Possible Limits of Conceptual Engineering: Magnetism, Fixed Points and Inescapability

Matti Eklund

matti.eklund@filosofi.uu.se

1. Introduction

Wilfrid Sellars famously contrasts the manifest image and the scientific image.[1] Roughly speaking, the manifest image is the image of the world of common sense and everyday life, and the scientific image is the image presented to us by science. There is a prima facie conflict between these two images, and Sellars, as well as many other philosophers, has seen the question of how to reconcile these images as a central task for philosophy. In discussions of these issues, there is much focus on the specific image of the world as presented by science and how it relates to the manifest image. But one can, and I will, ask more general questions about how the common sense image of the world relates to the image of the world *after systematic reflection and investigation*, whether as part of science properly so-called or not. And while the two "images" are naturally thought of as (in part although not exclusively) being two different theories and the issue is naturally framed as (in part) one of how to reconcile two theories, I will here focus on a different but related issue. Theories employ *concepts*. One can then ask probing questions about the relationship between the concepts that are used in the theory associated with the common sense, manifest image, and the concepts used in theory associated with the scientific image. In principle, two conflicting theories can use the same concepts – two theories may employ the same concept but simply say different

---

[1] Sellars (1962).

things about what falls under it – but it is natural from many perspectives to think that the theories do not employ exactly the same concepts. Some concepts used in the latter theory may be of a technical character, and will not be found in the former. There are general questions that can be asked about the relationship between what concepts the two kinds of theories use. What I will focus on here is this. We begin inquiry with the manifest image, and the concepts employed by the theory associated with that image. As we revise the theory we start with, and develop a scientific image of some sort or other, are all concepts we start with in principle candidates for replacement or revision, or are some concepts somehow immune to that?

There are many different ways to approach this sort of broad question, and also many ways to sharpen it. One can see Kant's philosophy as a systematic argument for the claim that some concepts are immune to revision. From a more naturalistic perspective one can speculate that some concepts are innate and for that reason immune to revision. From a much more practical perspective one can envisage arguments for why some or many concepts are simply very hard to revise. I will not attempt to consider these approaches. What I will aim to do here is to discuss theoretical rather than practical arguments regarding immunity to revision, while staying away from arguments whose success depends on very general theoretical commitments, like those found in Kant.

In the next section, I will introduce the issues just mentioned in a different way, by a discussion of what has come to be known as *conceptual engineering*.

2. Conceptual engineering

In the contemporary philosophical discussion, there is much talk of so-called conceptual engineering.[2] I will not here attempt an overview of all the various things that are discussed under that heading. Instead let me just emphasize what conceptual engineering, as I think of it, is, and what motivates it. The basic idea is simple. Sometimes the concept(s) we employ for a given purpose are not the ideal concepts to employ for that given purpose. There is a worthwhile project of figuring out which concept(s) are better for the purpose in question, and of going on to employ those concepts (and, one might add, of getting others to employ those concepts). It is this project I think of as conceptual engineering. Note the generality of conceptual engineering as described. It can occur anywhere in science, anywhere in philosophy – and also outside of any theoretical endeavour.

Some quick remarks may be in order, clarifying what conceptual engineering is meant to be. There is first a question about what concepts are. The label "concept", widely used, tends to be ambiguous. Sometimes it stands for things meant; sometimes it stands for some sort of mental counterpart of words. In the latter case, concepts do not have meanings but rather are meanings. Either understanding is fine so long as there is no unclarity. Understanding "concept" the first way, what conceptual engineering is focused on is revising concepts themselves. Understanding "concept" the second way, conceptual engineering is focused on revising the *meanings of* concepts. One could even eschew any talk of concepts and instead speak of, for example, *meanings* or *words* or, maybe, *representational devices*. Second, for many purposes, one could avoid all talk of *representations*, and instead focus on *what is represented*. ("In this kind of inquiry, should we focus on these entities or those? Should we focus on this property or that?") But talk of concepts and words is still sometimes

---

[2] The literature on conceptual engineering is already extensive. Two books devoted to the topic are Cappelen (2018) and the edited volume Burgess, Cappelen and Plunkett (2020). Relevant overviews include Cappelen and Plunkett (2020), Eklund (2021), and Chalmers (forthcoming).

useful, for sometimes *how* we represent something matters; and some thought and talk is not representational. Third, it is natural to gloss conceptual engineering as a matter of fixing individual concepts. Sometimes the fix is to introduce entirely new concepts, or to drop an old one; it is not just a matter of revising some existing concepts. It may be better to speak of (say) fixing systems of concepts.

Conceptual engineering is sometimes discussed as if it is revisionary of standard philosophical methodology: as if it amounts to a new way of doing philosophy. But to what extent is it actually so? Here's a pretty orthodox answer: Some engineering has obviously already been going on in philosophy. Sometimes this has also been explicitly discussed. Such discussions can for example be found in philosophers as otherwise different as Nietzsche, Heidegger, Carnap, and Quine.[3] I think the recent systematic attention to conceptual engineering actually is to some extent novel, but any claim to the effect that conceptual engineering is revisionary of current methodology in philosophy would have to be somewhat guarded.

To me, conceptual engineering as briefly motivated above is clearly a worthwhile enterprise. What sorts of reasons can there be not to engage in conceptual engineering? It would seem to be outright silly to deny that we should strive to have concepts that serve our purposes as well as possible. And it would seem silly to insist that our actual concepts already do serve our purposes as well as any concepts can.

