessment. To flesh this out, first consider the following passage from Nado:

Further, purposes can be discarded. If an engineer comes to believe that a given purpose is undesirable or at least unneeded, the successor she devises is not obliged to retain it. For instance, racial concepts used to be implicated in pseudo-scientific explanations of variation in intellectual ability; this is clearly a theoretical purpose best rejected. (Nado 2020, 13)

Recall that for Nado, purposes for using a given classification procedure are to play the same normatively constraining role that inferential consequences play in our

account. However, as Nado correctly observes, these purposes may themselves be bad or somehow in need of revision. In our account, the revisability of inferential consequences follows naturally from the idea that consequences of application are part of the content of a concept, since conceptual engineers are concerned with revising the content of concepts. If we ascribe all consequences of application to things we do with a concept (as opposed to its content), we need to give two separate accounts: One of when and how to revise concepts, and another of when and how to revise the purposes/consequences/functions for which we employ concepts. By contrast, the present account promises a cohesive and symmetrical treatment of conditions and consequences of application.

On our account, to revise a concept's consequences of application is to engage in conceptual engineering just as much as revising its application conditions is to engage in conceptual engineering. After all, one concept's inferential consequences is another concept's inferential antecedents. How to revise a concept depends upon what part of an inferential role one is holding fixed. If we hold fixed that a diagnosis of paranoid schizophrenia makes you exempt from liability to punishment, we need to make sure that the diagnostic criteria make this consequence appropriate. Conversely, if we hold fixed the diagnostic criteria, we need to make sure that there are not any consequences of being classified as paranoid schizophrenic that are inappropriate relative to the diagnostic criteria. Thus, when engaging in conceptual engineering, we need to treat some part of a concept's inferential role as fixed, then assess how we should fill out that role: Paradigmatically, by identifying what its application conditions ought to be in light of consequences we are holding fixed, but potentially also by holding fixed when we may infer that the concept applies and revise the consequences thereof.

5. Conclusion

In this paper, we have turned an objection against conceptual engineering into an argument for how we should understand the practice. According to the objection,

conceptual engineering is not worthwhile because it does not make sense to assess what is being engineered as better or worse. The objection rested on a representationalist view of what is being engineered, according to which conceptual engineering is the project of changing extensions via changes in intensions. Instead of taking this view on board as an assumption, we instead sought a view of what is being engineered that is consistent with the assumption that conceptual engineering is worthwhile. We have argued that the things that inferentialists consider to be concepts are things that are worthwhile to engineer. Inferentialism about conceptual engineering captures the fact that only in the context of having significance for further thought, speech and action does it matter what it takes for a concept to apply. On our proposal, how we should revise or engineer a concept depends on its role in a web of concepts that are tied together by inferences we carry out in our discursive practice. We have shown by example how this view promises to make sense of exercises in conceptual engineering. It turns out that exercises in conceptual engineering make more practical sense if they are construed as exercises of revising or constructing inferential devices.

Chapter 5

Against Amelioration as Revelation

1. Introduction

It is not straightforward to square commitment to an externalist metasemantics with commitment to conceptual engineering as a project. Cappelen (2018) observes that externalist determinants of meaning, such as past baptisms, are not typically within our control. On this basis, he argues that conceptual engineers are not in a position to effect the changes they aim for.¹ In this paper, I examine a different point of interaction between externalism and conceptual engineering. On certain externalist views, words tend to mean what they ought to mean, irrespective of what we take them to mean at any given point in time. If an externalist metasemantics secures goodness of meaning despite our semantic ignorance, the only remaining task for a would-be conceptual engineer is to *reveal* what a word means. Sally Haslanger suggests an interpretation along these lines:

Social constructionists can rely on externalist accounts of meaning to argue that their disclosure of an operative or a target concept is not *changing the subject*, but better reveals what we mean. By reflecting broadly on how we use the term ‘parent,’ we find that the cases, either as they stand or adjusted through ameliorative analysis, project onto an objective social, not natural, type. So although we tend to assume we are expressing the concept of *immediate progenitor* by the term ‘parent’ in fact we are expressing the concept of *primary caregiver* (or some such); the constructionist shows us that our assumptions about what we mean are false, given our practice. This is not to propose a new meaning, but to reveal an existing one. (Haslanger 2006, 110)

¹ For criticism and discussion, see (Koch 2021a), (Riggs 2019), (Deutsch 2020), (Pinder 2021) and (Jorem 2021).

There is no need to change what is already in a good state. If a particular set of terms ought to denote the “objective types” of social reality, but externalism has already made them do so, theorists are left with identifying and making us aware of the meaning of the terms in question.

Derek Ball and Theodore Sider support relevantly similar views. According to Ball’s *temporal externalism*, what a word means at any given point in time is partly determined by future, reasonable agreement (Ball 2020). To illustrate this view, consider an utterance of ‘Two men cannot marry’ by someone in our past; say, 150 years ago. On one interpretation, the person has stated a truth, but done so with a meaning for ‘marry’ that is morally objectionable, and different from the meaning it has now. On Ball’s view, the person has stated a plain falsehood. Since we now (and presumably further into the future) are in reasonable agreement that same-sex couples can marry, we (or our descendents) determine that ‘marry’ never failed to apply, semantically, to same-sex couples.² As in the case of Haslanger’s proposal, an externalist mechanism is supposed to be responsible for the term’s having the meaning it ought to have, it is only that speakers in our past have been ignorant of it. Sider may be grouped together with Haslanger and Ball. He holds both 1) that there is a metasemantic force of *reference magnetism* which makes words attain meanings that “carve nature at its joints”, and 2) that meaning candidates that carve nature at its joints are better than meaning candidates that do not carve nature at its joints (Sider 2011). Again, this is a view on which an externalist

² Note that Haslanger’s interpretation of this case is different from Ball’s, and not in line with her interpretation of the ‘parent’ case: “When activists have claimed that same sex couples can be married, or that LGBTQ domestic arrangements are families, it wasn’t based on what we have meant all along, but on what we should have meant.” (Haslanger 2020, 257). As Haslanger acknowledges, her views about the semantics and metasemantics of revisionary projects have changed over time. In (Haslanger 2020), she defends and develops the externalist view set out in (Haslanger 2006), but also gives the (seemingly) conflicting verdict of ‘marriage’ and ‘family’ just quoted. For present purposes, there is no need to resolve the tensions in and between Haslanger’s writings, so as to reach the soundest possible interpretation. Here, I simply register that they exist.

mechanism makes words mean what they ought to mean (irrespective of what we take them to mean).

Following Cappelen's (2018, 79) description of Haslanger's view, let us use 'Amelioration as Revelation' as a name for the view common to Haslanger, Ball and Sider. In this paper, I criticize Amelioration as Revelation. I argue that we should not accept a metasemantic view on which natural language expressions always attain the meaning they ought to have. Thus, I argue that we should not accept an extreme form of Amelioration as Revelation (section 3). The view generates implausible verdicts about semantic change and dubious commitments about how to do semantic and metasemantic theorizing. However, once we set aside the extreme version, the moderate versions also become less appealing. The distinction between normative and descriptive assessments of meaning should instead see us questioning the verdicts we make about the hypothetical scenarios that are supposed to motivate externalism in the first place (section 4). This distinction has significant explanatory power. By harnessing it, we may avoid postulating mysterious mechanisms of meaning determination. Moreover, we should expect both our normative and descriptive theorizing about language to improve when we do not blur the *is* and *ought* of meaning. I begin my discussion (section 2) by laying out an assumption that will sharpen our focus of the point of disagreement.

2. What a term ought to mean

Conceptual engineering is motivated by the idea that natural and cultural evolution has not gifted us with a perfect conceptual repertoire. To progress and flourish we may need to develop new concepts: This may e.g. help us form better theories, explanations and predictions, or foster better, more equitable social affairs. To operationalize this idea for the present context, understand our conceptual repertoire as the set of word-meaning pairs we use; our *lexicon*. In many cases, we may improve our lexicon by expanding it, i.e., by adding a new word-

meaning pair. However, there are good reasons for not expanding each and every time we identify a meaning we ought to be using. Specifically, expansion is not a viable option if the meaning of a word in use has pernicious effects on our thought or practice. In that circumstance, we should rather revise the meaning of the word (provided that we want to continue using it). In addition, cognitive limitations may restrict how many different word-meaning pairs we are actively able to grasp and use. Sometimes, then, we may need to replace or revise concepts, rather than expand our conceptual repertoire.³ What a term means and what it ought to mean can thus come apart. According to proponents of Amelioration as Revelation, however, these cases are few and far between, because the metasemantics of our languages effectively takes care of our engineering for us.

In order to engage critically with Amelioration as Revelation, we need to assume that there is such a thing as what a term ought to mean. To use what is hopefully an uncontroversial example, ‘marriage’ ought to have a meaning such that two men are not precluded, from the very meaning of the word, to count as married. If ‘marriage’ had a meaning such that only couples consisting of one man and one woman could count as married, this would breed or constitute injustice towards non-heterosexuals. For an entirely different example, I think it is fair to suppose that ‘cow’ ought to denote a particular species and not, say, the union of cows, chairs and afternoon bus rides. Or consider the claim that ‘planet’ ought to have a meaning in accordance with the definition provided by the International Astronomical Union (IAU) and not one on which Pluto counts as a planet. I take these to be plausible examples of what particular words ought to mean.

It is important to note that neither Haslanger, Ball nor Sider have an objection against assessing concepts or meaning normatively. In the case of Haslanger and Sider, we find explicit claims about what terms ought to mean or

³ For additional argument, see (Cappelen 2020, 141–46) and the illuminating discussion of “conceptual maximalism” in (Dever 2020).

which concepts we ought to have.⁴ Indeed, we would struggle to define the view common to these philosophers if we did not assume that there is such a thing as what a term ought to mean, or that meanings are the sort of things that can be better or worse. There could be no Amelioration as Revelation if amelioration were not even *logically* possible. I underscore this point to show that our assumption is (or should be) in the common ground.

In order to clarify the assumption about there being such a thing as what a term ought to mean, let me specify what I am not assuming. First, I do not assume that what a term ought to mean is something that holds independently of any further fact. In particular, one might think that the fact that ‘marriage’ ought to apply to same-sex couples holds only in virtue of the (contingent) purposes for which we have/use the word ‘marriage’, or in virtue of facts about, say, the role it plays in our practice or in reasoning (see e.g. Nado 2020). What a term ought to mean might be relative to such factors. Second, I do not assume that there is one meaning that is the uniquely best meaning for a given term. There can be multiple, equally good meanings for a term, all to be preferred over other meanings. This still leaves us with a fact to the effect that the term in question ought to have one of *these* (as opposed to one of *those*) meanings, or that it should mean *this* rather than *that*. Our assumption is thus compatible with the pluralist points made by Chalmers (2011) and Ludlow (2014, 82). Third, when I write about what a term ought to mean, I do not intend to privilege any particular locus of meaning. We can talk about what ought to be the standing semantic value of an expression in a given language like English, what its semantic value ought to be in some non-standardly individuated language such as an idiolect, or what speakers in a particular context

⁴ “The problem is that they’ve got the wrong concepts. They’re carving the world up incorrectly.” (Sider 2011, 13)

“A purely ‘descriptive’ approach to the analysis of knowledge, I’ve argued, either ignores the normative question of what epistemic concepts we ought to employ, or assumes implausibly that the epistemic concepts we do employ are the ones we ought to.” (Haslanger 1999, 467)

ought to communicate (“speaker-mean”) by an expression.⁵ It is conceivable that normative facts about these loci of meaning come apart, especially if facts about what a term ought to mean are sensitive to contingent factors like what our purposes for having/using the term are.

3. Is amelioration always revelation?

On an extreme version of Amelioration as Revelation, the meaning determinants of natural language never fail to make terms mean what they ought to mean. For ease of expression, let ‘Always Revelation’ be a name for this thesis. In the following, I develop two (connected) arguments against Always Revelation. The first argument centers on the problematic consequences the thesis generates for our assessments of semantic change, in particular cases that look like semantic improvements. Always Revelation either deprives us of the best explanation for these cases or commits us to implausible verdicts about them. The second argument is that Always Revelation undermines the explanatory purposes for which we have and develop a metasemantics.

3.1 Semantic change

The phenomenon of semantic change is widely recognized.⁶ As generations pass on, and as our technological, natural and social circumstances change, meaning evolves. Here are a few examples.

⁵ In (Jorem 2021) I discuss whether these different meaning phenomena make for feasible and worthwhile ways of implementing a designed meaning.

⁶ What is more controversial is just how much semantic change there is, and how gradually it happens. Dorr and Hawthorne’s (2014) discussion of semantic plasticity and its counter-thesis, semantic robustness, illustrates this nicely. Dorr and Hawthorne make the plausible assumption that semantic facts supervene nomologically on physical facts, but show that it is not obvious how much physical change there must be of the relevant kind for semantic change to occur. According to plasticity, small changes in

The colonists who traveled to America spoke a version of English in which the semantic value of ‘corn’ was *cereal grain*. In American English today, the word denotes a specific cereal grain, namely maize.