Some of the most central topics in the conceptual engineering literature are the following. (i) What is meant by *concept* here?[4] (ii) Is it one and the same concept that changes or is the old concept replaced by a new one? If it is sometimes the one thing and sometimes the other, when do we have to do with mere change and when is it a matter of

---

[3] For a helpful overview of the pre-history of contemporary conceptual engineering, see Jorem (2021), ch. 1.
[4] See e.g. Cappelen (2018), Koch (2021), and Isaac (forthcoming).

replacement?[5] (iii) Do we in some sense change the topic when revising/replacing the old concept? If, say, we are having a discussion of free will, and someone starts using a new concept FREE WILL* instead of the ordinary concept FREE WILL, is she then changing the topic, and if she is, is that a problem?[6] (iv) Can the proposed revisions be implemented, and, if so, how is this best achieved?[7] (v) When should the old *word* be retained for the new or revised concept?[8]

The main thing I want to stress is that none of these issues is pressing for conceptual engineering as a project. It may be important to be clear on what we mean by "concept", and on what sorts of changes a concept may undergo and still be the same concept, but regardless of what the answers to these questions may be, the motivation stands, so (i) and (ii) are in that way not pressing. Similarly, the motivation stands whatever we should say about when we should retain the old word (so (v) is in that way not pressing). The issues that may be thought to correspond to threats to conceptual engineering are (iii) and (iv). Consider first (iv), and suppose that the answer to question (iv) is the one that threatens to be the most worrisome for the conceptual engineer: that implementation is impossible. In some obvious ways, this limits the practical applicability of conceptual engineering. But the conceptual engineer's assessment of concepts is still relevant. The conceptual engineer might still say: such-and-such possible concepts would be better in thus-and-such ways than our actual concepts. And such truths can still be important even if we cannot come to adopt the other possible concepts. In general, knowing the limits of the tools I have is important even if I am unable to get better tools. If I have constructed artificial wings for flying, it is good for me to know their deficiencies before I take off from the roof, even if constructing better wings is not feasible. Issue (iii) could suggest a principled objection against conceptual engineering if changing the

---

[5] See e.g. Nado (forthcoming) and Richard (2020).
[6] See e.g. Cappelen (2018), Sundell (2020), Pinder (2021) and Belleri (forthcoming).
[7] See e.g. Cappelen (2018), Jorem (2021a), Pinder (2021), and Nimtz (forthcoming).
[8] For some relevant discussion see Chalmers (forthcoming) and Sterken (2020).

topic was somehow always bad. But there is no reason to assume that. However exactly we construe "topics", some topics may be such that they should be abandoned in favor of improved topics.

This does not mean that questions (i)-(v) need to be without interest. All I say is that the motivation for conceptual engineering as I have described it here remains unscathed regardless of how these questions are answered.

What principled objections to conceptual engineering could there be? One kind of principled objection might be a Wittgenstein-inspired objection to the effect that philosophy should "leave everything as it is" to use the common phrase here.[9] Let me say a few words about that. First, what could reasonably underlie this Wittgensteinian injunction? One thought, with recognizable Wittgensteinian pedigree, might be that philosophical problems never arise because of defects in language, but insofar as the problems have to do with language at all, what is needed to deal with the problems is only a clear view of how language works. But if that is what the thought amounts to, it is not incompatible with conceptual engineering as motivated here. One can perfectly consistently hold both that philosophical problems do not arise because of defects in language and that the expressions and concepts we actually employ are not the best ones for the purposes to which they are put. Now, second, Wittgenstein does say "Philosophy may in no way interfere with the actual use of language; it can in the end only describe it".[10] Taken at face value, this goes beyond any claim specifically about what underlies philosophical problems. It explicitly prohibits philosophers from attempting to reform language use, and the "only describe" part further seems to prohibit philosophers from even evaluating the expressions and concepts we use. I have a hard time seeing what might underlie this stronger claim. Does Wittgenstein mean merely to say something about what philosophy and philosophers may and may not do, or does he more

---

[9] Compare *Philosophical Investigations*, §124.
[10] See again Wittgenstein's *Philosophical Investigations*, §124.

generally issue a prohibition against interfering with, and evaluating, language? The formulation certainly suggests the former. If the claim is only the former, the question is what renders philosophy or philosophers especially unsuitable for interfering and evaluating. If the claim is the latter, it is worth noting that the latter claim is very strong, and implausibly so.

In general, I think there are are no objections to the general project of conceptual engineering to worry about. What work remains regarding conceptual engineering? Pretty much everything regarding the details, of course. Which concepts should be revised, and why? But what I want focus on is the question: what *principled* work remains? Here is one general question: What are the *principled limits*, if any, to what conceptual engineering is possible? Even when general concerns about the project of conceptual engineering have been set aside, one may wonder whether there are some significant limits. Here are two kinds of questions one might ask. First, what kinds of concepts can there be? Limits on what concepts there can be are limits of conceptual engineering. Second, there are questions about what kinds of revisions of the concepts we have are in principle possible. I will here focus on the second of these questions. The implementation problem, mentioned above, concerns practical obstacles, albeit of a general nature. I will here focus on theoretical, principled limitations.

This question about the limits of conceptual engineering is in essence the same question as I described in the last section, concerning the relationship between the concepts employed in the theories associated with the pre-reflective theory of the world and theories of the world arrived at by systematic reflection and investigation. It goes back to the relationship between Sellars' two images of the world, the manifest image and the scientific image.

In what follows I will discuss possible limits to conceptual engineering, in three ways. First, I will briefly discuss reference magnetism. Second, I will relate to Herman Cappelen's (2018) recent discussion of issues in this vicinity, where Cappelen brings up arguments due to David Chalmers and Matti Eklund. Third, I will discuss whether some concepts are

inescapable, in something like the sense recently characterized by Thomas Hofweber, where a concept is inescapable for a thinker if the thinker has the concept and cannot rationally revise or replace it.[11]

3. Reference magnetism

Some may think the phenomenon of *reference magnetism*, famously stressed by David Lewis, is relevant to the questions that have been raised. [12] The phenomenon of reference magnetism is the phenomenon that some entities somehow are intrinsically more eligible to be meant than others. If facts about some community's use of a given sign leave open both that the sign means F and that the sign means G, then if F is more reference-magnetic, the sign means F rather than G.