‘Livid’ has an interesting history, narrated by Merriam-Webster as follows:

The Latin adjective *lividus* means "dull, grayish, or leaden blue." From this came the French *livide* and eventually the English "livid," which was used to describe flesh discolored by a bruise when it was first recorded in the early 17th century. A slight extension of meaning gave it the sense "ashen or pallid," as used in describing a corpse. "Livid" eventually came to be used in this sense to characterize the complexion of a person pale with anger ("livid with rage"). From this meaning came two new senses in the 20th century. One was "reddish," as one is as likely to become red with anger as pale; the other was simply "angry" or "furious," the most common sense of the word today.⁷

‘Planet’ originates from Ancient Greek where it had the meaning *wanderer*. It developed to mean something like *large celestial body that orbits around a star*. In 2006, the IAU voted to define ‘planet’. According to the resolution,

A planet is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit.⁸

physical facts bring about small changes in semantic facts. According to robustness, semantic facts change in a more jerky fashion. The assumption that there is semantic change is simply the assumption that meanings are not robust to an absolute degree. For present purposes, however, we do not have to determine exactly how plastic they are.

⁷ <https://www.merriam-webster.com/dictionary/livid> (accessed August 17, 2021). The example is borrowed from Allison Koslow (2020, 30–31, 96–97), whose dissertation is rich on examples of semantic change.

⁸ https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf (accessed August 17, 2021).

This resolution demoted Pluto from the status of a planet. Indeed, the protocol explicitly acknowledges this consequence. After initial resistance, the broader public has by now adopted a usage in line with the IAU, and I think it is safe to say that the meaning of the word ‘planet’ has changed.⁹

Semantic change is more elusive than changes in phonetics, syntax, spelling and punctuation. The latter are directly observable, while detecting semantic change may require more extensive investigation or support from theoretical hypotheses, especially so if the changes are small. Typically, our observations of linguistic behavior underdetermine whether we are witnessing a case of semantic change or a change in belief. Since semantic change is comparatively elusive, there is space for many different but all reasonable views about its prevalence. By the same coin, there may be a relatively large extent of reasonable disagreement about individual cases. Even so, the very existence of semantic change is not to be denied. Changes in meaning are, just as other linguistic changes, events we should want explanations for.

Some changes in meaning are probably best explained by appeal to random events—akin to mutations in biological evolution—such as misinterpretations or unmotivated whims of the moment, that hearers go on to copy. Further changes may come about for sensible reasons, but not because they improve prior meanings, e.g., when a metaphorical usage becomes the new normal usage (recall ‘livid’). However, it is very plausible that changes in meaning also arise when we try to satisfy our needs for inquiry or comply with our normative convictions. If we realize that the existing meaning of a term breeds injustice, given facts about our practice, we should change it (recall ‘marriage’). These changes do not need to happen with much metalinguistic awareness, as when the IAU voted to define ‘planet’. According to Ludlow (2014), Plunkett and Sundell (2013) and Plunkett

⁹ Note that I use ‘semantic change’ exclusively in the sense of what Armstrong (2016) calls ‘lexical shift’, as opposed to ‘lexical expansion’. In linguistics, ‘semantic change’ may be used for both kinds of change (see e.g. Blank 1999).

(2015), speakers frequently negotiate or litigate the meaning of words by using them in particular ways, as opposed to mentioning words in explicit statements about what their meaning ought to be. Such conversational moves may be picked up by other speakers and proliferate, and eventually result in semantic change. Whether consciously or unconsciously, semantic change may come about when we tailor language to better fulfill our objectives.

Always Revelation has a problem with the cases where we modify language to reach our objectives. According to the thesis, semantic change never constitutes an improvement. But then, since terms always and already mean what they ought to mean, a term can never change meaning because speakers identify an improvement. The cases that look like improvements on prior meanings would have to be explained as non-rational changes. Alternatively, they would have to be explained away as not being genuine cases of semantic change. Let us assess these interpretational options in order, focusing on the example of the IAU's decision to (re-)define 'planet'.

The suggestion that 'planet' changed meaning, but not because of the reasons in favor of the change is in deep tension with the historical facts. The resolution itself came about as the result of rational and occasionally heated argument. One of the principal reasons in favor of the motion was to make it so that members of the extension of 'planet' shared more of the properties that are of interest to astronomers (Ludlow 2014, 46–47). Of particular interest is the property of having cleared its neighborhood of debris (condition c, quoted above). This is exactly the kind of semantic change that Carnap would commend, as it satisfies his “fruitfulness” criterion on good revisions, according to which engineered concepts (“explicata”) ought to support generalizations (Carnap 1962, chap. 1). I take it that Carnapian fruitfulness is a plausible desideratum for assessing what a term ought to mean. If we thus suppose that the change in question *would* be an improvement *if* the term 'planet' did not already have it, *and* we suppose, with support from historical fact, that the members of the IAU reasoned accordingly, then it is

extremely plausible to think that the change happened because they identified an improvement. Provided that ‘planet’ did change meaning, Always Revelation undermines what seems like the only plausible explanation of the change that occurred. There is just no good way to explain the change if we do not appeal to the informed discussions in and around the IAU.

The other option for a proponent of Always Revelation is to urge that the case of ‘planet’ is not, in fact, a case of semantic change. More than the previous option, this is true to the spirit of Ball’s temporal externalism and Sider’s reference magnetism. After all, scientists did reach a *reasonable agreement*. Ball could argue that they thereby determined the meaning of ‘planet’—not by changing it, but by determining the meaning it had all along. Sider, meanwhile, could argue that celestial bodies that have cleared the neighborhood around their orbit make up a (relatively) fundamental kind that ‘planet’ was tracking all along; the fact that astronomers recognized this joint in nature with was but a matter of catching up with our semantics. Now, these interpretations of the case may seem acceptable in isolation, especially if one is already drawn to endorse reference magnetism or temporal externalism. However, we need to put the case into perspective.

First, it is clear that ‘planet’, at some point in our past (in Greek, in English or in between), changed meaning from *wanderer* to something like *large celestial body that orbits around a star*. We do not want to conclude that the word never meant *wanderer*, but the accounts Ball and Sider put forward do not specify a temporal range for the semantic facts that are determined by, respectively, future reasonable agreement and reference magnetism. Their views then leave it mysterious how a word—gradually or suddenly?—becomes subject to strong externalistic forces that lock its meaning into what it ought to be, whereas previously it changed meaning unaffected by the mechanism in question.

Second, Always Revelation betrays the impression that ‘planet’ looks a lot like any other case of semantic change with respect to the impression that *it is* a case of semantic change. Just as ‘corn’ and ‘livid’, the word seems to mean something, and

then something different. There seems to be no difference over the fact that semantic change occurred. Always Revelation is inconsistent with this impression. To make this more concrete, consider the following. A proponent of Always Revelation would hold that people in our past spoke falsely when they called Pluto ‘a planet’. By contrast, they would hold that early colonists to America spoke truly when they called wheat ‘corn’. It seems strange to give diverging verdicts about these cases. To be sure, there is a sense in which the pre-revision ‘planet’-users were making a mistake, since they were using ‘planet’ with a meaning that could be improved upon, but once we are equipped with the distinction between what a term does mean and what it ought to mean, we do not have to classify the mistake as one of expressing a falsehood. What makes the situation more precarious for the proponent of Always Revelation is that their contrasting verdicts about the cases are forced by abstract theoretical assumptions about how meaning is determined, as opposed to observation of how the target terms are used, and how their uses have changed over time. Always Revelation puts a constraint on diachronic semantics to the effect that one has identified a case of semantic change (as opposed to a change in belief) only if the later meaning is not an improvement on the prior meaning. This is to put the cart before the horse. As I shall argue shortly, metasemantics answers to semantics rather than the other way around.

Third, we need to keep in mind that ‘planet’ is but one example where a term (presumably) changes meaning and where the change in question is an improvement. Other examples include our transition to a meaning for ‘marriage’ that includes same-sex couples; our transition from a broad, pre-Linnaean meaning of ‘fish’ to a more precise meaning based on taxonomical desiderata (roughly: *aquatic, cold-blooded, gill-bearing, craniate vertebrates that do not have limbs with digits*); and the long-stretch transition from the Ancient Greek ‘atomos’—which expressed the concept of a hypothesized, indivisible particle—to the current meaning of ‘atom’, which denotes the smallest particle of a chemical element (an entity which turns out to be divisible). Are really none of these cases instances of

semantic change? To appreciate why we should not answer this question affirmatively, let us first take stock.

We are assuming that there is such a thing as what a term ought to mean. Therefore, we cannot deny that it is *logically* possible to improve meaning. We are also assuming that semantic change exists, although we have not assumed which cases count as changes in meaning and which count as changes in belief. A proponent of Always Revelation thus finds herself in the awkward predicament of denying that there is semantic change in precisely all the cases where semantic change makes the most rational sense, i.e., when an improvement has been made. This is a bizarre consequence of the view, considering that the problematic cases in all other respects look like cases of semantic change. If instead the proponent of Always Revelation accepts that there is semantic change in these cases, she has to deny that they count as improvements, but that deprives us of the best explanation we could give for the changes, and we should be left wondering which cases could count as improvements to meaning on their view.

Any individual case of something that looks like a case of semantic change may be explained away as not being a genuine case of semantic change, and any genuine case of semantic change may be understood and explained as a non-improvement. Individually, such interpretations might be tolerable, but collectively they make for an unpleasant cost of being committed to an abstract theoretical view that is in tension with how actual speakers use words. This is not intended to be a knockdown argument against the externalist views in question. Rather, it shows that we need compensation. The views need to have other theoretical virtues to outweigh the problems identified, or strong, independent arguments in their favor. However, as I argue in section 4, recognizing the distinction between normative and descriptive assessments of meaning should instead see us questioning a central form of argument in favor of these externalist views. Before that, let us take a look at the general shape of the problem facing a proponent of Always Revelation.

3.2 An unacceptable constraint on semantics

A semantics for a language is an assignment of meaning to the expressions of that language. As a fundamental desideratum, the assignment should be broadly consistent with how expressions of the language are in fact used, and it should help explain facts about communication and action based on linguistic communication.¹⁰ For example, assigning *water* as the meaning of ‘water’ may help explain why speakers go fetch bottles of water when they say things like ‘I’m going to get us some water.’ Crucially, this is a descriptive enterprise, not a prescriptive one. Semanticists are not tasked with finding an assignment that solves for the best possible or most useful repertoire of meanings. That is, their assignment of meaning is not constrained by norms pertaining to which concepts we ought to be using or which meanings our terms ought to have. Rather, semanticists have to take meaning as it is, benign or harmful; useful or impractical.

The purpose of identifying a *metasemantics* for a language is to explain its semantics. It is to explain why the expressions of the language mean what they mean:

Insofar as linguistic semantics aspires to ‘specify’ or ‘report’ the meanings of sentences and sub-sentential expressions (in some systematic way), philosophy of language could go on to tell us how or why these symbols come to have those meanings—perhaps unearthing more basic or fundamental facts in virtue of which such semantic states of affairs obtain. We take this second sort of inquiry to be paradigmatic, if not exhaustive, of metasemantics. (Burgess and Sherman 2014, 1–2)

This order of explanation is important. We do not start off with hypotheses about how meaning is determined, then decide which semantic propositions to hold true. Admittedly, we may seek help from theoretical hypotheses when our observations

¹⁰ It should be *broadly* consistent with linguistic behavior, but it need not be perfectly consistent, since we must allow for production errors and factual mistakes.

of linguistic behavior do not answer the questions we have about meaning. The main order of duty is still to start off with assessments of what terms mean, then seek the best explanations for the facts we have found. But then, if semanticists are to describe meanings as they are—good or bad; useful or impractical—we cannot endorse a metasemantics that puts a constraint on semantics to the effect that terms always mean what they ought to mean. As we noted in the previous section, this is to put the cart before the horse. We should instead be prepared to accept that the semantic facts are not always as they should be.

If we should not accept the constraint Always Revelation puts on semantics, we should not accept Always Revelation. Admittedly, this does not shut the door for weaker versions of Amelioration as Revelation. To be clear, Ball's temporal externalism comes close to being an instance of Always Revelation, since *what speakers in the future reasonably agree to* strongly selects for the meanings our expressions ought to have. The notion of a reasonable agreement is a normative notion. Even so, there is a logical gap between what future speakers *reasonably* agree to and what they *ought* to agree to (so as to optimize our lexicon). The notion of an objective kind may only be borderline normative. Thus, the corresponding logical gap is arguably greater in the case of Haslanger's and Sider's views, according to which specific sets of terms latch on to objective social and natural kinds. But whether or not objectivity is inherently normative, the notion has clear-cut normative import in Haslanger's and Sider's accounts, because they both make the plausible assumption that our terms *ought* to carve natural and social reality at their objective joints.