Assuming that this phenomenon is real, something which of course can be questioned, then one may think that it is a relevant limit to conceptual engineering if some contents are more reference-magnetic than others. However, I will be brief about this. As already Lewis stressed, use can trump eligibility.[13] Sometimes we can use GRUE rather than GREEN, even if GREEN is more magnetic than GRUE. There may still in principle be the practical problem of how to manage to use an expression with a less eligible meaning. But first, and most importantly, that is a mere practical problem and does not present any kind of principled obstacle. Second, when replacing pre-theoretical concepts by more apt ones, we arguably go from less eligible to more eligible meanings not the other way around.

---

[11] Hofweber discusses inescapability in his (forthcoming) and (manuscript). He discusses similar themes, although not under that label, in his (2021).
[12] See Lewis (1983, 1984).
[13] Lewis (1983), p. 372.

4. Earlier discussions about principled limits to conceptual engineering

Towards the end of Herman Cappelen's (2018) book on conceptual engineering, something like the issue of the kinds of limits that I have described comes up. Cappelen says:

> Are some *terms so basic that they cannot be engineered*? Are some *terms so fundamental that we are stuck with them*—so basic that evolution, revision, and amelioration are impossible? […] You could argue for such views in various ways. One line of thought has it that, as a matter of empirical fact, there are certain meanings that we're born with and just can't get rid of—they're stuck in our brains and however much we try, they remain there. Another line of thought has it that there are specific features of specific meanings that make the idea of evolution problematic. In what follows I will focus on this second line of thought and I'll consider two arguments—one from Chalmers and one from Eklund. My conclusion will be that we should stick with the working hypothesis that *everything is in flux—that all representational devices can be revised, and there's no natural end point to conceptual engineering*.[14]

Cappelen presents Chalmers and Eklund as arguing that there are certain principled limits to conceptual engineering, and the italicized part presents Cappelen's response to the arguments. In what follows I will argue that the arguments from Chalmers and Eklund do not even purport to show that, in Cappelen's words, there is no natural end point to conceptual engineering. They do not even purport to contradict what Cappelen says. If there are real questions about Cappelen's view that everything is in flux, they lie elsewhere. If Chalmers' and Eklund's discussions are relevant, they are so at most indirectly.

---

[14] Cappelen (2018), p. 194.

Here, first, is the argument from Chalmers that Cappelen discusses. Central to Chalmers (2011) is the topic of verbal disputes and the following proposed strategy for deciding whether a given dispute is verbal (this is Chalmers' "method of elimination"): state the dispute with the key term replaced by other terms, and see if the dispute remains. If the dispute goes away, it was verbal. If it remains, the dispute wasn't verbal. Chalmers thinks that applying this strategy shows how central disputes about, e.g., what semantics is or what physicalism is are merely verbal. If two theorists have a dispute over what physicalism is, but when "physicalism" as each uses it is replaced by the theorists' respective understandings of it the dispute goes away, the dispute was merely verbal – it simply concerns how to use the label "physicalism". But then Chalmers goes on to note an apparent limitation of this strategy: in some cases it seems that the dispute cannot be faithfully restated without the use of the key term. He considers the possibility of barring "ought" and introducing "ought$_1$" and "ought$_2$" instead.[15] Here the outcome is different:

> In the case of 'semantics', 'physicalism', and so on, this situation suggested a verbal dispute. Should we likewise diagnose a verbal dispute here? Intuitively, the answer is no. For all we have said, moral disputes are substantive disputes. Instead, we have simply exhausted the relevant vocabulary. It appears that at a certain point (perhaps once we have fixed on the appropriate moral "ought"), we have reached bedrock: a substantive dispute involving a concept so basic that there is no hope of clarifying the dispute in more basic terms.[16]

Introducing some terminology that Chalmers uses, a *bedrock dispute* is "a substantive dispute for which no underlying dispute can be found by the method of elimination", and *bedrock*

---

[15] Chalmers (2011), p. 543.
[16] Chalmers (2011), p. 543.

expressions are expressions such that "some disputes are bedrock with respect to those expressions", where "a dispute is bedrock with respect to E when the dispute is substantive and there is no dispute not involving E (and analogous in other respects) that underlies the original dispute".[17]

Commenting on Chalmers, Cappelen says that in Chalmers' terminology, bedrock expressions are "expressions that cannot be engineered"; they are "conceptual foundations where there's no option of moving to a *neighboring property*" (my emphasis).[18] He goes on to claim that there are no bedrock expressions. There are two problems with how Cappelen discusses this, and the problems illustrate how what Chalmers discuses is not even a threat to conceptual engineering in the first place. First, if E is a bedrock expression then one cannot eliminate the expression without a sense of loss: the idea is that one cannot faithfully state what was at issue in theses involving E when E is not used. But even if that is so, it can be that E should be eliminated or revised. After all, the loss in question may be a loss we can and should live with. In other words, one can still rationally conceptually engineer things. Second, Cappelen misstates things when talking about *properties*. Chalmers is only talking about replacing one *expression* by another. But the expressions involved – the old expressions and the new expressions – express the same concept and ascribe the same property: the disputants' "ought$_1$ and "ought$_2$" are not supposed to mean anything different from their original "ought"'s. This points to a reason why Chalmers' discussion is actually irrelevant as far as the possible limits of conceptual engineering are concerned. He is not really talking about replacements of *concepts*. While there is something intuitive about the phenomenon that Chalmers describes, his discussion only concerns which expression gets attached to which concept.

---

[17] Chalmers (2011), pp. 545-6.
[18] Cappelen (2018), p. 194.