Always Revelation commits the mistake of forcing our semantic judgments to comply with our normative considerations about what the meaning of a term ought to be, and the weaker versions of Amelioration as Revelation make use of conceptually normative tools to identify how meaning is determined. Still, they avoid the consequence that semantic reality is *always* as it should be. Therefore, if they commit the same mistake, it is clear that they do so to a lesser extent. It would

be difficult to capture the precise extent to which these accounts constrain semantics unduly (assuming they do), partly because their metasemantic accounts are not fully spelled out and partly because the semantic assumptions that would help settle the matter may be called into dispute. Instead of embroiling ourselves in this difficult task, let us turn to the motivation for endorsing the views in question. What motivates the thought that meaning is determined by matters such as *what is reasonably agreed to in the future?* On my proposed diagnosis, recognizing that speech may be evaluated for the goodness of the meaning of the component terms—and not just by the truth of the sentences uttered—dispels the need for baking in norms of meaning selection into our metasemantics.

4. Evaluating speech

Let us again consider the example of ‘marriage’. Imagine a time in which people thought of marriage as a bond that could only hold between one man and one woman. They believed that it is not only morally unacceptable for two men to marry, but conceptually impossible. Whatever partnerships two men or two women might engage in, people at the time did not think these partnerships could qualify as marriages. For the sake of simplicity, we may suppose that every speaker of English held this belief, and that they used ‘marriage’ and its cognates accordingly. Now consider the utterance of a speaker at the time:

- 1) Two men cannot marry.

Theorists of all metasemantic persuasions should agree that the utterer of 1) is a making a mistake, or so I assume. Part of the appeal of externalism is that it can explain how and why she is making a mistake.

On internalist premises, the attitudes, beliefs and use-patterns in the speaker’s linguistic community would determine that ‘marriage’ does not apply to same sex couples. Therefore, an internalist is committed to think that the utterance of 1) was true. This seems like a bad consequence. To avoid it, we cannot remain committed

to internalism. That is, we would need to accept that meaning is determined by something beyond the past and current intentional states of speakers in the linguistic community of the utterer. For example, if meaning is partly determined by future reasonable agreement, we may resist the conclusion that the utterer of 1) spoke truly. On the assumption that we want to avoid the consequence that the utterer of 1) spoke truly, this provides support for externalism.

However, notice that the truth of 1) at the time of the utterance does not actually entail that two men cannot marry (when I spell this out without quotation marks). If the meaning of ‘marry’ has changed, as I have suggested, 1) expresses a different proposition pre-change than what it expresses post-change. If we were to utter 1) now, we would say something false. So although an internalist is committed to regard the pre-change utterance of 1) as true, she is not committed to regard current utterances of 1) as true. It follows that she is not committed to deny that two men can marry. Assessing the pre-change utterance of 1) as true is therefore less problematic than what one might think on first consideration. This is a first step. The second is more important.

One of great virtues of conceptual engineering as a methodological framework is that it allows us to evaluate speech by more sophisticated means than simply judging sentences to be true or false (and subsentential expressions as successfully referring or not). This was recognized by Carnap long ago, who viewed language selection and conceptual engineering as the primary task of philosophy, and thought of it as a pragmatic matter, as opposed to a matter of finding the truth (c.f. Creath 1990, 1–43). In contemporary writing, this line of thought is spelled out by Jennifer Nado, who argues that success in conceptual engineering should be measured by “conceptual efficacy” rather than truth (Nado 2021a). The idea advocated by Carnap and Nado is not that we should cease to evaluate speech as true or false. Once we have assigned meaning to the words of a language, truth is perhaps the most important tool we have to evaluate utterances in that language. Nonetheless, we should also evaluate utterances for the goodness of concepts

expressed. In the case of the pre-change utterance of 1) we may object that the concept of marriage expressed by the utterance is unfairly exclusive, and that same-sex couples ought to enjoy the right to marry. This is not to argue that 1) is false, but to argue that it *ought* to be false, because ‘marry’ ought to have a meaning that renders it false. Without resorting to externalism, then, one may do justice to the impression that the utterer of 1) was making a mistake. And if we take seriously the idea that there is such a thing as what terms ought to mean, there is no need to think that the mistake of using ‘marry’ with a morally harmful meaning is any less egregious than using ‘marry’ to say something false.

We are in a position to evaluate speech in a more precise way if we make normative assessments of concepts alongside traditional truth evaluations. This removes the need to bake norms of meaning selection into our metasemantics, and it provides a better framework for evaluating speech. The sorts of arguments one may give for the claim that a speaker has made use of a bad concept are different from the arguments one may give for the claim that the speaker has said something false. We had better not confuse these kinds of arguments. Instead of trying to do all our speech evaluation with descriptive assessments of meaning and alethic assessments of statements based on the meaning we have assigned, we should enrich our practice with normative assessments of meaning.

The argument I outlined in favor of externalism below 1) is ultimately unsound and nudges us to commit to an unduly impoverished way of evaluating speech. Admittedly, the traditional arguments in favor of externalism do not center on the word ‘marry’, but (perhaps most famously) on ‘water’ and ‘arthritis’ (Putnam 1975; Burge 1979). Still, the arguments based on Putnam’s Twin Earth case and Burge’s arthritis case mirror the structure of the former argument. In Putnam’s Twin Earth thought experiment, there is a planet just like ours, with people just like us, except for the fact that the watery stuff that exists on Twin Earth is not composed of H₂O. Instead, it is composed of a structurally complex molecule known as XYZ. Apart from that, the watery stuff on Twin Earth is just like water

here on Earth. We may then imagine that a speaker from before the discovery of water's chemical structure is transported to Twin Earth. Upon seeing a pond of watery liquid, she utters

2) There is water here.

The intuition we are supposed to have is that the speaker has said something false. Let us now withhold this particular judgment, and instead start off with the assumption that the speaker is *somehow* ignorant of *some* kind of fact. There are different kinds of mistakes the speaker could be making, which correspond to different kinds of corrections we could purport to give her. The first kind of correction concerns the speaker's background knowledge. We could lecture her on chemistry and inform her that the stuff she has hitherto called 'water' has a particular microstructure that does not match the microstructure of the liquid in front of her. The second kind of correction concerns her use of the word 'water' to talk about the liquid in front of her. Having taught her basic chemistry, we could try to convince her that 'water' applies to all and only samples of H_2O , and hence that what she said is false. It is not obvious how the speaker would or should respond to this second kind of correction. She might disagree, arguing that she uses 'water' for anything with the observable, nutritious and practical properties of the watery stuff back home, and hence that the watery stuff in front of her should qualify as water. After all, why should she think that 'water' was tracking properties she has never once been concerned about, as opposed to properties that matter to her and her fellow speakers every day? Remember, her earthly linguistic community has not yet begun to identify water with H_2O , so it is conceivable that this modal profile for 'water' has not yet settled in her language. The third kind of correction is to convince her that 'water' ought to have a meaning such that it applies to all and only samples of H_2O , e.g. because samples of H_2O constitute a natural kind that has semantic privilege, or because this meaning for water does good for our explanations of water-related events and states of affairs. As in the

case of ‘marry’, it is possible to argue that ‘water’ ought to have a particular meaning quite independently of what its actual meaning is.