Let me now turn to Eklund (2015). Eklund argued there that it may be that there in a sense are no alternatives to certain basic normative concepts, and to the notions of truth and existence. He used the label *conceptual fixed points* for the phenomenon he took himself to identify. It would take us on too much of a detour to describe the different cases in any detail, but let me very, very briefly describe what Eklund argued. He focused on three cases, the concept TRUTH, the concept EXISTS, and normative concepts like OUGHT.

First, when it comes to TRUTH, one question is what it might even be for some people to to employ a concept which plays the role of our concept TRUTH – being the aim of assertion and belief, the goal of inquiry, etc. – but still is not coextensive with it. Is this even possible? Seeming examples seem instead to involve people having attitudes toward different contents. Just to provide a simple illustration, if we encountered a community whose members kept saying "not p" in cases where it seems reasonable to believe that p, it is better to interpret them as meaning something other than we do by "not p" than it is to interpret them as aiming to have something other than truth as the aim of assertion. Second, in the case of EXISTENCE, a point is that differences in what EXISTENCE-like concept a community uses has ripple effects for the rest of that community's system of concepts. Among authors who speak of the possibility of different EXISTENCE-like concepts it is common to hold that what makes something an EXISTENCE-like concept in the first place is that it obeys the same inference rules as the existential quantifier, the standard kinds of introduction- and elimination-rules. But then there cannot be languages employing different EXISTENCE-like concepts but whose *atomic* sentences have the same truth-values. The differences which were supposed to have to do specifically with the existence concepts then spill over into many sentences that do not explicitly concern existence. Third, OUGHT. Suppose first there are different OUGHT-like concepts, in the following sense: there are different concepts which have the same normative role as OUGHT has but which are not coextensive. Then it seems that in some sense there is a

live, practical question of what OUGHT-like concept to use when guiding action. But that question can easily seem weird and awkward – even ineffable. For how are we to raise it properly? For it may well be that if the question is which concept we ought to use when guiding action, then the answer is OUGHT, but if the question is which concept we ought* to use (when OUGHT* is one of these alternatives), the answer is OUGHT*. But that is no help! One intuitively attractive way to get around with these issues is to deny the supposition mentioned and say that there aren't different ought-like concepts in the sense characterized: all concepts associated with the same normative role as ought are coextensive. Then OUGHT cannot be replaced by a different, non-coextensive concept with the same role: there are no such concepts.

The three cases are different. But *roughly*, a common theme is that there are certain *roles* that (for a thinker) can be played only by one concept. I am quick about the arguments, for my present concern isn't with how good they are but with what they purport to show.

Cappelen brings up Eklund's arguments alongside those of Chalmers as arguments to the effect that some concepts cannot be engineered. But obviously, even given the complete success of these earlier arguments of Eklund's, there are still questions about whether to employ a TRUTH-like concept at all, whether to employ an EXISTENCE-like concept at all and whether to employ any OUGHT-like concept at all. Maybe one should, but the arguments given by Eklund don't address the issue. They are only to the effect that for a thinker with a concept playing a given role, there are in a sense (in fact different senses) no alternatives.

5. Inescapability

What I have done so far is the following. I have introduced the general idea of principled limits to concept revision, or conceptual engineering. I have then brought up some earlier discussions from the literature which might be thought relevant to this issue. But these

discussions have turned out to be somewhat less relevant to the issue of concept revision than one might have thought. In this and the following sections I will consider a seemingly more promising suggestion for making progress on the issue, appealing to the notion of *inescapability*, discussed in recent work by Thomas Hofweber.[19] However, I will bring up problems and complications also regarding this suggestion.

At a first stab, a concept that a thinker has is inescapable for that thinker if and only if she cannot rationally revise or replace the concept. Compare,as an illustration, the stronger notion of *inevitability*: a concept is inevitable if every thinker must have it.[20] There are possible general reasons for doubting that any concept can be inevitable. What about possible brain injuries, or possible kinds of brain surgery? Whatever in the end to say, one can see how appeal to such things can provide an argument against the possibility of inevitable concepts. But even if, for reasons like these, there are no inevitable concepts, some concepts can still be inescapable in the sense characterized: these are not cases of rational revision or replacement. Compare too the notion of *rational* inevitability: a concept is rationally inevitable if, for every thinker, it is a failure of rationality not to have it. Some concepts could be rationally inevitable even if none are outright inevitable. But even rational inevitability is a very strong condition, and one may well think that it is considerably more plausible that some concepts are inescapable for some thinkers in the sense characterized than that some concepts are rationally inevitable.

I should add, before proceeding, that what matters from Hofweber's perspective when he discusses rationality is what is rational *by one's own lights* – what is rational from the

---

[19] Hofweber's main work on inescapability, his (forthcoming) and (manuscript), is as yet unpublished, but Hofweber discusses some of these themes, in relation to logic, in his (2021).

[20] Hofweber compares inescapability and inevitability in his (forthcoming) and (manuscript) and makes the points I am making in the main text.

agent's own perspective.[21] He emphasizes that he is operating with a notion of rationality that is in this way subjective. Doing so seems reasonable in the context. We want to focus on what to do from the agent's own perspective, not, for example, what is in fact conducive to externally given goals. (And if we understood rationality externalistically it would be more plausible to take some concepts to be rationally inevitable.)

A complication regarding the first characterization of inescapability has to do with scenarios like where a villain holds you hostage and says, "I will torture you unless you revise concept C".[22] To get around these complications one might instead ask: is the concept's worth *in and of itself* such that it would not be rational to revise it? The idea behind such a reformulation would be that appeal to external factors like the villain's threats can then be set aside. But an "in and of itself"-formulation threatens to block off too much. A concept's value often resides in what it lets us do, and that inevitably involves external factors. One might think that a clause like "for the purposes of inquiry" could do the trick. If I revise concept C because of the villain's threat I don't do so for the purposes of inquiry. Yet a clause like this allows for the relevance of the right external factors. But a revised version of the villain case shows that this will not do. What if the villain says to me "I will give you vast new intellectual capacity (and a vast amount of research funding) if you revise concept C"? Having noted the problems regarding how to formulate the condition for being inescapable, I will simply leave the matter here, assuming there must be some way or other to appropriately finesse the matter. There are connections here to the "wrong kind of reason"-type problems prominently discussed elsewhere in the literature.[23]

---

[21] This is a theme he repeatedly comes back to in his (2021) discussion of logic. It is a nice question whether this might require the agent to have a concept of rationality. But I think the answer to this question is no: I can assess what is rational from a dog's perspective without taking the dog itself to have a concept of rationality.

[22] This is a kind of complication that Hofweber discusses.

[23] See, e.g., Hieronymi (2005).

When we consider the possible limits of conceptual engineering, focusing on inescapability promises to be more relevant than focusing on magnetism or focusing on the earlier arguments due to Chalmers and Eklund promise to be. If some concepts really are inescapable in Hofweber's sense, then that seems to amount to a limit to conceptual engineering in the kind of sense we have been after. However, I will soon go on to problematize this.

6. Examples

Here are some examples of concepts that Hofweber (forthcoming) and (manuscript) takes to be inescapable. The general idea of inescapability does not need to be bound up with these examples, but focusing on examples is still helpful.

First, certain logical concepts. Suppose I have the concept classical negation, NOT$_C$. I can then prove to myself that double negation elimination (DNE) – meaning the inference from $\sim_c\sim_c p$ to p (where "$\sim_c$" stands for NOT$_C$) – is valid. This means that if I replace classical negation by something else, I lose a valid form of inference – and that would be bad. Suppose for example that I contemplate replacing the NOT$_C$, for which DNE is valid, with the concept intuitionistic negation, NOT$_I$, for which it is not valid (meaning: the inference from $\sim_i\sim_i p$ to p is not valid).[24] Since the switch would involve losing a valid form of inference, the switch would by my lights be irrational. The considerations are general, so the conclusion is that if I have and employ classical negation, that concept is inescapable for me.

Second example: certain normative concepts. Suppose I consider whether I should replace my concept OUGHT with an alternative, OUGHT*, and that I recognize the alternative to differ in extension from OUGHT. It seems that I can easily conclude that it would be irrational

---

[24] It is a presupposition of the argument as stated that there are two different negation concepts, one obeying classical rules and one obeying intuitionistic rules. I find this plausible myself and will not question it. But some theorists would deny this presupposition.

for me to give up my concept OUGHT in favor of OUGHT*. If I were to switch then I would be considering what I OUGHT* to do, and if my thinking would be effective then I would do that. But what I OUGHT* to do is not always what I ought to do. I might then end up doing things that I ought not to do even when I reason perfectly with the concepts that I would then employ. Switching seems then to lead me astray. I would end up doing what I ought not to do. It does not seem rational to decide on a course of action that likely has this consequence.[25]

The examples and the reasoning are from Hofweber, and I use these specific examples for the reason that they seem like the best kinds of examples of inescapable concepts.[26] But do they work? I will first discuss some reasons for doubt that are specific to each of these cases. Then I will bring up some general problems regarding what inescapability amounts to.

Here, first, is a complication specific to the logic example. What the argument regarding DNE can show that DNE, stated employing my actual classical concept NOT$_C$, is valid. What seems clearly to be true is that I would be irrational if I were to keep using that concept while refusing to employ DNE. But of course that is not what is at issue in the context. The question is whether to employ that concept at all. And that question is not answered that easily. Compare perhaps: I could find myself in a position where I employ a concept # such that P#Q is always true, regardless of the truth values of P and of Q, and the inference from P to P#Q is valid. The concept # is then rather trivial. So dropping the concept # means that I lose a valid form of inference. But this is consistent with # being eminently droppable. For all that has been said, it could be a rather useless concept. The general point is that the argument that DNE – meaning DNE for the negation concept I currently have – is valid is far from an ultimately compelling argument for retaining that concept.

---

[25] As Hofweber notes, the OUGHT/OUGHT* example is similar to what is discussed in Eklund (2015) and Eklund (2017). However, Hofweber does not make the same use of the example as Eklund does.
    Note that in the argument of Eklund's described above, it is stipulated that the alternatives to OUGHT are associated with the same normative role as OUGHT. Hofweber is more liberal in this regard.
[26] Hofweber discusses the logic case in (2021), (forthcoming) and (manuscript), and he discusses the normative case in his (forthcoming) and (manuscript).

Turn now to normative concepts. Since it is OUGHT and not OUGHT* that it is my actual concept for guiding action, then in order to raise the question of what concept to use it is OUGHT that I must use. The alternative would be not to raise the question at all. An intuitive complication regarding the example is that in a context where I am deliberating over whether to employ OUGHT or OUGHT* it would be odd for me to be satisfied by telling myself "I ought to use OUGHT rather than OUGHT*". The possibility of instead asking whether I ought* to use OUGHT or OUGHT* should be rather salient . But what Hofweber would emphasize is that OUGHT is the concept I find myself with, so even if the possibility of asking these other questions is salient those questions are not my questions.[27]

A complication of a different nature regarding the supposed inescapability of OUGHT is this. Suppose I reasonably believe of myself that I am bad at reasoning, in such a way that even if OUGHT and OUGHT* are not coextensive, for me to deliberate about what I ought* to do and then to go on to do what I conclude I ought* to do is a more reliable way for me to end up doing what I *ought* to do than employing OUGHT itself would be. Then it seems I can argue in favor of replacing OUGHT by OUGHT* as the concept for me to use in the relevant kind of deliberation.

The above remarks and complications are specific to the specific examples of inescapable concepts that I have brought up. What about more general considerations?