Clearly, the case of using ‘water’ to talk about a XYZ-composed liquid raises a host of different questions. It is well beyond the scope of the present paper to answer all of them. Although I suggested in passing how one might resist the externalist verdict expressed by the second kind of correction—by outlining a response the speaker could make—my present objective is not to provide a general argument against all forms of externalism. Rather, it is to highlight that we need to take care when we evaluate these cases so that we do not assimilate the different ways one may assess the target utterance. In the Twin Earth case, we need to distinguish lack of knowledge about the chemical composition of substances from the (alleged) mistake of using ‘water’ to talk about an XYZ-composed liquid. We also need to distinguish normative assessments of meaning from descriptive assessment of meaning. When we distinguish the normative assessment from the descriptive assessment more sharply, part of the intuitive pull of externalism disappears, in particular for the strands of externalism I have grouped as Amelioration as Revelation. The case of ‘marry’ brings this out in a more dramatic way, but it is still present in the Twin Earth thought experiment: There is a reasonable case to be made for the claim that ‘water’ ought to apply to all and only samples of H_2O , and this may affect our judgments about whether the utterance is true or false.

Setting aside the question of whether we ought to endorse an externalist or internalist metasemantics, we need to take note of the effect normative assessments of meaning make on our judgment about the truth of utterances. In ordinary discourse, ‘truth’ functions as an endorsement device. It requires both metalinguistic awareness and appreciation for the fact that there is such a thing as what a term ought to mean in order to see clearly that an ascription of truth need not entail endorsement of using the target terms with the target meanings. In the absence of metalinguistic awareness and awareness of conceptual ethics, it is

unsurprising that we are drawn to judge that an utterance is false when it contains a term that expresses something it ought not to. Since judgments about the truth of utterances are the central form of motivation for externalist views, it is unsurprising that some have thought that our languages have metasemantics with built-in optimization- or amelioration mechanisms, as is common to the views that fall in the Amelioration as Revelation category. However, normative assessments of meaning should be kept apart from our assessments of truth. When we can already explain what we need to explain without postulating mysterious determinants of meaning—such as events in the future or epistemically occluded objective kinds in the world—we have no good reason to think meaning is determined accordingly.

6. Conclusion

Metasemantic views of the kind I have called Amelioration as Revelation attempt to explain descriptive facts of meaning on normative grounds. The intuitive pull of such views may be ascribed to our lacking awareness (in philosophy and beyond) of normative assessments of meaning, and perhaps to a lack of metalinguistic carefulness. In attempting to bake normative facts about meaning into descriptive facts about meaning, the views impoverish our speech-assessment practices. They also threaten to weaken our ability to improve our conceptual repertoire insofar as they fudge the difference between normative and descriptive claims about meaning. In turn, this fudges the difference between the sorts of arguments one may advance in favor of normative and descriptive claims about meaning. There is a resurgence of interest in normative thinking about concepts and meaning in philosophy, evinced in the fast-growing literature on conceptual engineering. This resurgence promises to better equip us to improve meaning and better understand what we are doing when we try to change meaning. If my arguments hold true, Amelioration as Revelation is both an implausible view of what happens in

exercises in conceptual engineering and detrimental to our capacity to do conceptual engineering well.

References

- Armstrong, Josh. 2016. "The Problem of Lexical Innovation." *Linguistics and Philosophy* 39 (2): 87–118.
- Austin, J. L. 1956. "A Plea for Excuses: The Presidential Address." *Proceedings of the Aristotelian Society* 57: 1–30.
- Ball, Derek. 2020. "Revisionary Analysis without Meaning Change (Or, Could Women Be Analytically Oppressed?)." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett, 35–58. Oxford: Oxford University Press.
- Blank, Andreas. 1999. "Why Do New Meanings Occur? A Cognitive Typology of the Motivations for Lexical Semantic Change." In *Historical Semantics and Cognition*, edited by Andreas Blank and Peter Koch. Berlin: Mouton de Gruyter.
- Brandom, Robert. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass: Harvard University Press.
- . 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass: Harvard University Press.
- Brigandt, Ingo. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177 (1): 19–40. <https://doi.org/10.1007/s11229-009-9623-8>.
- Brun, Georg. 2016. "Explication as a Method of Conceptual Re-Engineering." *Erkenntnis* 81 (6): 1211–41. <https://doi.org/10.1007/s10670-015-9791-5>.
- Burge, Tyler. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1): 73–122.
- Burgess, Alexis, and David Plunkett. 2013a. "Conceptual Ethics I." *Philosophy Compass* 8 (12): 1091–1101.
- . 2013b. "Conceptual Ethics II." *Philosophy Compass* 8 (12): 1102–10.
- . 2020. "On the Relation between Conceptual Engineering and Conceptual Ethics." *Ratio (Oxford)* 33 (4): 281–94. <https://doi.org/10.1111/rati.12265>.
- Burgess, Alexis, and Brett Sherman. 2014. "Introduction: A Plea for the Metaphysics of Meaning." In *Metasemantics: New Essays on the Foundations of Meaning*, 1–15. Oxford: Oxford University Press.

- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- . 2020. “Conceptual Engineering: The Master Argument.” In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett, 132–51. Oxford: Oxford University Press.
- Cappelen, Herman, and David Plunkett. 2020. “Introduction - A Guided Tour of Conceptual Engineering and Conceptual Ethics.” In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett. Oxford: Oxford University Press.
- Carnap, Rudolf. (1934) 1937. *The Logical Syntax of Language*. London: Kegan Paul, Trench, Trubner & co. Ltd.
- . 1950. “Empiricism, Semantics, and Ontology.” *Revue Internationale de Philosophie* 4 (11): 20–40.
- . 1956. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd ed. Chicago: University of Chicago Press.
- . 1962. *Logical Foundations of Probability*. 2nd ed. Chicago: University of Chicago Press.
- . 1963a. “Intellectual Autobiography.” In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 3--84. Open Court.
- . 1963b. “Replies and Systematic Expositions.” In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 859–1013. La Salle, Ill: Open Court.
- . (1928) 1967. *The Logical Structure of the World*. Der Logische Aufbau Der Welt. London: Routledge & Kegan Paul.
- Chalmers, David. 2011. “Verbal Disputes.” *Philosophical Review* 120 (4): 515–66.
- . 2020. “What Is Conceptual Engineering and What Should It Be?” *Inquiry*. <https://doi.org/10.1080/0020174x.2020.1817141>.
- Craig, Edward. 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Clarendon.
- Creath, Richard. 1990. *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work: Edited and with an Introduction by Richard Creath*. Berkeley, Calif.: University of California Press.
- Cummins, Robert. 1975. “Functional Analysis.” *The Journal of Philosophy* 72 (20): 741–65. <https://doi.org/10.2307/2024640>.