Here is a first more general consideration, what one might call *the acquisition argument*, purporting to problematize any case with the same structure as Hofweber's examples: where actual concepts compete with alternatives and what concepts it is rational for an agent to use is tied to what concepts she finds herself starting with:

The acquisition argument. I can rationally (or at least without irrationality) come to *acquire* and *be able to employ* some new concept or set of concepts – whether or not it is rational to use them for any purpose at all. Then, just as I can give an argument using my old concepts for the employment or these old concepts, I can give an argument using my new concepts for the employment of the new concepts. But then I am in a symmetric situation with respect to what concepts to employ: I can use some of my concepts (the old ones) in an argument for employment of the old ones and use others (the new ones) in an argument for employment of the new concepts. There is a stalemate.

The first point can be strengthened. One may think there is prima facie reason to acquire and be able to employ many concepts, for that allows one to take up more perspectives. Again to use the tool analogy from earlier: acquiring more tools is better, for that increases the options regarding what tool to use in a given situation. This is so even if some tools acquired turn out never to be fit to use. In the case of real, physical tools, there is the issue that they may be heavy to carry around, but this complication does not apply to concepts.

Even if the thought behind the acquisition argument is generally reasonable there may be complications when it comes to applying it in certain specific cases. Consider again OUGHT. Matthew Vermaire (2021) has recently argued that when one employs OUGHT one undertakes the following commitments:

(1) If you affirm that S ought to φ, then you're committed to approving of S's φ-ing.

(2) If you deny that S ought to φ, then you're committed to refraining from approving of S's φ-ing.[28]

And if a concept OUGHT* is associated with the same normative role as OUGHT, then when one employs OUGHT* one undertakes corresponding commitments:

(1*) If you affirm that S ought* to φ, then you're committed to approving of S's φ-ing.

(2*) If you deny that S ought* to φ, then you're committed to refraining from approving of S's φ-ing.

This does not mean that it is irrational to *have* both concepts, so the very first part of the acquisition argument is unscathed. But it does mean that it is irrational to, for some S and some φ, affirm that S ought to φ and deny that S ought* to φ (as Vermaire stresses). This is what we get when we start *employing* the two concepts, for example when making judgments about what concepts ought/ought* to be employed. So even if we can acquire new concepts without irrationality, employment is a different matter.[29]

7. Relative to a purpose

A more general complication regarding inescapability as Hofweber discusses it is that it seems that one must *relativize to purposes*.

Focus first on the logic case. If or when my main aim is *to reason validly* – to not happen to go from true premises to false conclusions – then concluding that DNE is valid

---

[28] I should note that Vermaire does not actually commit to OUGHT working this way. He is more cautious. Also, there are some problems with the suggestion as it stands. For example, can't there be acts such that it is not true that one ought to perform them but it is still permissible to approve of performing them?

[29] Vermaire (2021) is a critical discussion in Eklund (2017). Eklund (forthcoming) responds to Vermaire. The use made of Vermaire's ideas in the text is independent of the debate between Vermaire and Eklund.

could at least reasonably be thought to settle the matter in favor of employing classical negation (even though the specific concern that I raised above regarding this example of course remains). But where my main aim is, for example, to believe loads of truths even at the expense of believing a few falsehoods, then other concepts, governed by other rules, not necessarily truth-preserving, may be better. Sometimes, the aims may be in conflict – like when I ask the general question *what, all-things-considered, ought I to believe?* But at least often the aim simply varies with context.

Suppose my actual logical expressions and my actual truth predicate are governed by rules that do *not* lead to contradiction. Depending on my overall epistemic aims it can still be rational for me to revise my logical expressions and/or my truth predicate and come to employ concepts governed by rules that jointly do lead to contradiction: I can for example reason that employing concepts goverened by the inconsistent rules would help me reason more *efficiently* in ordinary cases (perhaps the rules I actually employ are more complex), and the inconsistency only comes to light in special, rare circumstances. Employing quick and dirty rules is overall more rational.[30] Whether revision in this case is more rational may depend on my purposes. If my aim is to avoid concluding falsehoods and contradictions at all cost the revision is not more rational. But if my aim is to come to believe many truths in an efficient way and it suffices that the ratio of truth to falsehoods is good, the revision may well be rational.

The issue of relativization to purposes is relevant also to the case of OUGHT. The natural way to read the example Hofweber uses is that it concerns the so-called *all things considered* OUGHT (or OUGHT *simpliciter* – different labels are used in the literature). The two concepts OUGHT and OUGHT* are competing all-things-considered concepts. And even assuming that Hofweber is right about what happens when I ask myself which concept I

---

[30] Compare Cherniak (1984).

should use where all-things-considered evaluation is what is at issue, it can be that sometimes my focus is on some more specific normative domain - morality/prudence/etc. – and then I can ask whether to use OUGHT or OUGHT*, meaning what to do from the point of view of morality/prudence/etc. This question is not trivially answerable.

The fact that there is an apparent need to relativize to purposes suggests that the turn to inescapability was not the improvement it might have seemed to be. The reason why the discussion from Eklund (2015) did not seem to have any immediate implications as regards limits of conceptual engineering had to do with relativization. The arguments that Eklund presents show at most that *given that we want to employ a concept of such-and-such a kind* we have no alternatives. That is relativization to a purpose. Now we see that the considerations concerning inescapability run into the same kind of problem.

Both the discussion of the acquisition argument and the considerations regarding relativization to purposes serve to highlight that there is a distinction to heed between *having* a concept and *employing* it. One might, for example, have a concept without ever employing it. One might, obviously, employ a concept for some purposes but choose not to employ it for others. It is possible to ask both which concepts it is rational to have and which concepts it is rational to employ, and arrive at different answers. (One possibility is that no concept is such that it is irrational to *have* it, but some concepts are such that it is irrational for a given thinker to *employ* them.)

A different issue that arises when the need to relativize to purposes is taken into account concerns how finely to individuate purposes. Focus on the logic example. One purpose one might have is to engage in *deduction*; another purpose one might have is to engage in *classically valid deduction*. One purpose one might have is just to act well in some *general* sense; another purpose one might have is more specifically tied to acting *morally*. Of course, we only rarely explicitly spell out to ourselves what our purposes are. One might

describe an agent and her behavior both in the more general and the more specific terms. It may be that a concept I use is indispensable for a finely individuated purpose (e.g., engaging in classically valid deduction) but not for a corresponding more coarsely individuated purpose (e.g. deduction). I will not attempt to make progress on the issue of purpose individuation. For now I just want to flag the issue and note its potential importance.

Given the seeming need to relativize to purposes, any characterization of inescapability must take this into account. Here is a first, simple suggestion:

A concept is inescapable for a thinker if there is some purpose such that employing it is indispensable for that purpose.[31]

There are presumably good questions about just what "indispensable" means in a characterization like this, and "indispensable" is of course not a mile away from "inescapable". But the issues I will go on to discuss are actually orthogonal to exactly how to resolve such questions and I will simply rely on an intuitive notion of indispensability.

This first suggestion is obviously problematic. First, there is a triviality worry. For every concept C, there is the purpose of thinking C-thoughts, and C seems indispensable for that purpose. Second, if a thinker herself fails to have any purposes for which C is indispensable, then C is hardly in any intuitively reasonable sense inescapable for the thinker. A somewhat improved suggestion is:

A concept is inescapable for a thinker if there is some purpose *that the thinker has* such that employing it is indispensable for that purpose.

---

[31] This and the following proposals only state sufficient conditions for inescapability. Of course there is also a question of what is necessary for inescapability. But one thing at a time.

This gets around the second problem. But it still remains that this is somewhat cheap. What if among the purposes a given thinker has is the purpose of thinking C-thoughts and she has and employs concept C for that purpose, but simple reflection would convince the thinker that there is no real point to thinking C-thoughts as opposed to other kinds of thoughts? Concept C then does not intuitively seem inescapable for that thinker – but it is classed as inescapable for her given the suggestion at issue.

The simplest, quickest way to deal with this problem is to restrict which purposes are at issue. One might restrict to *fine* (good, reasonable) purposes and say e.g.:

A concept is inescapable for a thinker if there is some *fine* purpose *that the thinker has* such that employing it is indispensable for that purpose.

One might then, for example, argue that for some or many concepts C, the purpose of thinking C-thoughts and concluding C-truths is not a fine purpose.

Of course, "fine", as used here, is vague and schematic. But on many natural ways of precisifying it, a purpose may be rationally revisable despite being fine. My old purpose may have been fine but I can rationally decide to replace it by purposes that seem even more worthwhile from my perspective. So even if appeal to "fine" purposes is an improvement over earlier attempts, it does not seem to go far enough when our topic is limits to conceptual engineering. Consider the following stronger notion, *absolute inescapability*:

A concept is absolutely inescapable if it is indispensable for an *inescapable* purpose that one has.

A purpose is here inescapable if one cannot rationally revise or replace it. So long as some purposes we have may be fine but yet escapable, absolute inescapability is a stronger notion than mere inescapability. And if some purposes we have are inescapable, and some concepts we use are inescapable for those purposes, then we have arrived at non-conditional limits to conceptual engineering after all.

Are some purposes we have inescapable? Let me close by just making some remarks on this.

In their important (2008), David Enoch and Joshua Schechter ask the question of how basic belief-forming methods, like (to mention some cases they centrally discuss) standard inductive method, inference to the best explanation, and basic deductive rules such as modus ponens, are justified. Their general idea is the following:

> There are certain projects in which we rationally ought to engage. They are, as we will call them, *rationally required* projects. Such projects plausibly include explaining the world around us, deliberating about what to do, planning for the future, and evaluating our own patterns of thinking. We are justified in employing any belief-forming method needed for successfully engaging in a rationally required project.[32]

Here is an example illustrating the idea:

> Consider the explanatory project, the project of understanding and explaining the world around us. This project is of fundamental importance to us. Indeed, it seems that engaging in this project is central to rationality; a thinker who does not inquire about the world around him is intuitively doing something wrong. This counts in favor of

[32] Enoch and Schechter (2008), p. 549f.

employing whatever methods are necessary for successfully engaging in the explanatory project.[33]

This certainly sounds like it could provide us with examples of inescapable purposes: among the inescapable purposes are the purposes of engaging in rationally required projects. One example might be what they call the explanatory project. But here are two general problems.

First, as noted, in his discussions of inescapability, Hofweber makes clear that the notion of rationality he is employing is *subjective*: the issue is what is rational from the agent's epistemic perspective. And it is clear why Hofweber approaches things that way. He is talking about the cognitive situation from the point of view of the agent, describing how the agent herself can reason about conceptual revision. By contrast, Enoch and Schechter's account is, as they note, externalist, and it is so in part by relying on facts, not necessarily available to the thinker, about which projects really are "rationally required".[34]

Second, the question of the individuation of purposes is very much with us. Focus on what Enoch and Schechter call the explanatory project. Is our concept EXPLANATION indispensable for engaging in that project, and hence inescapable for that purpose? That depends on how broadly or narrowly we think of "the explanatory project". If, narrowly, the explanatory project is tied to explanation in the sense of our actual concept EXPLANATION then our concept EXPLANATION may indeed be inescapable for the purpose of engaging in the explanatory project. But with the explanatory project thus narrowly understood it is somewhat doubtful that this project is inescapable. If the explanatory project is understood more broadly, so that one engages in this project so long as one aims at something which is broadly explanation-like, then it is more plausible that the project is inescapable, but it is correspondingly less plausible that our concept EXPLANATION is inescapable for it.