- Davidson, Donald. 1986. "A Nice Derangement of Epitaphs." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest Lepore, 433–46. Oxford: Blackwell.
- . (1984) 2001. "In Defence of Convention T." In *Inquiries into Truth and Interpretation*, 2nd ed., 65–79. Oxford: Oxford University Press.
- Deutsch, Max. 2020. "Speaker's Reference, Stipulation, and a Dilemma for Conceptual Engineers." *Philosophical Studies* 177 (12): 3935–57.
<https://doi.org/10.1007/s11098-020-01416-z>.
- Dever, Josh. 2020. "Outer Limits of Conceptual Engineering." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett, 152–69. Oxford: Oxford University Press.
- Devitt, Michael. 1981. *Designation*. New York: Columbia University Press.
- Dorr, Cian, and John Hawthorne. 2014. "Semantic Plasticity and Speech Reports." *Philosophical Review* 123 (3): 281–338.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. London: Duckworth.
- . 1986. "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest LePore. Oxford: Blackwell.
- Dutilh Novaes, Catarina, and Erich Reck. 2017. "Carnapian Explication, Formalisms as Cognitive Tools, and the Paradox of Adequate Formalization." *Synthese* 194 (1): 195–215. <https://doi.org/10.1007/s11229-015-0816-z>.
- Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford: Oxford University Press.
- Foucault, Michel. 1973. *Madness and Civilization: A History of Insanity in the Age of Reason*. Folie et Dêraison : Histoire de La Folie à l'âge Classique. New York: Vintage.
- Frege, Gottlob. (1879) 1972. *Conceptual Notation: And Related Articles*. Oxford Scholarly Classics. Oxford: Clarendon Press.
- . 1979. *Posthumous Writings*. Oxford: Blackwell.
- Friedman, Michael. 1987. "Carnap's Aufbau Reconsidered." *Noûs* 21 (4): 521–45.
- Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121–23.
- Grice, Paul. 1989. *Studies in the Way of Words*. Cambridge, Mass: Harvard University Press.

- Gutting, Gary. 1998. “‘Rethinking Intuition’: A Historical and Metaphilosophical Introduction.” In *Rethinking Intuition - The Psychology of Intuition and Its Role in Philosophical Inquiry*, edited by Michael R. DePaul and William Ramsey, 3–13. Lanham, MD: Rowman & Littlefield.
- Hannon, Michael. 2019. *What’s the Point of Knowledge? A Function-First Epistemology*. Oxford: Oxford University Press.
- Haslanger, Sally. 1999. “What Knowledge Is and What It Ought to Be: Feminist Values and Normative Epistemology.” *Philosophical Perspectives* 13: 459–80.
- . 2000. “Gender and Race: (What) Are They? (What) Do We Want Them To Be?” *Noûs* 34 (1): 31–55. <https://doi.org/10.1111/0029-4624.00201>.
- . 2006. “What Good Are Our Intuitions? - Philosophical Analysis and Social Kinds.” *Aristotelian Society Supplementary Volume* 80 (1): 89–118.
- . 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
<https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199892631.001.0001/acprof-9780199892631>.
- . 2020. “Going On, Not in the Same Way.” In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett, 230–60. Oxford: Oxford University Press.
- Heidegger, Martin. (1927) 1962. *Being and Time*. Sein Und Zeit. Oxford: Basil Blackwell.
- Isaac, Manuel Gustavo. 2020. “How to Conceptually Engineer Conceptual Engineering?” *Inquiry*, 1–24. <https://doi.org/10.1080/0020174X.2020.1719881>.
- Jorem, Sigurd. 2021. “Conceptual Engineering and the Implementation Problem.” *Inquiry* 64 (1–2): 186–211. <https://doi.org/10.1080/0020174X.2020.1809514>.
- Kahneman, Daniel, Amos Tversky, and Paul Slovic. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.
- Kant, Immanuel. (1781–1787) 1998. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W Wood. Cambridge: Cambridge University Press.
- King, Jeffrey C., and Jason Stanley. 2005. “Semantics, Pragmatics, and the Role of Semantic Content.” In *Semantics Versus Pragmatics*, edited by Zoltan Gendler Szabo, 111–64. Oxford: Oxford University Press.
- Koch, Steffen. 2020. “Engineering What? On Concepts in Conceptual Engineering.” *Synthese*. <https://doi.org/10.1007/s11229-020-02868-w>.

- . 2021a. “The Externalist Challenge to Conceptual Engineering.” *Synthese* 198 (1): 327–48. <https://doi.org/10.1007/s11229-018-02007-6>.
- . 2021b. “Why Conceptual Engineers Should Not Worry About Topics.” *Erkenntnis*. <https://doi.org/10.1007/s10670-021-00446-1>.
- Koslow, Allison Robbins. 2020. “Meaning Change, in Theory and in Practice.” Doctoral dissertation, Massachusetts Institute of Technology. <https://dspace.mit.edu/bitstream/handle/1721.1/130209/1241254352-MIT.pdf?sequence=1&isAllowed=y>.
- Kripke, Saul A. 1980. *Naming and Necessity*. Rev. and Enlarged ed. Library of Philosophy and Logic. Oxford: Blackwell.
- Leitgeb, Hannes, and André Carus. 2020. “Rudolf Carnap.” In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021 Edition. <https://plato.stanford.edu/archives/sum2021/entries/carnap/>.
- Lewis, Clarence Irving. 1929. *Mind and the World-Order: Outline of a Theory of Knowledge*. London: Charles Scribner.
- Lewis, David. (1975) 1983. “Languages and Language.” In *Philosophical Papers Volume I*, 163–88. Oxford: Oxford University Press.
- . 1984. “Putnam’s Paradox.” *Australasian Journal of Philosophy* 62 (3): 221–36.
- Löhr, Guido. 2021. “Commitment Engineering: Conceptual Engineering without Representations.” *Synthese*. <https://doi.org/10.1007/s11229-021-03365-4>.
- Ludlow, Peter. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford: Oxford University Press.
- Machery, Edouard. 2009. *Doing Without Concepts*. Oxford: Oxford University Press.
- . 2017. *Philosophy Within Its Proper Bounds*. Oxford: Oxford University Press.
- Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, Mass: MIT Press.
- . 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass: MIT Press.
- Nado, Jennifer. 2020. “Taking Control: Conceptual Engineering without (Much) Metasemantics.” *Inquiry*, 1–27. <https://doi.org/10.1080/0020174X.2020.1850342>.

- . 2021a. “Conceptual Engineering, Truth, and Efficacy.” *Synthese* 198: 1507–27. <https://doi.org/10.1007/s11229-019-02096-x>.
- . 2021b. “Conceptual Engineering via Experimental Philosophy.” *Inquiry* 64 (1–2): 76–96. <https://doi.org/10.1080/0020174X.2019.1667870>.
- . MS. “Strawson’s Challenge and the Problem of Multiple Successors.” *Unpublished Manuscript*.
- Nietzsche, Friedrich. (1901) 1968. *The Will to Power*. Translated by Walter Arnold Kaufmann and R.J Hollingdale. Der Wille Zur Macht. New York: Vintage Books.
- O’Grady, Paul. 1999. “Carnap and Two Dogmas of Empiricism.” *Philosophy and Phenomenological Research* 59 (4): 1015–27. <https://doi.org/10.2307/2653567>.
- Pinder, Mark. 2021. “Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning.” *Mind* 130 (517): 141–63.
- Plunkett, David. 2015. “Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy.” *Inquiry* 58 (7–8): 828–74. <https://doi.org/10.1080/0020174X.2015.1080184>.
- Plunkett, David, and Tim Sundell. 2013. “Disagreement and the Semantics of Normative and Evaluative Terms.” *Philosopher’s Imprint* 13 (23): 1–37.
- Price, Huw. 2011. *Naturalism Without Mirrors*. Oxford: Oxford University Press.
- Prinzing, Michael. 2018. “The Revisionist’s Rubric: Conceptual Engineering and the Discontinuity Objection.” *Inquiry* 61 (8): 854–80. <https://doi.org/10.1080/0020174X.2017.1385522>.
- Putnam, Hillary. 1975. “The Meaning of ‘Meaning.’” *Minnesota Studies in the Philosophy of Science* 7: 131–93.
- Queloz, Matthieu. 2021. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198868705.001.0001>.
- Quine, Willard V. O. 1951. “Two Dogmas of Empiricism.” *Philosophical Review* 60 (1): 20–43.
- . 1963. “Carnap and Logical Truth.” In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 385–406. La Salle, Ill: Open Court.
- Recanati, François. 2010. *Truth-Conditional Pragmatics*. Oxford: Oxford University Press.

- Riggs, Jared. 2019. "Conceptual Engineers Shouldn't Worry about Semantic Externalism." *Inquiry*, 1–22. <https://doi.org/10.1080/0020174X.2019.1675534>.
- . 2021. "Deflating the Functional Turn in Conceptual Engineering." *Synthese*. <https://doi.org/10.1007/s11229-021-03302-5>.
- Rosch, Eleanor, and Carolyn B Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7 (4): 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9).
- Sawyer, Sarah. 2020. "Truth and Objectivity in Conceptual Engineering." *Inquiry* 63 (9–10): 1001–22. <https://doi.org/10.1080/0020174X.2020.1805708>.
- Scharp, Kevin. 2013. *Replacing Truth*. Oxford: Oxford University Press.
- . 2020. "Philosophy as the Study of Defective Concepts." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett. 396–416. Oxford: Oxford University Press. <https://www.oxfordscholarship.com/view/10.1093/oso/9780198801856.001.0001/oso-9780198801856-chapter-19>.
- Schiappa, Edward. 2003. *Defining Reality: Definitions and the Politics of Meaning*. Carbondale, IL: Southern Illinois University Press.
- Schroeter, Laura, and François Schroeter. 2020. "Inscrutability and Its Discontents." *Canadian Journal of Philosophy* 50 (5): 566–79.
- Searle, John. 1996. "Contemporary Philosophy in the United States." In *The Blackwell Companion to Philosophy*, edited by Nicholas Bunnin and E. P Tsui-James, 1–22. Oxford: Blackwell Publishers.
- Sellars, Wilfrid. 1953. "Inference and Meaning." *Mind* 62 (247): 313–38.
- . 1997. *Empiricism and the Philosophy of Mind, with an Introduction by Richard Rorty and a Study Guide by Robert Brandom*. Cambridge, Mass: Harvard University Press.
- Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Clarendon Press.
- Silberstang, Edwin. 1972. "Rape." In *American Jurisprudence*, 2nd ed., 65:757–838. Rochester, NY: Lawyers Cooperative.
- Simion, Mona. 2018. "The 'should' in Conceptual Engineering." *Inquiry* 61 (8): 914–28. <https://doi.org/10.1080/0020174X.2017.1392894>.
- Simion, Mona, and Christoph Kelp. 2020. "Conceptual Innovation, Function First." *Nous* 54 (4): 985–1002. <https://doi.org/10.1111/nous.12302>.

- Strawson, Peter F. 1963. "Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 503–18. La Salle, Ill: Open Court.
- Thomasson, Amie L. 2020a. "A Pragmatic Method for Normative Conceptual Work." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett, 435–58. Oxford: Oxford University Press.
- . 2020b. *Norms and Necessity*. New York: Oxford University Press.
<https://doi.org/10.1093/oso/9780190098193.001.0001>.
- . MS. "Conceptual Engineering: When Do We Need It? How Can We Do It?" *To Appear in a Special Issue of Inquiry*.
- Weatherson, Brian. 2003. "What Good Are Counterexamples?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 115 (1): 1–31.
- Weinberg, Jonathan M. 2006. "What's Epistemology For? The Case for Neopragmatism in Normative Metaepistemology." In *Epistemological Futures*, edited by Stephen Hetherington, 26–47. Oxford: Clarendon Press.
- Weinberg, Jonathan M, Shaun Nichols, and Stephen Stich. 2001. "Normativity and Epistemic Intuitions." *Philosophical Topics* 29 (1/2): 429–60.
<https://doi.org/10.5840/philtopics2001291/217>.
- Williams, J. Robert G. 2007. "Eligibility and Inscrutability." *The Philosophical Review* 116 (3): 361–99. <https://doi.org/10.1215/00318108-2007-002>.
- Williamson, Timothy. 2007. *The Philosophy of Philosophy*. The Blackwell/Brown Lectures in Philosophy. Malden, Mass: Blackwell.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Edited by G. E. M. Anscombe and G. H. von Wright. Translated by G. E. M. Anscombe. Oxford: Blackwell.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Wright, Larry. 1973. "Functions." *The Philosophical Review* 82 (2): 139–68.
<https://doi.org/10.2307/2183766>.