---

[33] Enoch and Schechter (2008), p. 549.
[34] Enoch and Schechter (2008), pp. 567ff.

The case that Enoch and Schechter discuss that is most similar to what Hofweber discusses regarding classical negation concerns modus ponens. Here is their argument why we are justified in employing modus ponens:

> Executing contingency plans requires drawing the relevant Modus Ponens inferences. So engaging in the project of planning for contingencies requires employing MP. We cannot successfully engage in this project unless MP is effective; if MP is unreliable, the project of planning for contingencies is doomed to systematic failure. It is in virtue of this that we are justified in employing MP.[35]

But what this kind of argument could establish is at most that it is rational for us to employ a connective $\rightarrow$ such that *generally* speaking, Q follows from P and P$\rightarrow$Q. An argument like that Enoch and Schechter provide can hardly establish that this reasoning is deductively valid.

8. Concluding remarks

Both Sellars' contrast between the manifest image and the scientific image and theoretical reflection on conceptual engineering invite the question: what sorts of limits are there to the rational revision of concepts? I discussed several ways of approaching this issue, finding the most promising approach to be to discuss things in terms of Hofweber's notion of inescapability. However, the discussion of this approach showed, among other things, that problems familiar from earlier in the discussion arise here too. Specifically, we run into difficult questions regarding the inescapability of different purposes for which we employ concepts.

---

[35] Enoch and Schechter (2008), p. 555.

REFERENCES

Belleri, Delia: forthcoming, "Downplaying the Change of Subject Objection to Conceptual Engineering", *Inquiry*.

Burgess, Alexis, Herman Cappelen, and David Plunkett (eds.): 2020, *Conceptual Engineering and Conceptual Ethics*, Oxford University Press.

Cappelen, Herman: 2018, *Fixing Language*, Oxford University Press.

Cappelen, Herman and David Plunkett: 2020, "Introduction: A Guided Tour of Conceptual Engineering and Conceptual Ethics", in Burgess, Cappelen and Plunkett (2020).

Chalmers, David: 2011, "Verbal Disputes", *Philosophical Review* 120: 515-66.

Chalmers, David: forthcoming, "What is Conceptual Engineering and What Should It Be?", *Inquiry*.

Cherniak, Christopher: 1984, "Computational Complexity and the Universal Acceptance of Logic", *Journal of Philosophy* 81, 739-58.

Eklund, Matti: 2015, "Intuitions, Conceptual Engineering, and Conceptual Fixed Points", in Christopher Daly (ed.), *The Palgrave Handbook of Philosophical Methods*, Palgrave Macmillan.

Eklund, Matti: 2017, *Choosing Normative Concepts*, Oxford University Press.

Eklund, Matti: 2021, "Conceptual Engineering in Philosophy", in Justin Khoo and Rachel Sterken (eds.), *The Routledge Handbook of Social and Political Philosophy of Language*, Routledge.

Eklund, Matti: forthcoming, "Schmoughts for Naught? Reply to Vermaire", *Journal of Philosophy*.

Enoch, David and Joshua Schechter: 2008, "How are Basic Belief-Forming Methods Justified?", *Philosophy and Phenomenological Research* 76: 547-79.

Gibbard, Allan: 2003, *Thinking How to Live*, Harvard University Press, Cambridge, Massachusetts.

Hieronymi, Pamela: 2005, "The Wrong Kind of Reason", *Journal of Philosophy* 102: 437-57.

Hofweber, Thomas: 2021, "The Unrevisability of Logic", *Philosophical Perspectives* 35: 251-74.

Hofweber, Thomas: forthcoming, *Idealism and the Harmony of Thought and Reality*, Oxford University Press.

Hofweber, Thomas: manuscript, "Inescapable Concepts".

Isaac, Manuel Gustavo: forthcoming, "Which Concept of Concept for Conceptual Engineering?", *Erkenntnis*.

Jorem, Sigurd: 2021, *Crafting Tools for Thought and Talk: Essays in the Theoretical Foundations of Conceptual Engineering*, doctoral dissertation, University of Oslo.

Jorem, Sigurd: 2021a, "Conceptual Engineering and the Implementation Problem", *Inquiry* 64: 186-211.

Koch, Steffen: 2021, "Engineering What? On Concepts in Conceptual Engineering", *Synthese* 199: 1955-75.

Lewis, David: 1983, "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 61: 343–77.

Lewis, David: 1984, "Putnam's Paradox", *Australasian Journal of Philosophy* 62: 221–36.

Nado, Jennifer: forthcoming, "Taking Control: Conceptual Engineering Without (Much) Metasemantics", *Inquiry*.

Nimtz, Christian: forthcoming, "Engineering Concepts by Engineering Social Norms: Solving the Implementation Challenge", *Inquiry*.

Pinder, Mark: 2021, "Conceptual Engineering, Metasemantic Externalism, and Speaker-Meaning", *Mind* 130: 141-63.

Richard, Mark: 2020, "The A-project and the B-project", in Burgess et al. (2020).

Sellars, Wilfrid: 1962, "Philosophy and the Scientific Image of Man," in Robert Colodny
(ed.), *Frontiers of Science and Philosophy*, University of Pittsburgh Press, pp. 35–78

Sterken, Rachel Katherine: 2020, "Linguistic Intervention and Transformative
Communicative Disruptions", in Burgess et al. (2020).

Sundell, Tim: 2020, "Changing the Subject", *Canadian Journal of Philosophy* 50: 580-
93.

Vermaire, Matthew: 2021, "Against Schmought", *Journal of Philosophy* 118: 465-85.

Wittgenstein, Ludwig: 1997, *Philosophical Investigations*, Blackwell, Oxford. Translated by
Elizabeth Anscombe. First published 1953.